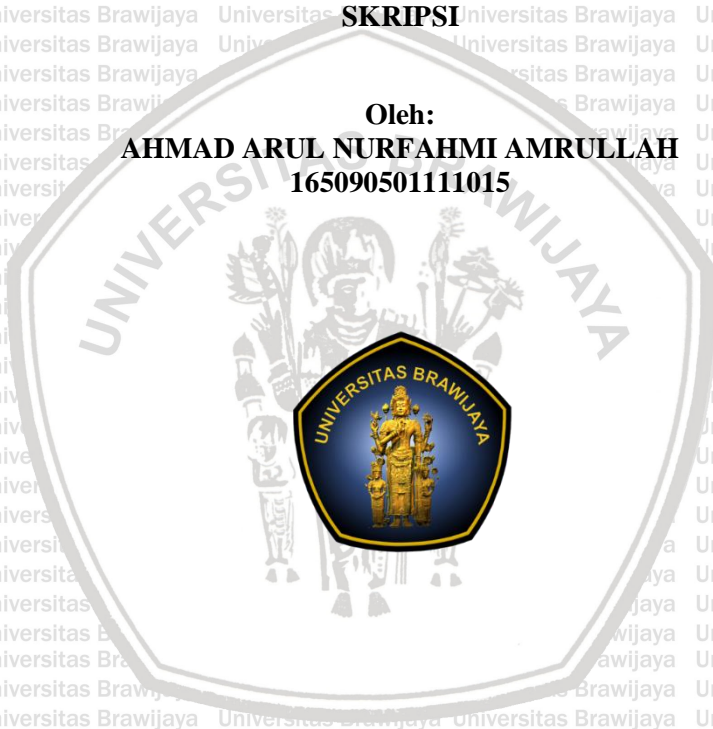


**AKURASI KLASIFIKASI MULTIVARIATE ADAPTIVE
REGRESSION SPLINE (MARS) PADA DATA SIMULASI
CLASS IMBALANCE PEUBAH RESPON BINER
MENGUNAKAN METODE *BOOTSTRAP* AGGREGATING
(*BAGGING*) DAN *SYNTHETIC MINORITY OVER-SAMPLING*
TECHNIQUE (SMOTE)**

SKRIPSI

Oleh:

**AHMAD ARUL NURFAHMI AMRULLAH
165090501111015**



PROGRAM STUDI SARJANA STATISTIKA

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN

ALAM

UNIVERSITAS BRAWIJAYA

MALANG

2019

HALAMAN JUDUL

AKURASI KLASIFIKASI *MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)* PADA DATA SIMULASI *CLASS IMBALANCE* PEUBAH RESPON BINER MENGGUNAKAN METODE *BOOTSTRAP AGGREGATING (BAGGING)* DAN *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)*

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh:

AHMAD ARUL NURFAHMI AMRULLAH

165090501111015



PROGRAM STUDI SARJANA STATISTIKA

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN

ALAM

UNIVERSITAS BRAWIJAYA

MALANG

2019



LEMBAR PENGESAHAN SKRIPSI
AKURASI KLASIFIKASI *MULTIVARIATE ADAPTIVE*
***REGRESSION SPLINE* (MARS) PADA DATA SIMULASI**
***CLASS IMBALANCE* PEUBAH RESPON BINER**
MENGGUNAKAN METODE *BOOTSTRAP AGGREGATING*
(BAGGING)* DAN *SYNTHETIC MINORITY OVER-SAMPLING
***TECHNIQUE* (SMOTE)**

Oleh:

AHMAD ARUL NURFAHMI AMRULLAH
165090501111015

Setelah dipertahankan di depan Majelis Penguji
pada tanggal 20 Desember 2019
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Statistika

Pembimbing

Achmad Efendi, S.Si., M.Sc., Ph.D.
NIP. 198102192005011001

Mengetahui,
a.n. Ketua Jurusan Statistika
Fakultas MIPA Universitas Brawijaya
Sekretaris,

Nurjannah, S.Si., M.Phil., Ph.D
NIP. 198009212005012001

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Ahmad Arul Nurfahmi Amrullah

NIM : 165090501111015

Jurusan : Statistika

Judul Skripsi :

AKURASI KLASIFIKASI *MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)* PADA DATA SIMULASI *CLASS IMBALANCE* PEUBAH RESPON BINER MENGGUNAKAN METODE *BOOTSTRAP AGGREGATING (BAGGING)* DAN *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)*

Dengan ini menyatakan bahwa:

1. Isi dari skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam Skripsi ini.
2. Apabila dikemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala resiko yang akan saya terima.

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 19 Desember 2019

Yang menyatakan,

Ahmad Arul Nurfahmi Amrullah

165090501111015

AKURASI KLASIFIKASI MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) PADA DATA SIMULASI CLASS IMBALANCE PEUBAH RESPON BINER MENGGUNAKAN METODE *BOOTSTRAP AGGREGATING (BAGGING)* DAN *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)*

ABSTRAK

Model *Multivariate Adaptive Regression Spline* (MARS) termasuk salah satu metode dalam regresi nonparametrik yang populer digunakan dalam permasalahan prediksi maupun klasifikasi. Permasalahan yang muncul dalam metode klasifikasi adalah adanya ketidakseimbangan banyaknya pengamatan antar kelas pada peubah respon yang cukup jauh sehingga terbentuk kelas data mayoritas dan kelas data minoritas. Model klasifikasi MARS pada umumnya efektif dalam mengklasifikasikan data yang tidak seimbang ini, karena model MARS merupakan model yang adaptif yaitu mampu membentuk model dengan cara menyesuaikan pola data yang ada. Untuk mengurangi kesalahan klasifikasi kelas minoritas pada data tidak seimbang dan untuk meningkatkan akurasi atau kekuatan klasifikasi pada model MARS dapat menggunakan dua metode yaitu *Bootstrap Aggregating (Bagging)* dan *Synthetic Minority Over-sampling Technique* (SMOTE). Penelitian ini bertujuan untuk mengetahui ketepatan klasifikasi menggunakan metode MARS, *Bagging*-MARS dan SMOTE-MARS pada data simulasi yang memiliki ketidakseimbangan banyaknya pengamatan antar kelas dengan proporsi kelas minoritas sebesar 10%, 15% dan 20%. Hasil penelitian ini menunjukkan ketepatan klasifikasi pada metode *Bagging*-MARS memiliki kekuatan klasifikasi yang paling bagus dibandingkan metode MARS dan SMOTE-MARS dengan melihat nilai akurasi, sensitifitas dan APER yang dihasilkan pada setiap tingkat proporsi kelas minoritas.

Kata kunci: MARS, *Bagging*, SMOTE, Ketidakseimbangan Kelas.



**ACCURACY MULTIVARIATE ADAPTIVE REGRESSION
SPLINE (MARS) FOR CLASSIFICATION OF SIMULATION
DATA IN CLASS IMBALANCE BINARY RESPONSE
VARIABLES USING BOOTSTRAP AGGREGATING
(BAGGING) AND SYNTHETIC MINORITY OVER-SAMPLING
TECHNIQUE (SMOTE)**

ABSTRACT

The Multivariate Adaptive Regression Spline (MARS) model is one of the methods in nonparametric regression that is popularly used in both prediction and classification problems. The problem that arises in the classification method is the imbalance of the number of observations between classes on the response variables far enough to form the majority data class and the minority data class. The MARS classification model is generally effective in classifying this unbalanced data, because the MARS model is an adaptive model that is able to form a model by adjusting existing data patterns. To reduce misclassification of minority classes on unbalanced data and to improve the accuracy or strength of classification on the MARS model, two methods can be used, namely Bootstrap Aggregating (Bagging) and Synthetic Minority Over-sampling Technique (SMOTE). This study aims to determine the accuracy of classification using the MARS, Bagging-MARS and SMOTE-MARS methods on simulation data that has a lot of imbalance between class observations with the proportion of minority classes by 10%, 15% and 20%. The results of this study indicate that the classification accuracy of the Bagging-MARS method has the best classification strength compared to the MARS and SMOTE-MARS methods by looking at the accuracy, sensitivity and APER values produced at each level of proportion of the minority classes.

Keywords: *MARS, Bagging, SMOTE, Class Imbalance.*





KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas berkat, rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Akurasi Klasifikasi *Multivariate Adaptive Regression Spline* (MARS) pada Data Simulasi *Class Imbalance* Peubah Respon Biner Menggunakan Metode *Bootstrap Aggregating (Bagging)* dan *Synthetic Minority Over-Sampling Technique (SMOTE)*”. Skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Statistika. Dalam penyusunan skripsi ini, banyak pihak telah memberikan bantuan maupun dukungan kepada penulis. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Bapak Achmad Efendi, S.Si., M.Sc., Ph.D. selaku dosen pembimbing atas waktu, bimbingan, kritik dan saran yang telah diberikan kepada penulis dalam penyusunan skripsi.
2. Ibu Rahma Fitriani, S.Si., M.Sc., Ph.D. selaku dosen penguji I yang telah memberikan banyak saran, arahan dan bimbingan kepada penulis dalam penyusunan skripsi.
3. Ibu Dr. Eni Sumarminingsih, S.Si., M.M. selaku dosen penguji II atas saran, arahan dan bimbingan yang diberikan kepada penulis dalam penyusunan skripsi.
4. Ibu Rahma Fitriani, S.Si., M.Sc., Ph.D. selaku Ketua Jurusan Statistika Fakultas MIPA Universitas Brawijaya.
5. Semua dosen, staf dan karyawan Jurusan Matematika Fakultas MIPA Universitas Brawijaya.
6. Keluarga tercinta, Bapak dan Ibu atas dukungan dan doa yang diberikan selama ini kepada penulis.
7. Teman-teman Statistika UB angkatan 2016 yang membantu dalam menyelesaikan penelitian ini dan yang telah memberikan doa serta semangat kepada penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, dengan segala kerendahan hati penulis mengharapkan kritik dan saran yang membangun demi perbaikan. Semoga skripsi ini bermanfaat bagi pembaca.

Malang, 19 Desember 2019

Penulis



DAFTAR ISI

	Hal.
HALAMAN JUDUL.....	iii
LEMBAR PENGESAHAN SKRIPSI.....	v
LEMBAR PERNYATAAN.....	vii
ABSTRAK.....	ix
<i>ABSTRACT</i>	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL.....	xix
DAFTAR LAMPIRAN.....	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian.....	3
1.4. Batasan Masalah.....	3
1.5. Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1. Klasifikasi.....	5
2.2. Ukuran Ketepatan Klasifikasi.....	6
2.3. Regresi Nonparametrik.....	6
2.4. <i>Multivariate Adaptive Regression Spline (MARS)</i>	7
2.4.1. Dekomposisi Analisis Ragam Model MARS.....	10
2.4.2. Multikolinieritas dalam Model MARS.....	12
2.4.3. Pendugaan Parameter Model MARS.....	13
2.4.4. Pengujian Parameter Model MARS.....	14
2.4.5. Pemilihan Model Terbaik MARS.....	16



2.4.6. Klasifikasi dengan Model MARS.....	17
2.5. <i>Bootstrap Agregating (Bagging)</i>	18
2.6. <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	19
BAB III METODE PENELITIAN	21
3.1. Sumber Data.....	21
3.2. Proses Pembangkitan Data.....	22
3.3. Metode Analisis Data.....	24
3.4. Diagram Alir.....	27
BAB IV HASIL DAN PEMBAHASAN	29
4.1. Proses Simulasi.....	29
4.2. Hasil Simulasi.....	31
4.2.1. Akurasi.....	31
4.2.2. Sensitifitas.....	35
4.2.3. APER.....	39
4.2.4. GCV.....	43
4.3. Pembahasan Hasil Simulasi.....	45
BAB V KESIMPULAN DAN SARAN	49
5.1. Kesimpulan.....	49
5.2. Saran.....	49
DAFTAR PUSTAKA	51
LAMPIRAN	55



DAFTAR GAMBAR

Hal.

Gambar 3.1. Persentase tiap kelas pada peubah respon kolektibilitas kredit data asli.....21

Gambar 3.2. Diagram Alir Pembangkitan Data.....27

Gambar 3.3. Diagram Alir Metode Analisis.....28

Gambar 4.1. Rata-rata akurasi (a) dan standar deviasi akurasi (b) pada data latih.....32

Gambar 4.2. Rata-rata akurasi (a) dan standar deviasi akurasi (b) pada data uji.....34

Gambar 4.3. Rata-rata sensitifitas (a) dan standar deviasi sensitifitas (b) pada data latih.....36

Gambar 4.4. Rata-rata sensitifitas (a) dan standar deviasi sensitifitas (b) pada data uji.....38

Gambar 4.5. Rata-rata APER (a) dan standar deviasi APER (b) pada data latih.....40

Gambar 4.6. Rata-rata APER (a) dan standar deviasi APER (b) pada data uji.....42

Gambar 4.7. Rata-rata GCV (a) dan standar deviasi GCV (b).....44

Gambar 4.8. Plot Akurasi Metode MARS (a), *Bagging*-MARS (b) dan SMOTE-MARS (c).....46

Gambar 4.9. Plot Sensitifitas Metode MARS(a),*Bagging*-MARS(b) dan SMOTE-MARS(c).....47

Gambar 4.10. Plot APER Metode MARS (a), *Bagging*-MARS (b) dan SMOTE-MARS (c).....48





DAFTAR TABEL

Hal.

Tabel 2.1. Tabel Kontingensi5

Tabel 2.2. Tabel Analisis Ragam..... 15

Tabel 4.1. Banyak pengamatan, ukuran sampel (n), tingkat proporsi kelas minoritas (p), peubah dan sebaran yang digunakan dalam proses simulasi.....30

Tabel 4.2. Rata-rata akurasi yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latihan.....31

Tabel 4.3. Standar deviasi akurasi yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latihan.....32

Tabel 4.4. Rata-rata akurasi yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji33

Tabel 4.5. Standar deviasi akurasi yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji34

Tabel 4.6. Rata-rata sensitifitas yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latihan.....35

Tabel 4.7. Standar deviasi sensitifitas yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latihan.....36

Tabel 4.8. Rata-rata sensitifitas yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji37

Tabel 4.9. Standar deviasi sensitifitas yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji38

Tabel 4.10. Rata-rata APER yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latihan.....39



Tabel 4.11. Standar deviasi APER yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latih..... 40

Tabel 4.12. Rata-rata APER yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji 41

Tabel 4.13. Standar deviasi APER yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji..... 42

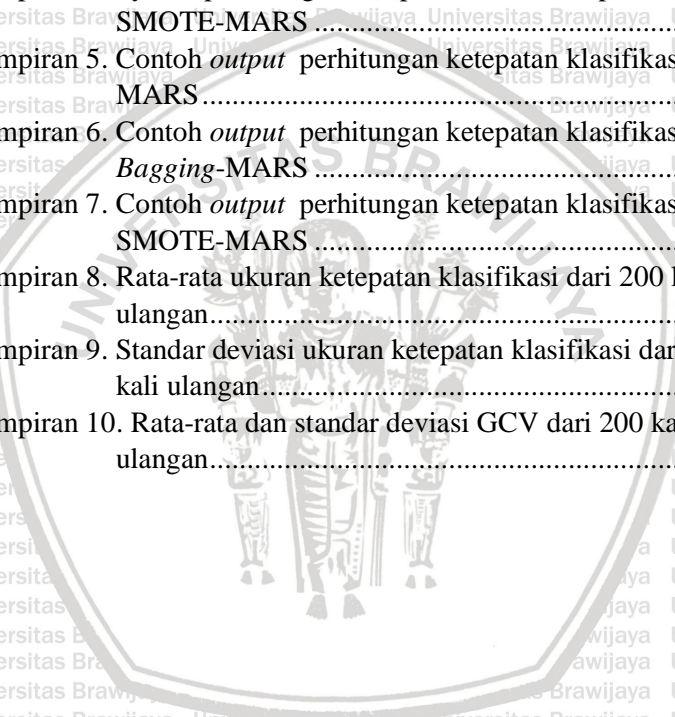
Tabel 4.14. Rata-rata GCV yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel 43

Tabel 4.15. Standar deviasi GCV yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel 44



DAFTAR LAMPIRAN

	Hal.
Lampiran 1. Syntax R untuk membangkitkan data tidak seimbang ..	55
Lampiran 2. Syntax perhitungan ketepatan klasifikasi pada metode MARS	57
Lampiran 3. Syntax perhitungan ketepatan klasifikasi pada metode <i>Bagging</i> -MARS	58
Lampiran 4. Syntax perhitungan ketepatan klasifikasi pada metode SMOTE-MARS	60
Lampiran 5. Contoh <i>output</i> perhitungan ketepatan klasifikasi metode MARS	62
Lampiran 6. Contoh <i>output</i> perhitungan ketepatan klasifikasi metode <i>Bagging</i> -MARS	64
Lampiran 7. Contoh <i>output</i> perhitungan ketepatan klasifikasi metode SMOTE-MARS	66
Lampiran 8. Rata-rata ukuran ketepatan klasifikasi dari 200 kali ulangan	69
Lampiran 9. Standar deviasi ukuran ketepatan klasifikasi dari 200 kali ulangan	71
Lampiran 10. Rata-rata dan standar deviasi GCV dari 200 kali ulangan	73





BAB I PENDAHULUAN

1.1. Latar Belakang

Analisis regresi merupakan salah satu metode statistika yang umum digunakan untuk melihat hubungan fungsional antara peubah respon yang dipengaruhi oleh peubah prediktor. Terdapat dua pendekatan yang dapat digunakan dalam menyelesaikan analisis regresi, yaitu regresi parametrik dan nonparametrik. Regresi parametrik adalah regresi yang pola data dan kurva regresi sudah diketahui, sedangkan regresi nonparametrik adalah regresi yang diasumsikan bentuk kurva tidak diketahui dan termuat dalam ruang fungsi sehingga memberikan fleksibilitas lebih besar dalam mengestimasi bentuk kurva regresi (Eubank, 1988). Ada beberapa pendekatan regresi nonparametrik yaitu *spline*, kernel, polinomial, MARS dan *Wavelet*.

Salah satu metode regresi nonparametrik adalah *Multivariate Adaptive Regression Spline* (MARS) dikembangkan oleh Friedman (1991) yang merupakan kombinasi yang kompleks antara *Spline Regression* dan *Recursive Partitioning Regression*. *Spline Regression* merupakan salah satu metode regresi yang memiliki sifat tersegmen dan memberikan fleksibilitas lebih dibanding regresi polinomial, sehingga bentuk kurva regresi dapat menyesuaikan diri dengan efektif terhadap karakteristik data yang ada. *Recursive Partitioning Regression* merupakan model regresi bersifat tersegmen namun setiap bagian fungsi saling lepas yang menyebabkan knot tidak kontinu. Menurut Friedman (1991), model MARS berguna untuk mengatasi masalah data berdimensi tinggi, yaitu data sampel yang berukuran $50 \leq n \leq 1000$ di mana n merupakan ukuran sampel dan memiliki peubah prediktor $3 \leq p \leq 20$ di mana p merupakan banyaknya peubah. Model MARS tidak hanya digunakan dalam permasalahan prediksi suatu peubah respon namun MARS juga dapat digunakan dalam permasalahan klasifikasi.

Permasalahan yang muncul dalam metode klasifikasi adalah adanya ketidakseimbangan kelas pada peubah respon. Ketidakseimbangan ini terletak pada timpangnya proporsi banyak pengamatan antar kategori pada peubah respon yang cukup jauh sehingga terbentuk kelas data mayoritas dan kelas data minoritas. Metode klasifikasi konvensional pada umumnya kurang efektif dalam

mengklasifikasikan data yang tidak seimbang ini, karena kurang memperhatikan kelas minoritas. Oleh karena itu, model klasifikasi akan cenderung lebih lemah dalam memprediksi kelas minoritas dari pada memprediksi kelas mayoritas atau dengan kata lain banyak pengamatan pada kelas minoritas salah diklasifikasikan dari pada kelas mayoritas (Yanmin dkk., 2007). Untuk mengurangi kesalahan klasifikasi kelas minoritas pada data tidak seimbang dan untuk meningkatkan akurasi atau kekuatan klasifikasi pada model klasifikasi konvensional dapat menggunakan dua metode yaitu *Bootstrap Aggregating (Bagging)* dan *Synthetic Minority Over-sampling Technique (SMOTE)*.

Bagging merupakan metode yang dikemukakan oleh Breiman (1994). *Bagging* adalah salah satu metode *ensemble learning* untuk memperbaiki kekuatan prediksi atau akurasi dari beberapa algoritma tertentu seperti regresi atau pohon klasifikasi. *Bagging* bekerja dengan cara penggabungan (*aggregating*) sejumlah k model yang dilatih menggunakan sejumlah k data latih yang dibangkitkan secara acak menggunakan metode *Bootstrap* agar masing-masing model saling independen (Suyanto, 2018).

SMOTE merupakan salah satu metode penanganan pada data tidak seimbang yang dikemukakan oleh Chawla dkk. (2002). SMOTE bekerja dengan cara membangkitkan data buatan untuk kelas data minoritas pada peubah respon (*oversampling*) dan juga menurunkan banyak pengamatan pada kelas data mayoritas (*undersampling*) sehingga proporsi kelas mayoritas dan minoritas menjadi lebih seimbang.

Penelitian sebelumnya, Adiangga (2015), membahas tentang penanganan ketidakseimbangan kelas dengan menggunakan SMOTE dan *Cost Sensitive Learning* dalam mengklasifikasikan pekerja anak di Jakarta. Pada penelitian tersebut diperoleh hasil bahwa penggunaan SMOTE pada data tidak seimbang memberikan hasil yang optimal dan mampu meningkatkan klasifikasi pada kelas minoritas. Pada penelitian lain, Arleina dan Otok (2014), membahas tentang penggunaan *Bagging-MARS* untuk mengklasifikasikan rumah tangga miskin di Kabupaten Jombang dan diperoleh kesimpulan bahwa metode *Bagging-MARS* lebih tepat digunakan karena memiliki kekuatan klasifikasi lebih tinggi dibandingkan metode MARS tanpa *Bagging*. Berdasarkan dua hasil penelitian diatas, pada penelitian ini dilakukan simulasi ketidakseimbangan banyaknya pengamatan antar



kategori pada peubah respon biner dengan menerapkan metode *Bagging* dan SMOTE pada model MARS.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang ada, rumusan masalah pada penelitian ini adalah bagaimana perbandingan ukuran ketepatan klasifikasi metode MARS, *Bagging*-MARS dan SMOTE-MARS dari data simulasi pada setiap tingkat ukuran sampel data tidak seimbang dan setiap tingkat proporsi kelas minoritas peubah respon biner?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang ada, penelitian ini memiliki tujuan untuk membandingkan ukuran ketepatan klasifikasi metode MARS, *Bagging*-MARS dan SMOTE-MARS dari data simulasi pada setiap tingkat ukuran sampel data tidak seimbang dan setiap tingkat proporsi kelas minoritas peubah respon biner.

1.4. Batasan Masalah

Berdasarkan permasalahan diatas, batasan masalah pada penelitian ini yaitu:

1. Banyaknya maksimum fungsi basis (BF) adalah 3 kali banyaknya variabel prediktor yang ada, banyaknya maksimum interaksi (MI) yang digunakan sebanyak 2 interaksi dan minimum observasi (MO) antara knot yang digunakan sebesar satu.
2. *Resampling* menggunakan *Bootstrap* sebanyak 25 kali pada metode *Bagging*.
3. Parameter pembangkitan data sintesis (*perc.over*, *perc.under* dan *k*) pada metode SMOTE secara berurutan bernilai 900, 600 dan 5.

1.5. Manfaat Penelitian

Manfaat yang dapat diperoleh dari hasil penelitian ini yaitu:

1. Menambah pengetahuan dalam bidang statistika mengenai model MARS dalam mengklasifikasikan data tidak seimbang berbasis data simulasi *cross section*.
2. Memberikan informasi mengenai ketepatan klasifikasi menggunakan metode *Bagging* dan SMOTE pada model MARS dalam kasus ketidakseimbangan kelas.
3. Dapat dijadikan sebagai pertimbangan dalam memilih metode statistika yang tepat pada kasus klasifikasi data yang memiliki



ketidakseimbangan kelas seperti pada kasus data kolektibilitas kredit bank, kasus data pasien terkena penyakit HIV AIDS dan lain-lain.



BAB II TINJAUAN PUSTAKA

2.1. Klasifikasi

Klasifikasi adalah proses dalam mendapatkan model atau fungsi yang dapat digunakan untuk membedakan kelas pada suatu data. Klasifikasi juga dapat digunakan untuk memprediksi kelas pada data yang tidak diketahui kelasnya (Han dan Kamber, 2006). Menurut Agresti (2007), metode klasifikasi yang baik akan menghasilkan peluang kesalahan alokasi yang kecil. Metode yang biasa digunakan untuk klasifikasi antara lain *Logistic Regression*, *Discriminant Analysis*, *Naïve Bayes*, *Decision Trees*, *Support Vector Machines*, *Neural Network* dan lain-lain.

Metode klasifikasi merupakan salah satu metode *supervised learning* yang digunakan dalam memecahkan permasalahan klasifikasi pada data. Metode *supervised learning* adalah metode yang mempunyai tujuan untuk mengidentifikasi model atau fungsi sesuai *input* dan *output* yang ada pada data, yang berarti terdapat peubah target yang dispesifikasikan sebelumnya. Algoritma ini membentuk suatu fungsi yang memetakan *input* ke *output* yang diinginkan. Kualitas hasil pemetaan sangat bergantung pada kesesuaian *input* dan *output* yang diberikan. Oleh karna itu, algoritma jenis ini disebut sebagai pembelajaran terawasi atau *supervised learning* (Suyanto, 2018). Bila variabel respon bersifat kategorik biner maka penentuan ketepatan klasifikasi dapat diketahui dan disajikan melalui tabel kontingensi seperti berikut:

Tabel 2.1. Tabel Kontingensi

Hasil Observasi (<i>Actual Class</i>)	Hasil Prediksi (<i>Predicted Class</i>)	
	$\hat{y}=0$	$\hat{y}=1$
$y=0$	n_{11}	n_{12}
$y=1$	n_{21}	n_{22}

di mana:

- 0 : bila peubah respon tidak terjadi (kejadian gagal)
- 1 : bila peubah respon terjadi (kejadian sukses)
- n_{11} : banyak kejadian gagal dari observasi yang tepat diprediksikan sebagai kejadian gagal.
- n_{12} : banyak kejadian gagal dari observasi yang salah diprediksikan sebagai kejadian sukses



n_{21} : banyak kejadian sukses dari observasi yang salah diprediksikan sebagai kejadian gagal.

n_{22} : banyak kejadian sukses dari observasi yang tepat diprediksikan sebagai kejadian sukses

2.2. Ukuran Ketepatan Klasifikasi

Setelah diperoleh hasil prediksi klasifikasi, langkah selanjutnya adalah mengukur ketepatan hasil prediksi dengan hasil observasi. Menurut IRIS (2005), ukuran ketepatan klasifikasi yang dapat digunakan yaitu:

1. Akurasi yaitu persentase ketepatan klasifikasi yang dilakukan oleh suatu algoritma/fungsi klasifikasi secara keseluruhan.

$$\text{Akurasi} = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\% \quad (2.1)$$

2. Sensitifitas yaitu persentase amatan yang tepat diprediksikan sebagai kejadian gagal ($y=0$) dari semua observasi kejadian gagal

$$\text{Sensitifitas} = \frac{n_{11}}{n_{11} + n_{12}} \times 100\% \quad (2.2)$$

3. *Apparent Error Rate* (APER) adalah persentase kesalahan klasifikasi yang dilakukan oleh suatu algoritma/fungsi klasifikasi.

$$\text{APER} = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\% \quad (2.3)$$

2.3. Regresi Nonparametrik

Analisis regresi merupakan salah satu metode statistika yang umum digunakan untuk melihat hubungan fungsional antara peubah respon yang dipengaruhi oleh peubah prediktor. Menurut Eubank (1988), regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon dan prediktor yang tidak diketahui bentuk fungsinya, hanya diasumsikan fungsi *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu, sehingga regresi nonparametrik memiliki fleksibilitas yang tinggi. Karena bentuk kurva regresi tidak diketahui, maka regresi nonparametrik dapat membentuk fungsi apapun dalam menduga bentuk kurva regresi baik linier atau nonlinier.

Menurut Eubank (1988), model regresi nonparametrik secara umum dapat disajikan sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i, \text{ untuk } i = 1, 2, \dots, n \quad (2.4)$$

di mana:

- y_i : nilai peubah respon pada pengamatan ke- i
- $f(x_i)$: fungsi regresi tanpa bentuk
- x_i : nilai peubah prediktor pada pengamatan ke- i
- ε_i : sisaan ke- i
- n : banyak pengamatan

2.4. *Multivariate Adaptive Regression Spline (MARS)*

MARS dikembangkan oleh Friedman (1991). MARS merupakan pendekatan model regresi multivariat 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.

Menurut Friedman (1991), model MARS berguna untuk mengatasi masalah data berdimensi tinggi, yaitu data sampel yang berukuran $50 \leq N \leq 1000$ di mana N merupakan ukuran sampel dan memiliki peubah prediktor $3 \leq n \leq 20$ di mana n merupakan banyaknya peubah. Model MARS juga berguna untuk mengatasi masalah ketidakkontinuan pada data.

Beberapa hal yang perlu diperhatikan dalam membangun model MARS adalah:

1. Knot.

Nash dan Bradford (2001) menyatakan bahwa knot adalah sebuah titik di mana kemiringan dari sebuah garis berubah. Sebuah knot menandai akhir dari satu wilayah data dan awal dari wilayah data yang lain. Disetiap knot, diharapkan adanya kontinuitas dari fungsi basis antar bagian.

Pemilihan jumlah dan penempatan knot pada MARS dilakukan melalui prosedur pemilihan terlebih dahulu berdasarkan *straight forward search problem*, yaitu dengan menguji sejumlah knot dan memilih satu atau sekumpulan knot yang memberikan R^2 terbesar (Steinberg dkk., 2001). MARS dalam menemukan lokasi dan banyaknya knot yang dibutuhkan melalui prosedur *forward and backward stepwise*. Mula-mula membangun model dengan sangat banyak knot yang menyebabkan model MARS *overfitting*, kemudian





knot yang berkontribusi paling sedikit terhadap model menyeluruh dipindahkan. Pemilihan knot pada prosedur *forward* akan memasukan banyak lokasi knot yang tidak tepat, tetapi knot yang tidak tepat ini pada akhirnya akan dihapus dari model melalui prosedur pemangkasan *backward* (Abraham dan Steinberg, 2002).

MARS mencari lokasi knot dengan menentukan minimal amatan antar knot. Cara untuk menghasilkan model MARS yang *adaptive* adalah dengan menggunakan minimum amatan antar knot yang relatif kecil dengan tetap mempertimbangkan ukuran contoh (Stainberg dkk., 2001). Moisen dan Frescino (2002) menetapkan banyaknya amatan minimal antar knot adalah satu sehingga MARS akan mempertimbangkan setiap titik pada ruang peubah prediktor sebagai kandidat lokasi knot.

2. Fungsi Basis (BF, *Base Function*)

Sekumpulan fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor yang dipisahkan oleh titik-titik knot. Fungsi basis pada dasarnya menyatakan kembali hubungan antara prediktor dengan respon (Steinberg dkk., 2001).

Pada pemodelan MARS pemilihan fungsi basis dapat dijelaskan dalam metode *forward and backward stepwise*. Prosedur *forward stepwise* dimulai dengan hanya sebuah konstanta dalam model dan kemudian menambahkan fungsi basis ke dalam model. Pada prosedur ini fungsi ditambahkan sampai banyak fungsi basis maksimal yang diperbolehkan telah tercapai. Kebaikan model pada prosedur *forward* diukur dengan *Mean Square Error* (MSE). Penambahan fungsi basis akan selalu mengurangi nilai MSE (Steinberg dkk., 2001). Mina (2008) menyatakan bahwa setelah membentuk model dengan sangat banyak fungsi basis, prosedur *backward stepwise* diterapkan yaitu dengan memindahkan fungsi basis yang berkontribusi terkecil terhadap model. Pada tahap ini, peubah prediktor dapat dikeluarkan dari model jika tidak ada satupun fungsi basisnya yang berkontribusi signifikan terhadap kinerja prediksi model. Pemilihan fungsi basis pada prosedur *backward stepwise* berdasarkan nilai *Generalized Cross Validation* (GCV).

Friedman (1991), menyarankan banyaknya maksimum fungsi basis (BF) adalah 2 sampai 4 kali banyaknya variabel prediktor yang ada. Banyaknya maksimum interaksi (MI) yang digunakan sebanyak

0 sampai 3 interaksi. Interaksi merupakan hasil kali antar peubah yang saling berkorelasi. Jika jumlah interaksi lebih dari 3 akan dihasilkan model yang kompleks dan model akan sulit untuk diinterpretasi. Minimum observasi (MO) antara knot yang digunakan sebesar 0, 1, 2 dan 3.

Menurut Hastie dkk. (2009), MARS dapat dinyatakan dalam bentuk perluasan fungsi basis linier tersegmen:

$$T_1(x, t) = (x - t)_+ = \begin{cases} x - t, & \text{jika } x > t \\ 0, & \text{selainnya} \end{cases} \quad (2.5)$$

dan

$$T_2(x, t) = (t - x)_+ = \begin{cases} t - x, & \text{jika } x < t \\ 0, & \text{selainnya} \end{cases} \quad (2.6)$$

di mana: x : peubah prediktor
 t : titik knot pada peubah x

Persamaan (2.5) dan (2.6) disebut *reflected pair* di mana indeks positif menunjukkan bahwa hanya nilai positif saja yang digunakan, jika tidak maka ditetapkan nilai nol.

Model MARS merupakan modifikasi model *Recursive Partitioning Regression* (RPR) dengan kombinasi *spline* yang digunakan untuk mengatasi kelemahan RPR sehingga menghasilkan model yang kontinu pada knot dan dapat mengidentifikasi adanya fungsi linear dan aditif. Kontinuitas diperlukan untuk mendapatkan fungsi basis yang baik yaitu memberikan keakuratan yang tinggi.

Perbaikan yang ada dalam model MARS untuk mengatasi keterbatasan RPR yaitu menghasilkan fungsi basis sebagai berikut:

$$B_m^{(q)}(x) = \prod_{k=1}^{K_m} [s_{km} \cdot (x_{i(km)} - t_{km})]_+^q \quad (2.7)$$

di mana: $B_m^{(q)}(x)$: basis fungsi ke- m yang merupakan fungsi dari x
 K_m : maksimum banyaknya derajat interaksi
 s_{km} : konstanta bernilai +1 bila data disebelah kanan titik knot dan bernilai -1 bila data disebelah kiri titik knot
 $x_{i(k,m)}$: peubah prediktor ke- i
 t_{km} : nilai knot dari peubah prediktor $x_{i(k,m)}$

Persamaan (2.7) disebut *truncated power spline basis function* (Friedman, 1991).



2.4.1. Dekomposisi Analisis Ragam Model MARS

Menurut Friedman (1991), model umum MARS dapat ditulis:

$$f(x_i) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km}(x_{i(k,m)} - t_{km})]_+ + \varepsilon_i \quad (2.8)$$

$m = 1, 2, \dots, M$ dan $k = 1, 2, \dots, K_m$

di mana: a_0 : konstanta regresi (*intersep*)
 a_m : koefisien dari fungsi basis ke- m
 M : maksimum banyaknya fungsi basis
 ε_i : sisaan ke- i

Hasil penjabaran dari persamaan (2.8) adalah:

$$f(x_i) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km}(x_{i(k,m)} - t_{km})]$$

$$f(x_i) = a_0 + \sum_{m=1}^M a_m [s_{1m}(x_{i(1,m)} - t_{1m})] +$$

$$\sum_{m=1}^M a_m [s_{1m}(x_{i(1,m)} - t_{1m})][s_{2m}(x_{i(2,m)} - t_{2m})] +$$

$$\sum_{m=1}^M a_m [s_{1m}(x_{i(1,m)} - t_{1m})][s_{2m}(x_{i(2,m)} - t_{2m})][s_{3m}(x_{i(3,m)} - t_{3m})] + \dots + \varepsilon_i \quad (2.9)$$

Persamaan (2.9) secara umum dapat ditulis sebagai berikut:

$$f(x_i) = a_0 + \sum_{m=1}^M f_i(x_i) + \sum_{m=1}^M f_{ij}(x_i, x_j) + \sum_{m=1}^M f_{ijk}(x_i, x_j, x_k) + \dots + \varepsilon_i \quad (2.10)$$

Persamaan (2.10) menunjukkan bahwa penjumlahan pertama saat $k=1$ meliputi semua fungsi basis untuk satu peubah. Penjumlahan kedua saat $k=2$ meliputi semua fungsi basis untuk interaksi antara 2 peubah, hingga penjumlahan ke- M dimana $k=K_m$ meliputi semua fungsi basis untuk interaksi sebanyak K_m peubah.



Persamaan (2.10) merupakan dekomposisi dari analisis ragam pada model MARS. Interpretasi model MARS melalui dekomposisi analisis ragam adalah merepresentasikan peubah prediktor yang masuk ke dalam model termasuk interaksi antar peubah prediktor serta level interaksinya. Interpretasi lebih lanjut dapat dilakukan dengan merepresentasikan dekomposisi analisis ragam secara grafis (Friedman, 1991).

Berdasarkan persamaan (2.8) model MARS dapat ditulis sebagai berikut:

$$y_i = f(x_i) = a_0 + \sum_{m=1}^M a_m B_m(x_i) + \varepsilon_i, \quad (2.11)$$

di mana $m = 1, 2, \dots, M$

dengan

$$B_m(x_i) = \prod_{k=1}^{K_m} [s_{km}(x_{i(k,m)} - t_{km})]_+, \quad (2.12)$$

di mana $k = 1, 2, \dots, K_m$

Persamaan (2.11) dapat dituliskan dalam bentuk matriks:

$$y = B\alpha + \varepsilon \quad (2.13)$$

di mana:

$$y_{n \times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \varepsilon_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \alpha_{(M+1) \times 1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{bmatrix}$$

$$B_{n \times (M+1)} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{1(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [s_{kM}(x_{1(k,M)} - t_{kM})]_+ \\ 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{2(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [s_{kM}(x_{2(k,M)} - t_{kM})]_+ \\ \vdots & \vdots & \dots & \vdots \\ 1 & \prod_{k=1}^{K_1} [s_{k1}(x_{n(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [s_{kM}(x_{n(k,M)} - t_{kM})]_+ \end{bmatrix}$$



2.4.2. Multikolinieritas dalam Model MARS

MARS membangun fungsi basis melalui seleksi *forward*, kemudian seleksi *backward* untuk mendapatkan model yang lebih sederhana. Jika kedua peubah prediktor berkorelasi, prosedur *forward* pada MARS akan menempatkan knot pada salah satu dari dua peubah tersebut, sehingga jika berdasarkan GCV dan *penalized residual sum of square* akan menghasilkan hasil yang kurang akurat (De Veaux dan Ungar, 1994).

Untuk mengetahui adanya multikolinieritas dapat dilihat dari nilai *Variance Inflation Factor* (VIF) yang dihitung berdasarkan nilai koefisien determinasi *auxiliary regression*. *Auxiliary Regression* merupakan bentuk regresi yang digunakan untuk mengetahui apakah terdapat korelasi antara peubah-peubah prediktor yang digunakan dengan cara membuat model regresi baru dengan menggunakan salah satu peubah prediktor sebagai peubah respon. Apabila VIF bernilai lebih dari atau sama dengan 10 maka terdapat indikasi adanya multikolinieritas. Menurut Li (2000), nilai *Variance Inflation Factor* (VIF) dihitung dengan rumus:

$$VIF_j = \frac{1}{1 - R_j^2}, \text{ untuk } j = 1, 2, \dots, p \quad (2.14)$$

di mana:

- R_j^2 : koefisien determinasi *auxiliary regression* dengan peubah prediktor x_j sebagai peubah respon untuk $j = 1, 2, \dots, p$
- x_j : Peubah prediktor ke- j untuk $j = 1, 2, \dots, p$
- p : banyaknya peubah prediktor

Friedman (1991) mengusulkan penanganan dalam mengatasi multikolinieritas yaitu dengan meningkatkan orde interaksi dalam model, kemudian membandingkan nilai GCV yang paling minimum, selanjutnya memilih model dengan interaksi terendah. Penanganan yang lain adalah dengan melibatkan peubah prediktor tertentu pada model sehingga mengurangi kolinieritas saat memasukkan peubah prediktor. Namun menurut De Veaux dan Ungar (1994) tak satupun strategi tersebut mengatasi permasalahan di antara dua peubah yang berkorelasi, tetapi hanya menghalangi agar salah satu tidak menghasilkan nilai optimum.

Tindakan untuk perbaikan multikolinieritas pada MARS dapat dilakukan seperti halnya multikolinieritas pada regresi berganda, yaitu mengeluarkan satu peubah prediktor dari penelitian, penambahan data



baru, melakukan transformasi data, melakukan analisis komponen utama atau lain sebagainya.

2.4.3. Pendugaan Parameter Model MARS

Menurut Kriner (2007), pendugaan parameter model MARS dengan peubah respon kontinu dilakukan dengan metode kuadrat terkecil, namun apa bila peubah respon merupakan peubah kategorik biner pendugaan parameter model MARS dapat dilakukan dengan metode *maximum log likelihood*. Pada MARS dengan peubah respon biner, antar pengamatan saling bebas dan nilai peluang peubah responnya $P(Y_i = 1) = \pi$ dan $P(Y_i = 0) = 1 - \pi$, maka fungsi kemungkinan yang akan dimaksimalkan adalah sebagai berikut:

$$L(\alpha) = \prod_{i=1}^N \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} = \prod_{i=1}^N \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right]^{y_i} [1 - \pi(x_i)]$$

$$L(\alpha) = \prod_{i=1}^N \left[\exp \left(\sum_{m=1}^M a_m B_m \right) \right]^{y_i} \left[1 + \exp \left(\sum_{m=1}^M a_m B_m \right) \right]^{-1} \quad (2.15)$$

Agar persamaan (2.15) dapat dimaksimumkan, maka kedua ruas harus dilogartimkan, sehingga diperoleh:

$$l(\alpha) = \ln L(\alpha)$$

$$\ln L(\alpha) = \ln \prod_{i=1}^N \left[\exp \left(\sum_{m=1}^M a_m B_m \right) \right]^{y_i} \left[1 + \exp \left(\sum_{m=1}^M a_m B_m \right) \right]^{-1}$$

$$l(\alpha) = \ln \left[\exp \left(\sum_{m=1}^M a_m B_m \right) \right]^{\sum_{i=1}^N y_i} + \ln \left[1 + \exp \left(\sum_{m=1}^M a_m B_m \right) \right]^{-N}$$

$$l(\alpha) = \sum_{i=1}^N y_i \cdot \ln \left[\exp \left(\sum_{m=1}^M a_m B_m \right) \right] - N \cdot \ln \left[1 + \exp \left(\sum_{m=1}^M a_m B_m \right) \right]$$

$$l(\alpha) = \sum_{i=1}^N y_i \cdot \left(\sum_{m=1}^M a_m B_m \right) - N \cdot \ln \left[1 + \exp \left(\sum_{m=1}^M a_m B_m \right) \right] \quad (2.16)$$

Persamaan (2.16) agar mencapai maksimal maka turunan parsial terhadap parameter α_m di mana $m=1,2,\dots,M$ harus disamakan dengan nol

$$\frac{\partial l(\alpha)}{\partial l(\alpha_1)} = \frac{\partial l(\alpha)}{\partial l(\alpha_2)} = \frac{\partial l(\alpha)}{\partial l(\alpha_3)} = \dots = \frac{\partial l(\alpha)}{\partial l(\alpha_M)} = 0 \quad (2.17)$$

Dalam menyelesaikan turunan parsial pada persamaan (2.17) diperlukan metode iterasi *Newton-Raphson* di dalam perhitungannya karena $\hat{\alpha}_M$ bersifat nonlinier.

2.4.4. Pengujian Parameter Model MARS

Pengujian terhadap parameter model bertujuan untuk memutuskan secara statistika apakah model yang telah terbentuk layak sebagai penduga hubungan sebenarnya antara peubah prediktor dengan peubah respon dengan melihat parameter model yang dihasilkan. Pengujian signifikansi parameter model MARS dapat dilakukan secara parsial maupun secara simultan.

a. Pengujian secara simultan

Untuk menguji signifikansi penduga parameter model MARS secara simultan digunakan uji-*F* melalui tabel analisis ragam dengan hipotesis:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

$$H_1: \text{minimal terdapat satu } \alpha_m \neq 0, \text{ di mana } m = 1,2,\dots,M$$

Statistik uji-*F* dapat dihitung dengan rumus:

$$F_{\text{hitung}} = \frac{KT_{\text{Regresi}}}{KT_{\text{Sisa}}} \sim F_{(M,n-M-1)} \quad (2.18)$$

Hipotesis nol ditolak apabila nilai statistik uji $F_{\text{hitung}} > F_{(M,n-M-1)}^{\alpha/2}$ artinya minimal terdapat satu $\alpha_m \neq 0$. Nilai KT_{Regresi} dan KT_{Sisa} didapatkan dari perhitungan tabel analisis ragam yang terdapat pada tabel 2.2.



Tabel 2.2. Tabel Analisis Ragam

Sumber Keragaman	Derajat Bebas	Jumlah Kuadrat (JK)	Kuadrat Tengah (KT)	F_{hitung}
Regresi	M	$\hat{\alpha}^t B^t y$	$\frac{\hat{\alpha}^t B^t y}{M}$	$\frac{\hat{\alpha}^t B^t y}{M}$
Sisa	$n-M-1$	$y^t y - \hat{\alpha}^t B^t y$	$\frac{y^t y - \hat{\alpha}^t B^t y}{n-M-1}$	$\frac{y^t y - \hat{\alpha}^t B^t y}{n-M-1}$
Total	$n-1$	$y^t y$		

di mana:

- $\hat{\alpha}$: vektor nilai pendugaan parameter
- B : Matriks fungsi basis
- y : vektor respon
- M : banyak fungsi basis
- n : banyak amatan

b. Pengujian secara parsial

Untuk menguji signifikansi penduga parameter model MARS secara parsial digunakan uji-t dengan hipotesis:

$$H_0: \alpha_m = 0$$

$$H_1: \alpha_m \neq 0, \text{ di mana } m = 1, 2, \dots, M$$

Statistik uji-t dapat dihitung dengan rumus:

$$t_{hitung} = \frac{\hat{\alpha}_m}{Se(\hat{\alpha}_m)} \sim t_{(n-M)} \tag{2.19}$$

di mana:

$\hat{\alpha}_m$: penduga parameter dari fungsi basis ke- m

$Se(\hat{\alpha}_m)$: Salah baku dari $\hat{\alpha}_m = \sqrt{var(\hat{\alpha}_m)}$

$var(\hat{\alpha}_m)$: diagonal utama matriks $[B^t V^{-1} B]^{-1}$

B : Matriks fungsi basis

V : Matriks dengan diagonal utama $[\pi_k(1 - \pi_k)]$



$$\text{Var}(\hat{\alpha}_m)_{(M+1) \times (M+1)}$$

$$= \begin{bmatrix} \text{var}(\hat{\alpha}_0) & \text{covar}(\hat{\alpha}_1, \hat{\alpha}_0) & \dots & \text{covar}(\hat{\alpha}_m, \hat{\alpha}_0) \\ \text{covar}(\hat{\alpha}_0, \hat{\alpha}_1) & \text{var}(\hat{\alpha}_1) & \dots & \text{covar}(\hat{\alpha}_m, \hat{\alpha}_1) \\ \vdots & \vdots & \ddots & \vdots \\ \text{covar}(\hat{\alpha}_0, \hat{\alpha}_m) & \text{covar}(\hat{\alpha}_1, \hat{\alpha}_m) & \dots & \text{var}(\hat{\alpha}_m) \end{bmatrix}$$

$$B_{n \times (M+1)}$$

$$= \begin{bmatrix} 1 & \prod_{k=1}^{K_1} [S_{k1}(x_{1(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [S_{kM}(x_{1(k,M)} - t_{kM})]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} [S_{k1}(x_{2(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [S_{kM}(x_{2(k,M)} - t_{kM})]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} [S_{k1}(x_{n(k,1)} - t_{k1})]_+ & \dots & \prod_{k=1}^{K_M} [S_{kM}(x_{n(k,M)} - t_{kM})]_+ \end{bmatrix}$$

$$V_{n \times n}$$

$$= \begin{bmatrix} \pi_k(x_1)(1 - \pi_k(x_1)) & 0 & \dots & 0 \\ 0 & \pi_k(x_2)(1 - \pi_k(x_2)) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi_k(x_n)(1 - \pi_k(x_n)) \end{bmatrix}$$

Hipotesis nol ditolak apabila nilai statistik uji $|t_{hitung}| < |t_{(n-M)}^{\alpha/2}|$ artinya penduga parameter α_m signifikan tidak sama dengan nol dengan kata lain fungsi basis ke- m berpengaruh nyata terhadap peubah respon.

2.4.5. Pemilihan Model Terbaik MARS

Pemilihan model terbaik dilakukan dengan metode *stepwise*. *Foward stepwise* dilakukan untuk mendapat model dengan banyaknya fungsi basis maksimum. Kebaikan model pada prosedur *forward* diukur dengan MSE. *Backward stepwise* diterapkan dengan memindahkan fungsi basis yang berkontribusi terkecil terhadap model. Pada tahap ini, peubah prediktor dapat dikeluarkan dari model jika tidak ada satupun fungsi basisnya yang berkontribusi signifikan terhadap kinerja prediksi model. Pemilihan fungsi basis pada prosedur *backward stepwise* berdasarkan nilai *Generalized Cross Validation* (GCV) yang didefinisikan sebagai berikut:



$$\text{LOF}(\hat{f}_M) = \text{GCV}(M) = \frac{N^{-1} \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{\hat{C}(M)}{N}\right]^2} \quad (2.20)$$

di mana:

- y_i : peubah respon pada pengamatan ke- i
- $\hat{f}_M(x_i)$: penduga peubah respon dengan fungsi basis M
- N : banyak pengamatan
- $\hat{C}(M)$: $C(M) + dM$, nilai d yang baik berada dalam interval $2 \leq d \leq 4$
- $C(M)$: $\text{trace}[\mathbf{B}(\mathbf{B}^t \mathbf{B})^{-1} \mathbf{B}^t] + 1$ adalah banyak parameter yang diduga
- M : banyak fungsi basis

Satu dari rangkaian model yang bervariasi dipilih sebagai model yang terbaik dengan GCV yang minimal. Alasan penggunaan GCV adalah karena efisiensinya dalam hal komputasi (Breiman, 1991).

2.4.6. Klasifikasi dengan Model MARS

Terdapat dua kemungkinan klasifikasi dengan model MARS. Kemungkinan pertama yaitu berhubungan dengan pasangan klasifikasi yang memiliki *output* biner dikodekan sebagai 0 dan 1. Kemungkinan kedua meliputi kalsifikasi dengan *output* tidak biner atau memiliki lebih dari 2 kelas (Hastie dkk., 2009). Klasifikasi yang memiliki *output* biner didasarkan pada pendekatan analisis regresi logistik dengan menggunakan model peluang dengan persamaan sebagai berikut:

$$\pi = \frac{e^{f(x_i)}}{1 + e^{f(x_i)}} \text{ dan } (1 - \pi) = \frac{1}{1 + e^{f(x_i)}} \quad (2.21)$$

- di mana: π : peluang kejadian sukses
- $(1 - \pi)$: peluang kejadian gagal
- $f(x_i)$: model regresi

Dalam penelitian ini, Y adalah peubah respon biner dengan m peubah prediktor. Model MARS dalam mengklasifikasikan 2 kelas diperlukan adanya transformasi logit yang dituliskan sebagai berikut:

$$\text{logit } \pi(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \hat{f}(x) \quad (2.22)$$



$\hat{f}(x)$ pada persamaan (2.22) merupakan model MARS, sehingga model MARS untuk klasifikasi adalah:

$$\ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km}(x_{i(k,m)} - t_{km})]_+ \quad (2.23)$$

Apabila peubah respon memiliki dua kategori maka dapat digunakan titik potong sebesar 0,5 dengan ketentuan jika $\pi(x) \geq 0,5$ maka hasil prediksi kelas adalah 1, apabila $\pi(x) < 0,5$ maka hasil prediksi kelas adalah 0.

2.5. Bootstrap Agregating (Bagging)

Bagging diperkenalkan oleh Breiman (1996) yang merupakan suatu metode untuk memperbaiki kekuatan prediksi dari beberapa penduga atau algoritma tertentu seperti regresi atau pohon klasifikasi. Secara umum *Bagging* banyak digunakan untuk model yang tidak stabil seperti *Decision Tree* dan *Artificial Neural Network*. Yang dimaksud tidak stabil ialah apabila terdapat sedikit perubahan pada data latih akan berpengaruh besar pada model pembelajaran yang dihasilkan. *Bagging* sangat baik digunakan pada data berdimensi tinggi di mana sering terdapat kesulitan dalam melakukan pemodelan klasifikasi ataupun regresi. Pada penelitian ini *Bagging* digunakan untuk memperbaiki kekuatan klasifikasi pada model MARS dengan cara melakukan resampling pada data.

Bagging bekerja dengan cara penggabungan (*aggregating*) sejumlah r model yang dibentuk menggunakan sejumlah r data latih yang dibangkitkan secara acak menggunakan metode *Bootstrap* agar masing-masing model saling independen (Suyanto, 2018). Menurut Buhlmann (2003), prinsip kerja *Bagging* adalah dengan cara mereduksi ragam peubah prediktor. Proses ini disebut sebagai *Bagging* prediktor yaitu merupakan metode untuk membangkitkan peubah prediktor yang dibentuk dengan membuat pengulangan *Bootstrap* terhadap kumpulan data sehingga didapatkan kumpulan data baru. Metode *Bootstrap* atau bisa disebut *random with replacement* adalah pembangkitan data menggunakan *sampling* acak dengan distribusi seragam, dimana subhimpunan pada data yang sudah terpilih secara acak bisa dikembalikan lagi ke sumber data sehingga dapat terpilih kembali.

Algoritma *Bagging* adalah sebagai berikut:

1. Sebuah data L terdiri dari $\{(y_{i,x_i}), i=1,2,\dots,n\}$ dilakukan pengambilan sampel berukuran n dengan pengembalian sehingga didapatkan data baru $L^* = \{(y_i^*, x_i^*), i=1,2,\dots,n\}$.
2. Melakukan algoritma klasifikasi atau regresi pada data baru L^* .
3. Mengulangi langkah 1 dan 2 sampai B kali penggandaan *Bootstrap*.
4. Diperoleh ketepatan klasifikasi dan prediksi pada setiap penggandaan *Bootstrap*.
5. Hasil prediksi peubah respon merupakan rata-rata dari prediksi yang dihasilkan setiap penggandaan *Bootstrap* apabila peubah respon bersifat kontinu dan merupakan hasil dari modus (*majority voting*) apabila peubah respon bersifat kategorik.

2.6. *Synthetic Minority Over-sampling Technique (SMOTE)*

Ketidakeimbangan data terjadi jika jumlah objek suatu kelas data lebih banyak dibandingkan dengan kelas lain. Kelas data yang objeknya lebih banyak disebut kelas mayoritas sedangkan lainnya disebut kelas minoritas. Pengaruh penggunaan data tidak seimbang untuk membuat model sangat besar pada hasil model yang diperoleh. Pengolahan algoritma yang tidak menghiraukan ketidakeimbangan data akan cenderung diliputi oleh kelas mayoritas dan mengacuhkan kelas minoritas (Chawla dkk., 2002). Salah satu metode dalam mengatasi permasalahan data tidak seimbang adalah dengan SMOTE.

SMOTE diperkenalkan oleh Chawla dkk. (2002) sebagai salah satu solusi dalam menangani data tidak seimbang. SMOTE merupakan salah satu turunan dari metode *oversampling* yang berprinsip memperbanyak pengamatan secara acak. Metode SMOTE menambah banyaknya pengamatan pada data (replikasi) kelas minoritas sehingga proporsi kelas data mayoritas dan minoritas menjadi lebih seimbang dengan cara membangkitkan data buatan (*synthetic data*) berdasarkan kedekatan ruang antar data pada kelas minoritas.

SMOTE merupakan salah satu metode resampling dengan memperbanyak sampel dengan perhitungan KNN. Data buatan pada metode SMOTE dibuat berdasarkan ketetanggaan terdekat data sebanyak k (*k-nearest neighbors*) untuk setiap data di kelas minoritas, setelah itu dibuat data sintesis sebanyak persentase duplikasi yang diinginkan. Prosedur pembangkitan data buatan dengan metode SMOTE adalah dengan persamaan sebagai berikut:



$$x_{baru} = x_i + (\hat{x}_i - x_i) \times rand[0,1] \quad (2.24)$$

di mana x_i adalah nilai peubah prediktor, $rand[0,1]$ adalah nilai acak yang bernilai antara 0 dan 1 serta \hat{x}_i adalah salah satu K -Nearest Neighbors (KNN). KNN digunakan untuk membuat data sintesis pada kelas minoritas. Dalam pembangkitan data buatan berdasarkan KNN dihitung jarak kedekatannya menggunakan metode jarak *Euclidean* dengan persamaan sebagai berikut:

$$d(x, y) = \sqrt{(x - y)^t(x - y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.25)$$

di mana:
 x : vektor peubah x
 y : vektor peubah y

Menurut Adiangga (2015), teknik SMOTE tidak hanya meningkatkan jumlah kelas data minoritas (*oversampling*) tetapi juga menurunkan jumlah kelas data mayoritas (*undersampling*) untuk menjaga keseimbangan kelas pada data.

Algoritma pembangkitan data buatan dengan metode SMOTE adalah sebagai berikut:

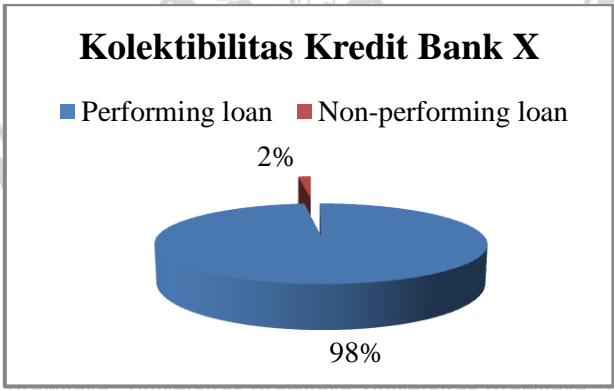
1. Menentukan nilai k pada KNN yang berarti setiap data akan dihitung KNN dengan mempertimbangkan k tetangga terdekat.
2. Menghitung jarak antar pengamatan pada kelas minoritas dengan metode jarak Euclidean pada persamaan (2.25).
3. Menghitung nilai data buatan pada persamaan (2.24).
4. Menentukan nilai *oversampling* dalam ratus persen. Nilai *oversampling* 400% berarti setiap data pada kelas minoritas akan dibangkitkan sebanyak 4 data buatan.
5. Menentukan nilai *undersampling* dalam ratus persen. Nilai *undersampling* 200% berarti sebanyak 2 kali jumlah data minoritas akan digunakan untuk mengambil contoh dari kelas mayoritas.
6. Data baru akan terdiri dari gabungan seluruh data minoritas ditambah hasil SMOTE dan contoh dari kelas mayoritas.

BAB III METODE PENELITIAN

3.1. Sumber Data

Untuk memahami pengaruh jumlah ukuran sampel, perbedaan banyaknya pengamatan pada peubah respon berskala nominal dengan dua kategori dan ada tidaknya pola distribusi tertentu terhadap akurasi yang dihasilkan metode klasifikasi maka perlu dilakukan simulasi, sehingga data yang digunakan merupakan data bangkitan. Data simulasi dibangkitkan pada satu model yaitu model yang terdiri dari satu peubah respon dan enam peubah prediktor.

Membangkitkan satu peubah respon berskala nominal biner yang berdistribusi bernauli menggunakan *Software R*. Parameter model diperoleh berdasarkan pemodelan regresi logistik dari data yang ada pada skripsi Naufal (2019) tentang *Ensemble Adaboost* pada metode CART dalam mengatasi *class imbalance* pada data kolektibilitas kredit bank. Banyaknya pengamatan yang ada pada data tersebut sebanyak 9737 pengamatan dengan peubah kolektibilitas kredit sebagai peubah respon, yaitu *performing loan* (label 0) sebanyak 9565 pengamatan dan *non-performing loan* (label 1) sebanyak 172 pengamatan.



Gambar 3.1. Persentase tiap kelas pada peubah respon kolektibilitas kredit data asli

Dalam penelitian ini diambil satu peubah respon yaitu kolektibilitas kredit dengan dua kategori, yaitu *performing loan* (label 0) dan *non-performing loan* (label 1) dan diambil enam peubah

prediktor yaitu jangka waktu kredit (dalam bulan), angsuran kredit (dalam juta rupiah), suku bunga (dalam persen), umur kredit (dalam tahun), umur debitur (dalam tahun) dan jenis kelamin untuk dilakukan analisis dengan menggunakan model regresi logistik. Dari hasil analisis menggunakan model regresi logistik akan didapatkan penduga parameter distribusi regresi logistik yang digunakan sebagai parameter dalam membangkitkan data simulasi.

Hasil pendugaan parameter model regresi logistik pada peubah respon kolektibilitas kredit dengan enam peubah prediktor (jangka waktu kredit, angsuran kredit, suku bunga, umur kredit, umur debitur dan jenis kelamin) kemudian akan dibulatkan hingga 2 angka di belakang koma. Hasil pendugaan parameter ini yang nantinya digunakan untuk membangkitkan peubah respon dari enam peubah prediktor diatas yang diambil secara acak dengan ukuran contoh ($n = 100, 200, 500$ dan 1000) dan proporsi kelas minoritas sebesar ($p = 10\%, 15\%$ dan 20%).

3.2. Proses Pembangkitan Data

Langkah-langkah dalam membangkitkan data dengan satu peubah respon yang berskala nominal dengan dua kategori adalah sebagai berikut:

1. Melakukan pendugaan parameter model regresi logistik pada data sehingga mendapatkan hasil pendugaan parameter yang dapat digunakan sebagai koefisien awal dalam metode simulasi
2. Menentukan koefisien awal model regresi logistik biner dengan cara pembulatan 2 angka di belakang koma didapatkan hasil: $\beta_0 = 4,01$; $\beta_1 = 0,01$; $\beta_2 = -0,06$; $\beta_3 = -5,91$; $\beta_4 = -0,26$; $\beta_5 = 0,02$ dan $\beta_6 = 0,37$ yang digunakan untuk membentuk vektor β seperti berikut:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix} = \begin{bmatrix} 4,01 \\ 0,01 \\ -0,06 \\ -5,91 \\ -0,26 \\ 0,02 \\ 0,37 \end{bmatrix}$$

3. Membuat matriks X (peubah prediktor) dengan kolom pertama bernilai 1 dan kolom lain merupakan nilai dari enam peubah prediktor pada data asli yang dituliskan seperti berikut:



$$X_{n \times (6+1)} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{16} \\ 1 & x_{21} & x_{22} & \dots & x_{26} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{n6} \end{bmatrix}$$

4. Membangkitkan peubah respon yang dibangkitkan dari distribusi *Bernaulli* yang menghasilkan dua kemungkinan, yaitu sukses atau gagal dengan fungsi kepadatan peluang seperti berikut:

$$p(y) = \begin{cases} \pi^y(1 - \pi)^{1-y}, & y = 0,1 \\ 0, & y = \text{lainnya} \end{cases}$$

di mana π = peluang kejadian sukses

Parameter distribusi Bernaulli ditunjukkan dengan persamaan (3.1) yang digunakan untuk membangkitkan peubah respon sebanyak jumlah pengamatan yang ada pada data asli.

$$\pi = \frac{\exp(X\beta)}{1 + \exp(X\beta)} = \frac{\exp(\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_6x_6)}{1 + \exp(\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_6x_6)} \quad (3.1)$$

5. Membentuk data *frame* yang terdiri dari peubah prediktor dari data awal dan peubah respon baru dari hasil bangkitan pada langkah 4.

6. Mengambil contoh pengamatan dari data *frame* yang dihasilkan pada langkah 5 sebanyak ukuran contoh ($n = 100, 200, 500$ dan 1000) dengan cara:

- a. Membagi data menjadi 2 bagian yaitu data dengan peubah respon kelas minoritas (label 0) dan data dengan peubah respon kelas mayoritas (label 1).
- b. Mengambil contoh acak pada data dengan peubah respon kelas minoritas (label 0) sebanyak tingkat proporsi yang digunakan ($p = 10\%, 15\%$ dan 20%) kali ukuran contoh ($n = 100, 200, 500$ dan 1000).
- c. Mengambil contoh acak pada data dengan peubah respon kelas mayoritas (label 1) sebanyak tingkat proporsi yang digunakan ($1-p = 95\%, 90\%$ dan 85%) kali ukuran contoh ($n = 100, 200, 500$ dan 1000).
- d. Menggabungkan contoh acak yang dihasilkan poin b dan c yang nantinya digunakan sebagai data simulasi.

3.3. Metode Analisis Data

Pada penelitian ini menggunakan tiga pendekatan metode yang terdiri dari:

A. Metode MARS

Langkah-langkah dalam analisis klasifikasi dengan metode MARS adalah sebagai berikut:

- 1) Membagi data simulasi menjadi 2, yaitu data latih dan data uji. Data dibagi secara acak dengan perbandingan data latih banding data uji sebesar 80%:20% atas dasar prinsip pareto.
- 2) Data latih digunakan untuk membentuk model klasifikasi MARS dengan mengkombinasikan nilai fungsi basis (BF) 3 kali banyaknya peubah prediktor, dengan maksimum interaksi (MI) 2 interaksi dan banyaknya observasi maksimum (MO) 1 observasi.
- 3) Menetapkan model MARS terbaik berdasarkan nilai GCV terkecil.
- 4) Memprediksi menggunakan model MARS terbaik pada data latih dan data uji.
- 5) Membentuk tabel dua arah antara nilai prediksi klasifikasi dengan nilai klasifikasi aktual.
- 6) Menghitung ketepatan klasifikasi metode MARS dengan menghitung nilai *Accuracy*, *Sensitivity* dan *APER*.

B. Metode *Bagging*-MARS

Langkah-langkah dalam analisis klasifikasi dengan metode *Bagging*-MARS adalah sebagai berikut:

- 1) Membagi data simulasi menjadi 2, yaitu data latih dan data uji. Data dibagi secara acak dengan perbandingan data latih banding data uji sebesar 80%:20% atas dasar prinsip pareto.
- 2) Melakukan *Bagging* dari pasangan peubah respon dan peubah prediktor pada data latih dengan tahapan:
 - a. Melakukan *Bootstrap* dari data latih, yaitu mengambil contoh acak sebanyak pengamatan pada data latih sesuai dengan sebaran seragam. Didapatkan data baru hasil *Bootstrap*.
 - b. Membentuk model klasifikasi MARS pada data hasil *Bootstrap* dengan BF, MI dan MO sama dengan model MARS terbaik pada data latih awal.



- c. Melakukan prediksi hasil klasifikasi model MARS pada data hasil *Bootstrap*.
 - d. Mengulangi langkah a hingga c sebanyak $B=25$ kali sehingga mendapatkan 25 hasil prediksi.
 - e. Hasil prediksi kelas merupakan *majority voting* dari 25 hasil prediksi data *Bootstrap* untuk ukuran sampel.
- 3) Membentuk tabel dua arah antara nilai prediksi klasifikasi dengan nilai klasifikasi aktual.
 - 4) Menghitung ketepatan klasifikasi metode *Bagging-MARS* dengan menghitung nilai *Accuracy*, *Sensitivity* dan *APER*.

C. Metode SMOTE-MARS

Langkah-langkah dalam analisis klasifikasi dengan metode SMOTE-MARS adalah sebagai berikut:

- 1) Membagi data simulasi menjadi 2, yaitu data latih dan data uji. Data dibagi secara acak dengan perbandingan data latih banding data uji sebesar 80%:20% atas dasar prinsip pareto.
- 2) Melakukan metode SMOTE pada data latih yaitu dengan membuat data buatan untuk kelas minoritas dengan tahapan:
 - a. Menentukan nilai k tetangga terdekat
 - b. Menghitung jarak antar k pengamatan pada kelas minoritas dengan metode jarak *Euclidean* persamaan (2.27) sehingga menghasilkan nilai KNN
 - c. Menghitung data buatan dengan persamaan (2.26)
 - d. Menentukan nilai *oversampling* dalam ratus persen. Nilai *oversampling* 900% menunjukkan bahwa setiap pengamatan pada kelas minoritas akan dibangkitkan data buatan sebanyak sembilan pengamatan
 - e. Menentukan nilai *undersampling* dalam ratus persen. Nilai *undersampling* 300% berarti kelas mayoritas akan diambil contoh seak sebanyak tiga kali jumlah data minoritas.
 - f. Data baru merupakan gabungan seluruh data kelas minoritas ditambah data buatan untuk kelas minoritas serta contoh acak dari kelas mayoritas.
- 3) Data hasil metode SMOTE digunakan untuk membentuk model klasifikasi MARS dengan BF, MI dan MO sama dengan model MARS terbaik pada data latih awal.
- 4) Memprediksi kelas menggunakan model MARS dengan SMOTE pada data latih dan data uji.



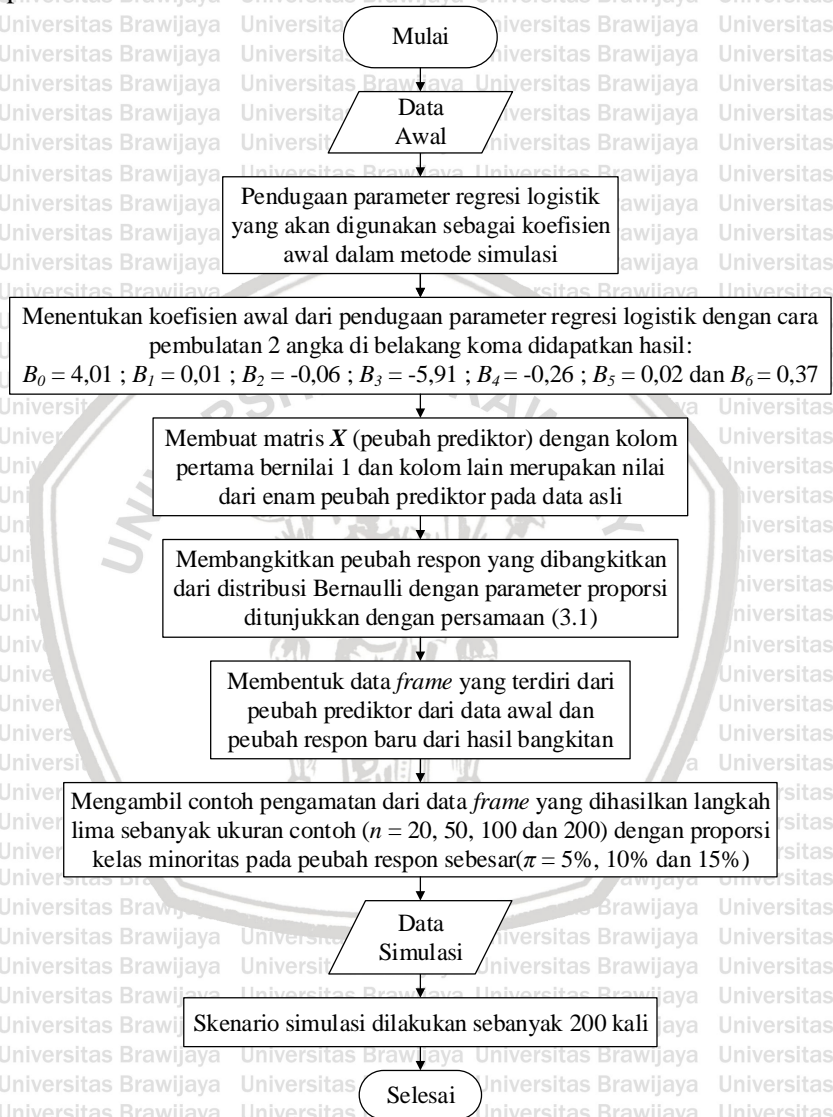


- 5) Membentuk tabel dua arah antara nilai prediksi klasifikasi dengan nilai klasifikasi aktual.
- 6) Menghitung ketepatan klasifikasi model MARS dengan menghitung nilai *Accuracy*, *Sensitivity* dan *APER*.



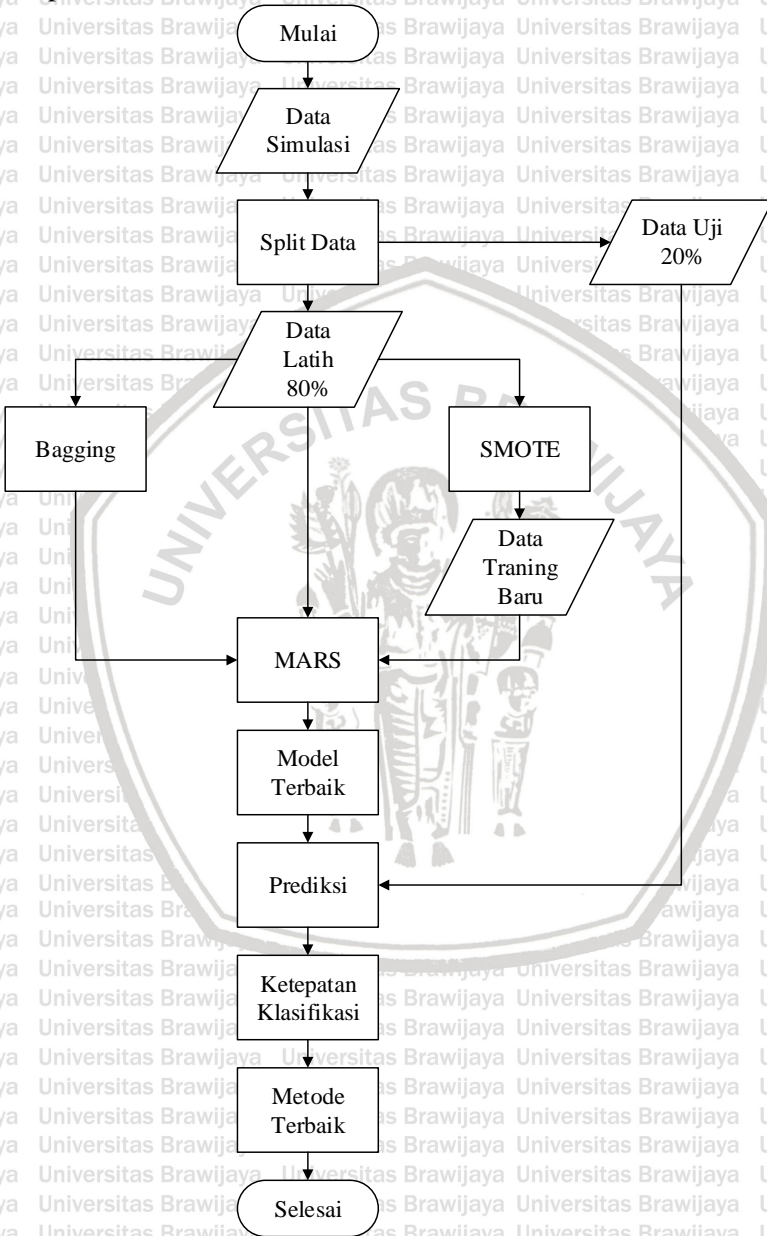
3.4. Diagram Alir

Gambar 3.1. merupakan diagram alir pembangkitan data dalam penelitian ini



Gambar 3.2. Diagram Alir Pembangkitan Data

Gambar 3.2 merupakan diagram alir metode analisis dalam penelitian ini



Gambar 3.3. Diagram Alir Metode Analisis

BAB IV HASIL DAN PEMBAHASAN

4.1. Proses Simulasi

Pada penelitian ini simulasi pembangkitan data menggunakan empat ukuran sampel yaitu 100, 200, 500 dan 1000 dengan tiga tingkat proporsi kelas minoritas yaitu 10%, 15% dan 20% yang ada pada setiap ukuran sampel. Dilakukan perulangan sebanyak 200 kali pada setiap kombinasi ukuran sampel dan proporsi kelas minoritas.

Proses simulasi dilakukan dengan beberapa langkah. Langkah awal yaitu melakukan pendugaan parameter regresi logistik biner pada data awal yaitu data yang ada pada skripsi Naufal (2019) dengan peubah respon merupakan kolektibilitas kredit dan enam peubah prediktor yaitu jangka waktu kredit, angsuran kredit, suku bunga, umur kredit, umur debitur dan jenis kelamin. Didapatkan hasil pendugaan parameter sebagai berikut: $\beta_0 = 4,01$; $\beta_1 = 0,01$; $\beta_2 = -0,06$; $\beta_3 = -5,91$; $\beta_4 = -0,26$; $\beta_5 = 0,02$ dan $\beta_6 = 0,37$. Hasil dari pendugaan parameter digunakan sebagai parameter awal pada proses simulasi yaitu membangkitkan nilai peubah respon sebanyak jumlah data awal (m) yang mengikuti distribusi *Bernaulli* dengan parameter π yang dihitung dari persamaan 3.1 di mana x_1, x_2, x_3, x_4, x_5 dan x_6 merupakan nilai variabel prediktor yang ada pada data awal. Setelah membangkitkan peubah respon langkah berikutnya yaitu membuat data *frame* antara peubah prediktor dan peubah respon hasil pembangkitan.

Dalam pembangkitan data pada empat ukuran sampel diatur agar memunculkan kasus ketidakseimbangan proporsi banyak pengamatan antar kategori pada peubah respon, di mana penelitian ini menggunakan tiga proporsi kelas minoritas yaitu 10%, 15% dan 20% dengan cara membagi dua data *frame* yaitu data *frame* yang memiliki respon 0 (kelas mayoritas) dan respon 1 (kelas minoritas). Pada data *frame* yang memiliki respon 1 (kelas minoritas) diambil secara acak sebanyak 10%, 15% dan 20% dari ukuran sampel yang digunakan pada penelitian ini. Pada data *frame* yang memiliki respon 0 (kelas mayoritas) diambil secara acak sebanyak 90%, 85% dan 80% dari ukuran sampel yang digunakan pada penelitian ini. Tabel 4.1 merupakan penjelasan dari banyak pengamatan, ukuran sampel, tingkat proporsi kelas minoritas, peubah dan sebaran yang digunakan dalam proses simulasi.



Tabel 4.1. Banyak pengamatan, ukuran sampel (n), tingkat proporsi kelas minoritas (p), peubah dan sebaran yang digunakan dalam proses simulasi

Peubah dan sebaran	p	n	Banyaknya pengamatan
- peubah prediktor (x_1, x_2, x_3, x_4, x_5 dan x_6) bersifat <i>fix</i> diambil dari data asli. - peubah respon (y) dibangkitkan sebanyak pengamatan pada data asli (m) dengan sebaran $y \sim \text{Bernaulli}(\pi)$	10%	100	Kelas minoritas : $10\% \times 100 = 10$ Kelas mayoritas : $90\% \times 100 = 90$
		200	Kelas minoritas : $10\% \times 200 = 20$ Kelas mayoritas : $90\% \times 200 = 180$
		500	Kelas minoritas : $10\% \times 500 = 50$ Kelas mayoritas : $90\% \times 500 = 450$
		1000	Kelas minoritas : $10\% \times 1000 = 100$ Kelas mayoritas : $90\% \times 1000 = 900$
	15%	100	Kelas minoritas : $15\% \times 100 = 15$ Kelas mayoritas : $85\% \times 100 = 85$
		200	Kelas minoritas : $15\% \times 200 = 30$ Kelas mayoritas : $85\% \times 200 = 170$
		500	Kelas minoritas : $15\% \times 500 = 75$ Kelas mayoritas : $85\% \times 500 = 425$
		1000	Kelas minoritas : $15\% \times 1000 = 150$ Kelas mayoritas : $85\% \times 1000 = 850$
20%	100	Kelas minoritas : $20\% \times 100 = 20$ Kelas mayoritas : $80\% \times 100 = 80$	
	200	Kelas minoritas : $20\% \times 200 = 40$ Kelas mayoritas : $80\% \times 200 = 160$	
	500	Kelas minoritas : $20\% \times 500 = 100$ Kelas mayoritas : $80\% \times 500 = 400$	
	1000	Kelas minoritas : $20\% \times 1000 = 200$ Kelas mayoritas : $80\% \times 1000 = 800$	

Setelah mengambil sampel sebanyak ukuran sampel, langkah selanjutnya yaitu data disimpan dan kemudian dibagi menjadi 2 kelompok data yaitu data latih dan data uji di mana proporsi banyak pengamatan pada data latih yaitu 80% dan pada data uji yaitu 20%. Data latih digunakan untuk membangun model klasifikasi dan data uji digunakan untuk melihat performa model klasifikasi pada data lain yang mencerminkan populasi. Pada data latih dilakukan pemodelan klasifikasi menggunakan MARS, *Bagging*-MARS dan SMOTE-MARS dan kemudian dihitung ukuran ketepatan klasifikasinya berupa



perhitungan *Accuracy*, *Sensitivity* dan *Apparent Error Rate (APER)* pada data latih dan data uji. Dalam mengukur kebaikan model juga dihitung nilai GCV yang dihasilkan dari pemodelan pada data latih. Simulasi ini dilakukan 200 kali pada setiap kombinasi ukuran sampel dan tingkat proporsi kelas minoritas.

4.2. Hasil Simulasi

4.2.1. Akurasi

1) Pada data latih

Akurasi adalah persentase ketepatan klasifikasi yang dilakukan oleh suatu model dalam mengklasifikasikan suatu data. Sehingga semakin tinggi nilai akurasi maka model tersebut semakin tepat dalam mengklasifikasikan suatu data. Pada data latih dihitung akurasi model untuk setiap satu kali simulasi. Dari 200 akurasi pada data latih yang dihasilkan dari proses simulasi, dihitung rata-ratanya yang disajikan pada Tabel 4.2.

Tabel 4.2. Rata-rata akurasi yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latih

Metode		MARS			Bagging-MARS			SMOTE-MARS		
		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,95	0,91	0,87	0,97	0,95	0,93	0,97	0,95	0,93
		30	69	90	36	71	70	47	44	18
	200	0,93	0,90	0,87	0,95	0,93	0,91	0,93	0,91	0,88
		79	61	49	43	24	21	94	08	34
	500	0,92	0,89	0,85	0,93	0,90	0,87	0,91	0,88	0,85
		71	12	95	44	31	52	43	36	27
	1000	0,91	0,88	0,85	0,92	0,89	0,86	0,90	0,87	0,84
		99	66	45	42	21	15	69	73	37

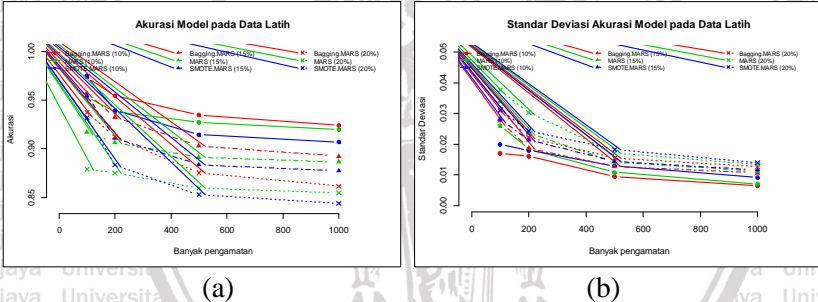
Berdasarkan Tabel 4.2. terlihat bahwa nilai akurasi yang dihasilkan setiap metode pada setiap kombinasi ukuran sampel dan tingkat proporsi kelas minoritas memiliki nilai akurasi diatas 80%. Untuk melihat kekonsistenaan (kehomogenan) nilai akurasi yang dihasilkan suatu model pada data latih maka perlu dihitung standar deviasinya yang disajikan pada Tabel 4.3.



Tabel 4.3. Standar deviasi akurasi yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latih

Metode		MARS			Bagging-MARS			SMOTE-MARS		
		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,02	0,03	0,03	0,01	0,02	0,03	0,01	0,02	0,03
		60	38	76	70	76	08	99	81	13
	200	0,01	0,02	0,03	0,01	0,01	0,02	0,01	0,02	0,02
		84	36	04	61	87	22	80	14	43
	500	0,01	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,01
		08	42	71	95	27	54	29	43	82
	1000	0,00	0,01	0,01	0,00	0,01	0,01	0,00	0,01	0,01
		70	11	34	64	06	26	91	16	39

Untuk melihat pola pergerakan nilai akurasi dan standar deviasinya pada data latih maka perlu dibuat *line plot* kedua statistik tersebut yang ditampilkan pada Gambar 4.1.



Gambar 4.1. Rata-rata akurasi (a) dan standar deviasi akurasi (b) pada data latih

Berdasarkan Gambar 4.1 (a) nilai akurasi pada data latih yang dihasilkan oleh setiap model memiliki pola menurun seiring bertambahnya ukuran sampel. Perbedaan bentuk garis menunjukkan perbedaan tingkat proporsi kelas minoritas. Semakin seimbang data latih maka hasil akurasi model pada data latih semakin kecil untuk setiap model. Perbedaan warna garis menunjukkan perbedaan model yang digunakan. Model *Bagging*-MARS menghasilkan nilai akurasi pada data latih yang lebih tinggi dibandingkan 2 metode lain dan model *SMOTE*-MARS menghasilkan nilai akurasi pada data latih yang paling kecil dibandingkan 2 metode lain.

Berdasarkan Gambar 4.1 (b) terlihat bahwa keragaman nilai akurasi pada data latih yang dihasilkan oleh setiap model semakin mengecil seiring bertambahnya ukuran sampel, sehingga semakin bertambahnya ukuran sampel nilai akurasi yang dihasilkan semakin homogen. Perbedaan bentuk garis menunjukkan perbedaan tingkat proporsi kelas minoritas, di mana keragaman nilai akurasi yang dihasilkan tidak jauh berbeda pada setiap tingkat proporsi kelas minoritas. Perbedaan warna garis menunjukkan perbedaan model yang digunakan. Dari 200 kali simulasi, model *Bagging-MARS* menghasilkan nilai akurasi lebih homogen pada data latih dibandingkan 2 metode lain karena memiliki nilai simpangan baku yang paling kecil pada setiap kombinasi ukuran sampel dan tingkat proporsi kelas minoritas.

2) Pada data uji

Model yang dibangun pada data latih digunakan untuk klasifikasi pada data uji sehingga didapatkan nilai akurasi model pada data uji. Dari 200 akurasi yang dihasilkan pada data uji, dihitung rata-ratanya yang disajikan pada Tabel 4.4.

Tabel 4.4. Rata-rata akurasi yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji

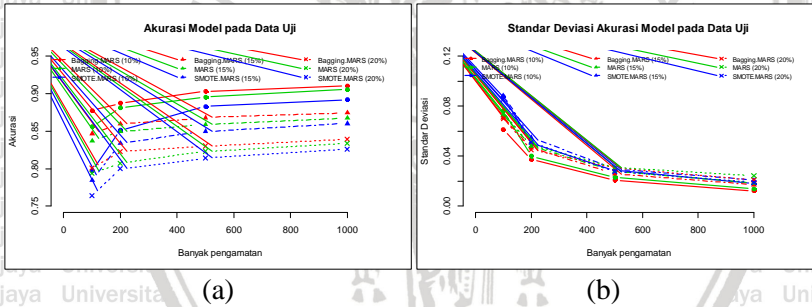
Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,85	0,83	0,79	0,87	0,84	0,80	0,79	0,78	0,76
		60	70	38	75	70	07	70	45	40
	200	0,88	0,84	0,80	0,88	0,85	0,82	0,85	0,83	0,79
		14	96	71	75	99	28	14	38	99
	500	0,89	0,85	0,82	0,90	0,86	0,83	0,88	0,84	0,81
		55	90	32	32	81	03	30	99	43
	1000	0,90	0,86	0,83	0,91	0,87	0,83	0,89	0,86	0,82
		60	76	38	04	47	96	21	02	57

Berdasarkan Tabel 4.4 terlihat bahwa nilai akurasi yang dihasilkan setiap metode pada setiap kombinasi ukuran sampel dan tingkat proporsi kelas minoritas memiliki nilai akurasi di antara 75% dan 95%. Untuk melihat kekonsistenaan (kehomogenan) nilai akurasi yang dihasilkan suatu model pada data uji maka perlu dihitung standar deviasinya yang disajikan pada Tabel 4.5.

Tabel 4.5. Standar deviasi akurasi yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji

Metode		MARS			Bagging-MARS			SMOTE-MARS			
		10%	15%	20%	10%	15%	20%	10%	15%	20%	
n	100	0,07	0,07	0,07	0,06	0,07	0,07	0,08	0,08	0,08	
		63	25	43	15	06	01	76	85	57	
		0,04	0,04	0,04	0,03	0,04	0,04	0,05	0,05	0,05	
	200	00	80	88	74	65	53	08	48	11	
		0,02	0,02	0,03	0,02	0,02	0,03	0,02	0,02	0,02	
		33	80	11	08	62	07	86	97	78	
	500	100	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,02	0,01
		0	39	86	46	23	72	12	82	08	89

Untuk melihat pola pergerakan nilai akurasi dan standar deviasinya pada data uji maka perlu dibuat *line plot* kedua statistik tersebut yang ditampilkan pada Gambar 4.2.



Gambar 4.2. Rata-rata akurasi (a) dan standar deviasi akurasi (b) pada data uji

Berdasarkan Gambar 4.2 (a) nilai akurasi pada data uji yang dihasilkan oleh setiap model memiliki pola naik seiring bertambahnya ukuran sampel sehingga model yang dibangun pada data latih memiliki performa yang semakin baik pada data uji seiring bertambahnya ukuran sampel. Perbedaan bentuk garis menunjukkan perbedaan tingkat proporsi kelas minoritas. Semakin seimbang data uji maka hasil akurasi model semakin kecil untuk setiap model, sehingga model yang dibangun pada data latih memiliki performa yang semakin baik dalam mengklasifikasikan data uji seiring berkurangnya keseimbangan data. Perbedaan warna garis



menunjukkan perbedaan model yang digunakan. Model *Bagging-MARS* menghasilkan nilai akurasi pada data uji yang lebih tinggi dibandingkan 2 metode lain dan model *SMOTE-MARS* menghasilkan nilai akurasi pada data uji yang paling kecil dibandingkan 2 metode lain.

Berdasarkan Gambar 4.2 (b) terlihat bahwa keragaman nilai akurasi pada data uji yang dihasilkan oleh setiap model memiliki pola turun seiring bertambahnya ukuran sampel, sehingga semakin bertambahnya ukuran sampel nilai akurasi yang dihasilkan semakin homogen. Perbedaan bentuk garis menunjukkan perbedaan tingkat proporsi kelas minoritas, di mana keragaman nilai akurasi yang dihasilkan tidak jauh berbeda pada setiap tingkat proporsi kelas minoritas. Perbedaan warna garis menunjukkan perbedaan model yang digunakan. Dari 200 kali simulasi, model *Bagging-MARS* menghasilkan nilai akurasi lebih homogen pada data uji dibandingkan 2 metode lain.

4.2.2. Sensitifitas

1) Pada data latih

Sensitifitas adalah persentase amatan yang tepat diprediksikan sebagai kejadian gagal ($y=0$) dari semua observasi kejadian gagal. Pada penelitian ini kejadian gagal merupakan respon dengan kelas mayoritas. Pada data latih dihitung sensitifitas model pada setiap satu kali simulasi. Dari 200 nilai sensitifitas yang dihasilkan dari proses simulasi, dihitung rata-ratanya yang disajikan pada Tabel 4.6.

Tabel 4.6. Rata-rata sensitifitas yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latih

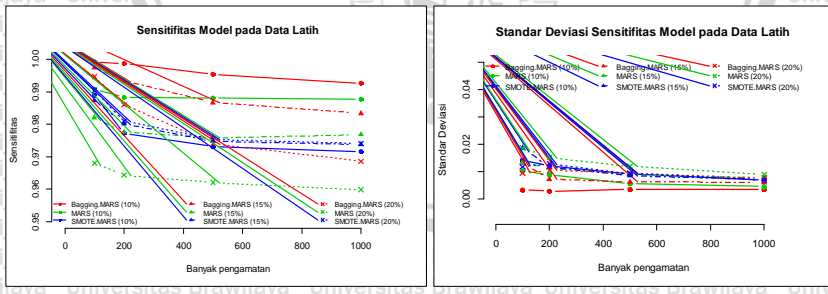
Meto- de		MARS			<i>Bagging-MARS</i>			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,99 08	0,98 21	0,96 80	0,99 92	0,99 74	0,99 47	0,98 91	0,98 73	0,99 05
	200	0,98 83	0,97 76	0,96 43	0,99 87	0,99 34	0,98 61	0,97 72	0,98 02	0,98 11
	500	0,98 81	0,97 58	0,96 20	0,99 54	0,98 67	0,97 40	0,97 31	0,97 48	0,97 53
	1000	0,98 77	0,97 68	0,95 98	0,99 26	0,98 34	0,96 86	0,97 15	0,97 38	0,97 41

Berdasarkan Tabel 4.6 terlihat bahwa nilai sensitifitas pada data latih yang dihasilkan setiap model pada setiap kombinasi ukuran sampel dan tingkat proporsi kelas minoritas memiliki nilai sensitifitas diatas 95%. Untuk melihat kekonsistenan (kehomogenan) nilai sensitifitas yang dihasilkan suatu model pada data latih maka perlu dihitung standar deviasinya yang disajikan pada Tabel 4.7.

Tabel 4.7. Standar deviasi sensitifitas yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latih

Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,01	0,01	0,01	0,00	0,01	0,00	0,01	0,01	0,01
		03	33	86	32	32	93	41	86	17
	200	0,00	0,01	0,01	0,00	0,00	0,01	0,01	0,01	0,01
		88	13	52	28	71	04	20	15	28
	500	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00
		56	89	19	34	62	94	91	85	93
100	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
0	44	75	90	35	59	77	69	70	68	

Untuk melihat pola pergerakan nilai sensitifitas dan standar deviasinya pada data latih maka perlu dibuat *line plot* kedua statistik tersebut yang ditampilkan pada Gambar 4.3.



(a) (b)

Gambar 4.3. Rata-rata sensitifitas (a) dan standar deviasi sensitifitas (b) pada data latih

Berdasarkan Gambar 4.3 (a) nilai sensitifitas pada data latih yang dihasilkan oleh setiap model semakin mengecil seiring bertambahnya ukuran sampel. Semakin seimbang data latih maka

hasil sensitifitas model semakin kecil untuk model *Bagging*-MARS dan MARS, namun untuk model SMOTE-MARS nilai sensitifitas relatif homogen pada setiap tingkat proporsi kelas minoritas. Pada tingkat proporsi kelas minoritas 10% dan 15% model *Bagging*-MARS efektif digunakan karena memiliki nilai sensitifitas yang selalu tinggi dibandingkan 2 metode lain. Pada tingkat proporsi 20% model MARS memiliki nilai sensitifitas yang paling kecil pada setiap ukuran sampel dibanding 2 metode lain.

Berdasarkan Gambar 4.3 (b) terlihat bahwa seiring bertambahnya ukuran sampel, keragaman nilai sensitifitas ketiga model semakin sama. Semakin seimbang data latih maka keragaman nilai sensitifitas model semakin sama untuk model *Bagging*-MARS dan MARS, namun untuk model SMOTE-MARS nilai sensitifitas relatif homogen pada setiap tingkat proporsi kelas minoritas. Pada tingkat proporsi 20% model MARS memiliki keragaman nilai sensitifitas yang paling kecil besar untuk setiap ukuran sampel dibanding 2 metode lain.

2) Pada data uji

Model yang dibangun pada data latih digunakan untuk klasifikasi pada data uji kemudian dihitung sensitifitas model pada data uji. Didapatkan 200 nilai sensitifitas yang dihasilkan pada data uji dan dihitung rata-ratanya yang disajikan pada Tabel 4.8.

Tabel 4.8. Rata-rata sensitifitas yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji

Meto- de	MARS			<i>Bagging</i> -MARS			SMOTE-MARS			
	p	10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,91	0,92	0,90	0,93	0,93	0,91	0,84	0,85	0,84
		36	15	78	83	34	78	22	29	94
	200	0,94	0,94	0,91	0,96	0,95	0,93	0,91	0,91	0,91
		83	10	95	19	38	47	31	85	50
500	0,96	0,95	0,93	0,97	0,96	0,94	0,95	0,95	0,94	
	99	73	97	86	76	75	25	07	67	
1000	0,97	0,96	0,94	0,98	0,97	0,95	0,96	0,96	0,96	
	89	59	62	51	35	52	14	22	11	

Berdasarkan Tabel 4.8 terlihat bahwa sensitifitas yang dihasilkan memiliki nilai diatas 80%. Untuk melihat kekonsistenan

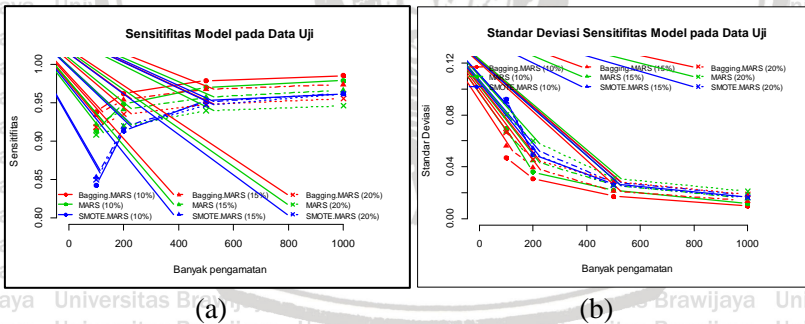


(kehomogenan) nilai sensitifitas yang dihasilkan suatu model pada data uji maka perlu dihitung standar deviasinya yang disajikan pada Tabel 4.9.

Tabel 4.9. Standar deviasi sensitifitas yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji

Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,06	0,06	0,08	0,04	0,05	0,06	0,09	0,09	0,09
		95	68	02	69	61	65	19	02	11
	200	0,03	0,04	0,06	0,03	0,03	0,04	0,04	0,05	0,05
		60	49	01	09	95	57	96	41	03
	500	0,02	0,02	0,03	0,01	0,02	0,02	0,02	0,02	0,02
		18	58	10	73	14	90	66	81	57
	1000	0,01	0,01	0,02	0,00	0,01	0,01	0,01	0,01	0,01
		0	18	54	10	97	37	86	68	82

Untuk melihat pola pergerakan nilai sensitifitas dan standar deviasinya pada data uji maka perlu dibuat *line plot* kedua statistik tersebut yang ditampilkan pada Gambar 4.4.



Gambar 4.4. Rata-rata sensitifitas (a) dan standar deviasi sensitifitas (b) pada data uji

Berdasarkan Gambar 4.4 (a) nilai sensitifitas pada data uji yang dihasilkan oleh setiap model semakin meningkat seiring bertambahnya ukuran sampel. Perbedaan bentuk garis menunjukkan perbedaan tingkat proporsi kelas minoritas, di mana semakin seimbang data uji maka hasil akurasi model semakin kecil untuk model *Bagging-MARS* dan *MARS*, namun untuk model *SMOTE-*



MARS nilai sensitifitas relatif sama pada setiap tingkat proporsi kelas minoritas. Model *Bagging*-MARS menghasilkan nilai sensitifitas pada data uji yang lebih tinggi dibandingkan 2 metode lain.

Berdasarkan Gambar 4.4 (b) terlihat bahwa seiring bertambahnya ukuran sampel, keragaman nilai sensitifitas pada data uji yang dihasilkan oleh setiap model semakin kecil sehingga semakin homogen nilai sensitifitas seiring bertambahnya ukuran sampel. Semakin besar ukuran sampel, keragaman sensitifitas yang dihasilkan pada data uji akan bernilai relatif sama dengan keragaman sensitifitas yang dihasilkan pada data latih. Pada tingkat proporsi kelas minoritas kecil (10% dan 15%) metode *Bagging*-MARS mampu memiliki nilai sensitifitas paling tinggi dan memiliki keragaman yang paling kecil.

4.2.3. APER

1) Pada data latih

Terdapat kaitan antara nilai akurasi yang dihasilkan suatu model dengan nilai APER. Semakin tinggi akurasi yang dihasilkan maka nilai APER semakin kecil dan begitu pula sebaliknya. Model yang baik apabila memiliki nilai APER yang semakin kecil karena APER merupakan persentase kesalahan klasifikasi yang dilakukan oleh suatu model dalam mengklasifikasikan data. Pada data latih dihitung nilai APER untuk setiap satu kali simulasi sehingga didapatkan 200 nilai APER pada data latih yang dihasilkan dari proses simulasi. Dihitung rata-rata nilai APER untuk setiap model yang disajikan pada Tabel 4.10.

Tabel 4.10. Rata-rata APER yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latih

Meto- de		MARS			<i>Bagging</i> -MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,04 70	0,08 31	0,12 09	0,02 64	0,04 29	0,06 30	0,02 53	0,04 56	0,06 82
	200	0,06 21	0,09 39	0,12 51	0,04 57	0,06 76	0,08 79	0,06 06	0,08 93	0,11 66
	500	0,07 30	0,10 88	0,14 05	0,06 56	0,09 69	0,12 48	0,08 57	0,11 64	0,14 73
	1000	0,08 01	0,11 34	0,14 55	0,07 58	0,10 79	0,13 85	0,09 31	0,12 27	0,15 63

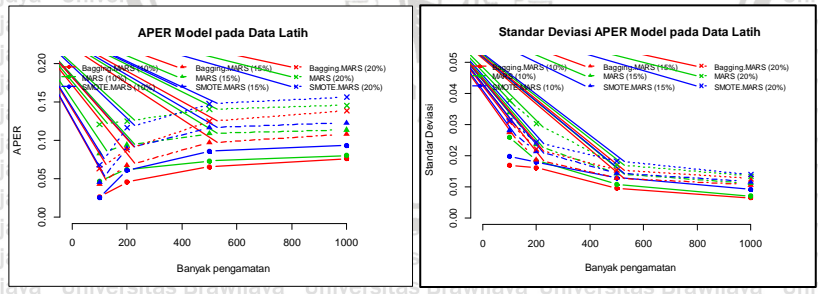


Berdasarkan Tabel 4.10 terlihat bahwa nilai APER yang dihasilkan setiap metode pada setiap kombinasi ukuran sampel dan tingkat proporsi kelas minoritas memiliki nilai APER dibawah 20%. Untuk melihat kekonsistenaan (kehomogenan) nilai APER yang dihasilkan suatu model pada data latih maka perlu dihitung standar deviasinya yang disajikan pada Tabel 4.11.

Tabel 4.11. Standar deviasi APER yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data latih

Metode		MARS			Bagging-MARS			SMOTE-MARS		
		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,02	0,03	0,03	0,01	0,02	0,03	0,01	0,02	0,03
		60	38	76	70	76	08	99	81	13
	200	0,01	0,02	0,03	0,01	0,01	0,02	0,01	0,02	0,02
		84	36	04	61	87	22	80	14	43
	500	0,01	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,01
		08	42	71	95	27	54	29	43	82
	1000	0,00	0,01	0,01	0,00	0,01	0,01	0,00	0,01	0,01
		70	11	34	64	06	26	91	16	39

Untuk melihat pola pergerakan nilai APER dan standar deviasinya pada data latih maka perlu dibuat *line plot* kedua statistik tersebut yang ditampilkan pada Gambar 4.5.



Gambar 4.5. Rata-rata APER (a) dan standar deviasi APER (b) pada data latih

Berdasarkan Gambar 4.5 (a) nilai APER pada data latih yang dihasilkan oleh setiap model memiliki pola yang naik seiring bertambahnya ukuran sampel. Semakin seimbang data latih nilai



APER model pada data latih semakin besar untuk setiap model. Perbedaan warna garis menunjukkan perbedaan model yang digunakan. Model *Bagging*-MARS menghasilkan nilai APER pada data latih yang lebih kecil dibandingkan 2 metode lain dan model SMOTE-MARS menghasilkan nilai APER pada data latih yang paling besar dibandingkan 2 metode lain.

Berdasarkan Gambar 4.5 (b) terlihat bahwa keragaman nilai APER pada data latih yang dihasilkan oleh setiap model semakin mengecil seiring bertambahnya ukuran sampel, sehingga nilai APER yang dihasilkan semakin homogen. Perbedaan bentuk garis menunjukan perbedaan tingkat proporsi kelas minoritas, di mana keragaman nilai APER yang dihasilkan tidak jauh berbeda pada setiap tingkat proporsi kelas minoritas. Dari 200 kali simulasi, model *Bagging*-MARS menghasilkan nilai APER lebih homogen pada data latih dibandingkan 2 metode lain karena memiliki nilai simpangan baku yang paling kecil pada setiap kombinasi ukuran sampel dan tingkat proporsi kelas minoritas.

2) Pada data uji

Model yang dibangun pada data latih digunakan untuk memprediksi kelas pada data uji sehingga didapatkan nilai APER model pada data uji. Dari 200 nilai APER yang dihasilkan pada data uji, dihitung rata-ratanya yang disajikan pada Tabel 4.12.

Tabel 4.12. Rata-rata APER yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji

Metode	MARS			<i>Bagging</i> -MARS			SMOTE-MARS			
	p	10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,14 40	0,16 30	0,20 63	0,12 25	0,15 38	0,19 93	0,20 30	0,21 55	0,23 60
	200	0,11 86	0,15 04	0,19 29	0,11 25	0,14 01	0,17 73	0,14 86	0,16 63	0,20 01
	500	0,10 46	0,14 10	0,17 68	0,09 68	0,13 19	0,16 98	0,11 71	0,15 01	0,18 58
	1000	0,09 40	0,13 25	0,16 62	0,08 97	0,12 53	0,16 05	0,10 80	0,13 98	0,17 43

Berdasarkan Tabel 4.12 terlihat bahwa nilai APER yang dihasilkan setiap metode pada setiap kombinasi ukuran sampel dan

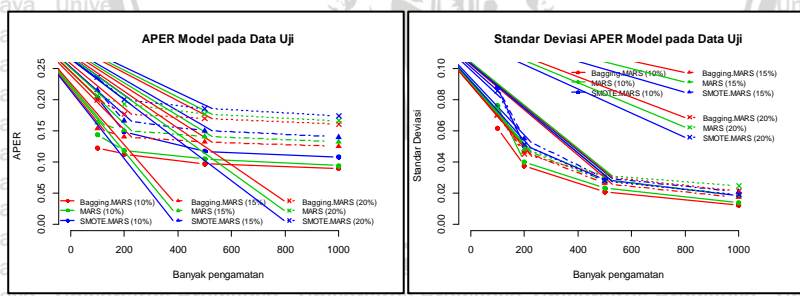


tingkat proporsi kelas minoritas memiliki nilai akurasi bawah 25%. Untuk melihat kekonsistenaan (kehomogenan) nilai APER yang dihasilkan suatu model pada data uji maka perlu dihitung standar deviasinya yang disajikan pada Tabel 4.13.

Tabel 4.13. Standar deviasi APER yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel pada data uji

Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,07	0,07	0,07	0,06	0,07	0,07	0,08	0,08	0,08
		63	25	45	15	00	01	76	85	57
	200	0,04	0,04	0,04	0,03	0,04	0,04	0,05	0,05	0,05
		00	80	88	74	65	53	08	48	11
	500	0,02	0,02	0,03	0,02	0,02	0,03	0,02	0,02	0,02
		33	80	11	08	62	07	86	97	78
1000	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,02	0,01	
	0	39	86	46	23	72	12	82	08	89

Untuk melihat pola pergerakan nilai APER dan standar deviasinya yang dihitung dari data uji maka perlu dibuat *line plot* kedua statistik tersebut yang ditampilkan pada Gambar 4.6.



(a) (b)

Gambar 4.6. Rata-rata APER (a) dan standar deviasi APER (b) pada data uji

Berdasarkan Gambar 4.6 (a) nilai APER pada data uji yang dihasilkan oleh setiap model memiliki pola turun seiring bertambahnya ukuran sampel, sehingga model yang dibangun pada data latih memiliki performa yang semakin baik pada data uji seiring

bertambahnya ukuran sampel. Semakin seimbang data uji maka nilai APER model semakin membesar untuk setiap model, sehingga model yang dibangun pada data latih memiliki performa yang semakin baik dalam mengklasifikasikan data uji yang memiliki sifat tidak seimbang. Model *Bagging*-MARS menghasilkan nilai APER yang lebih rendah dibandingkan 2 metode lain dan model SMOTE-MARS menghasilkan nilai APER pada data uji yang paling besar dibandingkan 2 metode lain.

Berdasarkan Gambar 4.6 (b) terlihat bahwa keragaman nilai APER yang dihitung pada data uji memiliki pola turun seiring bertambahnya ukuran sampel, sehingga semakin bertambahnya ukuran sampel nilai APER yang dihasilkan semakin homogen. Keragaman nilai APER yang dihasilkan tidak jauh berbeda pada setiap tingkat proporsi kelas minoritas. Dari 200 kali simulasi, model *Bagging*-MARS menghasilkan nilai akurasi lebih homogen pada data uji dibandingkan 2 metode lain.

4.2.4. GCV

Untuk melihat kebaikan model yang dihasilkan pada data simulasi maka perlu dihitung nilai GCV. Semakin kecil nilai GCV menunjukkan semakin kecilnya galat yang dihasilkan oleh model. Model yang memiliki galat kecil merupakan model yang memiliki performa baik dalam mengklasifikasikan data. Pada 200 kali simulasi dihitung nilai GCV. Dari 200 nilai GCV yang dihasilkan kemudian dihitung rata-ratanya pada setiap kombinasi ukuran sampel dan tingkat proporsi kelas minoritas yang disajikan pada Tabel 4.14.

Tabel 4.14. Rata-rata GCV yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel

Metode		MARS			<i>Bagging</i> -MARS			SMOTE-MARS		
		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,06	0,09	0,12	0,03	0,06	0,08	0,04	0,04	0,05
		24	63	28	48	51	42	18	95	35
		0,06	0,09	0,12	0,04	0,07	0,09	0,06	0,06	0,06
	200	71	53	01	77	19	39	24	44	64
		0,06	0,09	0,11	0,05	0,08	0,10	0,07	0,07	0,08
		85	57	87	84	40	57	36	80	07
	500	0,07	0,09	0,11	0,06	0,08	0,11	0,08	0,08	0,08
		02	52	66	44	86	00	28	46	61
		0								

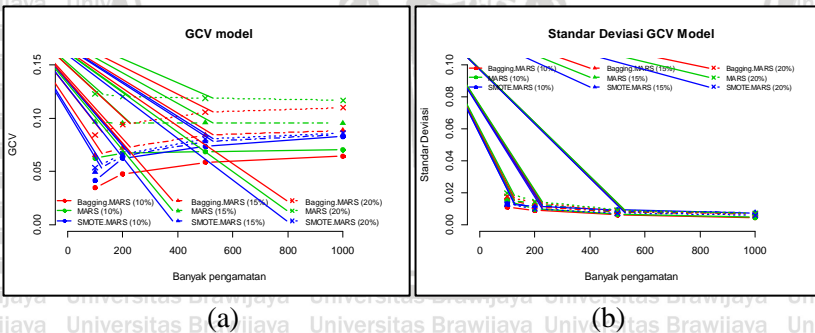


Untuk melihat kekonsistenaan (kehomogenan) nilai GCV yang dihasilkan suatu model maka perlu dihitung standar deviasinya yang disajikan pada Tabel 4.15.

Tabel 4.15. Standar deviasi GCV yang dihasilkan pada setiap metode, proporsi data tidak seimbang dan ukuran sampel

Metode	MARS			Bagging-MARS			SMOTE-MARS			
	10%	15%	20%	10%	15%	20%	10%	15%	20%	
p	10%	15%	20%	10%	15%	20%	10%	15%	20%	
n	10	0,01	0,01	0,01	0,01	0,06	0,01	0,01	0,01	
	0	51	64	98	08	71	75	31	28	27
	20	0,00	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,00
	0	98	28	46	90	10	38	16	19	99
	50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	0	65	83	91	62	79	85	93	86	72
	10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	00	46	72	70	45	68	68	71	75	59

Untuk melihat pola pergerakan nilai GCV dan standar deviasinya yang dihitung dari data uji maka perlu dibuat *line plot* kedua statistik tersebut yang ditampilkan pada Gambar 4.7



Gambar 4.7. Rata-rata GCV (a) dan standar deviasi GCV (b)

Berdasarkan Gambar 4.7 (a) terlihat bahwa nilai GCV pada model MARS cenderung konstan pada setiap ukuran contoh namun pada model *Bagging-MARS* dan *SMOTE-MARS* nilai GCV cenderung naik seiring bertambahnya ukuran sampel. Nilai GCV yang dihasilkan model *SMOTE-MARS* cenderung sama pada setiap tingkat proporsi kelas minoritas namun nilai GCV yang dihasilkan model

MARS dan *Bagging*-MARS cenderung turun seiring turunnya keseimbangan data.

Berdasarkan Gambar 4.7. (b) terlihat bahwa keragaman nilai GCV yang dihasilkan oleh setiap model memiliki nilai yang kecil yaitu dibawah 0,02. Keragaman nilai GCV ketiga model tidak berbeda jauh. Keragaman nilai GCV cenderung konstan pada setiap tingkat proporsi kelas minoritas. Terlihat bahwa keragaman nilai GCV yang dihasilkan dari ketiga model dan ketiga tingkatan kelas minoritas hampir sama.

4.3. Pembahasan Hasil Simulasi

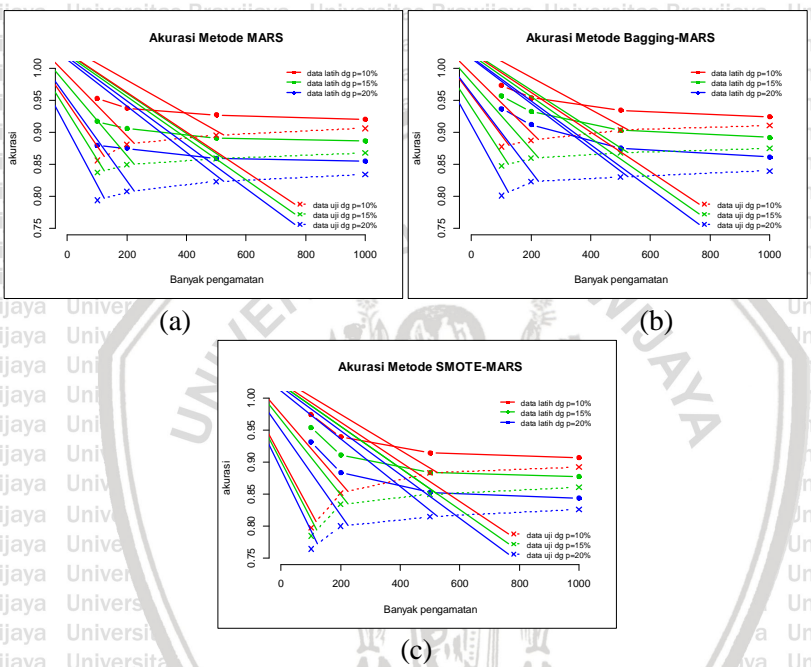
Dari hasil akurasi yang didapatkan pada data latih terlihat nilai akurasi yang dihasilkan semakin menurun seiring dengan bertambahnya jumlah sampel. Ini disebabkan karena model yang digunakan yaitu MARS memiliki fleksibilitas yang tinggi dalam membangun model klasifikasi. Dengan bertambahnya ukuran sampel menyebabkan semakin banyak pula jumlah pengamatan pada kelas minoritas yang digunakan dalam membangun model MARS sehingga model yang terbentuk cenderung semakin buruk dalam mengklasifikasikan data pada kelas mayoritas, dapat dilihat pada nilai sensitifitas yang dihasilkan setiap model memiliki nilai yang turun seiring bertambahnya jumlah sampel pada data latih. Kesalahan klasifikasi yang dihasilkan oleh setiap metode dilihat dari nilai APER yang dihasilkan pada data latih yang cenderung naik seiring bertambahnya ukuran sampel.

Dari hasil akurasi dan sensitifitas yang didapatkan pada data uji terlihat bahwa nilai yang dihasilkan semakin tinggi seiring bertambahnya jumlah sampel. Ini disebabkan karena model MARS yang dibangun pada data latih yang memiliki ukuran sampel lebih besar mendapatkan lebih banyak informasi dalam membangun model dibandingkan data latih dengan ukuran sampel lebih kecil. Model yang dibentuk memiliki performa yang semakin baik dalam memprediksi kelas pada data uji yang merepresentasikan populasi yang tidak diketahui karakteristiknya. Kebaikan performa pada setiap metode yang digunakan untuk mengklasifikasikan data uji dapat dilihat dari nilai APER yang semakin kecil seiring bertambahnya ukuran sampel.

Dari hasil akurasi dan sensitifitas yang dihasilkan pada data latih dan data uji terlihat bahwa nilai yang dihasilkan semakin turun seiring bertambahnya jumlah pengamatan pada kelas minoritas

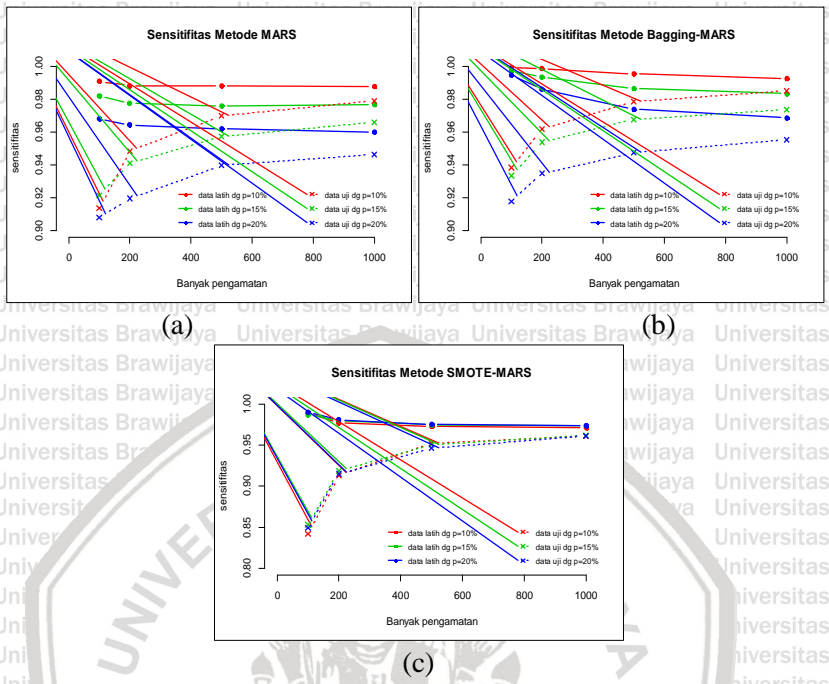


(semakin seimbang data). Ini dikarenakan model MARS memiliki fleksibilitas yang tinggi dalam membangun model sehingga semakin banyak jumlah pengamatan pada kelas minoritas yang digunakan dalam membangun model MARS maka model yang terbentuk cenderung semakin buruk dalam mengklasifikasikan data pada kelas mayoritas.



Gambar 4.8. Plot Akurasi Metode MARS (a), *Bagging*-MARS (b) dan SMOTE-MARS (c)

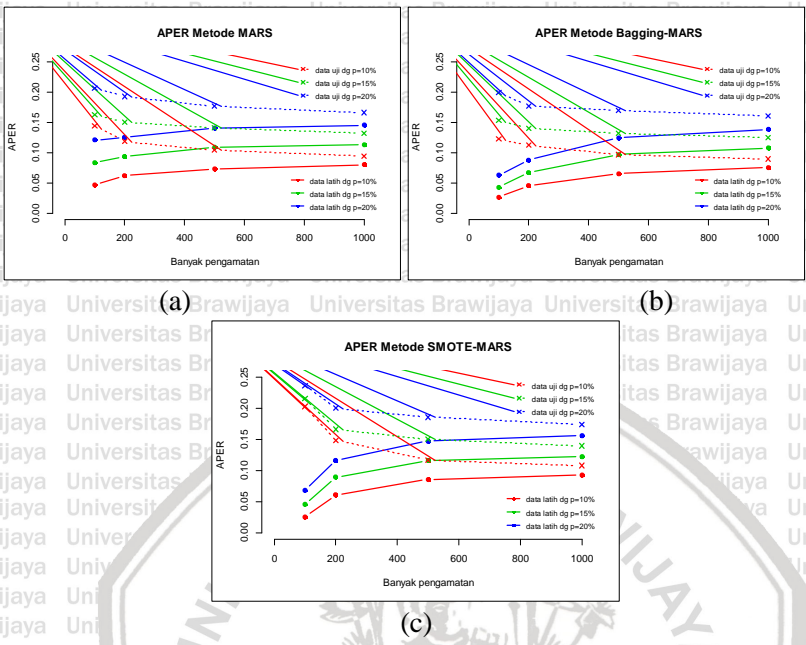
Gambar 4.8 (a) merupakan plot nilai akurasi terhadap ukuran sampel pada data latih dan data uji untuk metode MARS. Gambar 4.8 (b) merupakan plot nilai akurasi terhadap ukuran sampel pada data latih dan data uji untuk metode *Bagging*-MARS. Gambar 4.8 (c) merupakan plot nilai akurasi terhadap ukuran sampel pada data latih dan data uji untuk metode SMOTE-MARS. Dari ketiga plot tersebut terlihat bahwa nilai akurasi akan menuju satu nilai tertentu untuk setiap tingkatan proporsi kelas minoritas.



Gambar 4.9. Plot Sensitifitas Metode MARS(a),*Bagging*-MARS(b) dan SMOTE-MARS(c)

Gambar 4.9 (a) merupakan plot nilai sensitifitas terhadap ukuran sampel pada data latih dan data uji untuk metode MARS. Gambar 4.9 (b) merupakan plot nilai sensitifitas terhadap ukuran sampel pada data latih dan data uji untuk metode *Bagging*-MARS. Gambar 4.9 (c) merupakan plot nilai sensitifitas terhadap ukuran sampel pada data latih dan data uji untuk metode SMOTE-MARS. Dari ketiga plot tersebut terlihat bahwa nilai sensitifitas akan menuju satu nilai tertentu untuk setiap tingkatan proporsi kelas minoritas. Metode SMOTE-MARS menghasilkan nilai sensitifitas yang relatif sama pada setiap tingkat proporsi kelas minoritas. Ini disebabkan karena metode SMOTE-MARS membangun model bukan pada data latih dan data uji melainkan pada data yang telah dilakukan SMOTE.





Gambar 4.10. Plot APER Metode MARS (a), *Bagging*-MARS (b) dan SMOTE-MARS (c)

Gambar 4.10 (a) merupakan plot nilai APER terhadap ukuran sampel pada data latih dan data uji untuk metode MARS. Gambar 4.8 (b) merupakan plot nilai APER terhadap ukuran sampel pada data latih dan data uji untuk metode *Bagging*-MARS. Gambar 4.8 (c) merupakan plot nilai APER terhadap ukuran sampel pada data latih dan data uji untuk metode SMOTE-MARS. Dari ketiga plot tersebut terlihat bahwa nilai akurasi akan menuju satu nilai tertentu untuk setiap tingkatan proporsi kelas minoritas.



BAB V KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari hasil analisis yang dilakukan dengan menghitung ukuran ketepatan klasifikasi yang dihitung dari pemodelan MARS, *Bagging*-MARS dan SMOTE-MARS dapat ditarik kesimpulan bahwa semakin besar ukuran sampel yang digunakan nilai akurasi dan sensitifitas ketiga model pada data uji cenderung naik dan nilai APER yang dihasilkan ketiga model tersebut semakin turun. Semakin banyak pengamatan pada kelas minoritas nilai akurasi dan sensitifitas ketiga model pada data uji cenderung turun dan nilai APER yang dihasilkan ketiga model tersebut semakin naik. Metode *Bagging*-MARS menghasilkan nilai akurasi dan sensitifitas paling tinggi dibandingkan kedua model lain. Ini juga dapat dilihat dari nilai APER yang dihasilkan metode *Bagging*-MARS paling kecil dibandingkan kedua metode lainnya. Nilai GCV untuk mengukur kebaikan model yang dihasilkan model SMOTE-MARS cenderung sama pada setiap tingkat proporsi kelas minoritas namun nilai GCV yang dihasilkan model MARS dan *Bagging*-MARS cenderung turun seiring turunnya keseimbangan data.

5.2. Saran

Saran yang dapat diberikan penulis setelah penelitian ini yaitu:

1. Bagi studi kasus yang memiliki permasalahan ketidakseimbangan data seperti pada kasus data kolektibilitas kredit bank, kasus data pasien terkena penyakit HIV AIDS dan lain-lain dapat menggunakan metode *Bagging*-MARS karena hasil klasifikasi pada metode *Bagging*-MARS menghasilkan ketepatan klasifikasi yang lebih tinggi dari pada metode MARS dan SMOTE-MARS.
2. Bagi penelitian selanjutnya dapat menggunakan metode *ensemble* lain seperti *Boosting* atau *AdaBoost* untuk meningkatkan hasil akurasi klasifikasi pada data tidak seimbang.



DAFTAR PUSTAKA

- Abraham, A. & Steinberg, D. 2002. MARS: Still an Alien Planet in Soft Computing?. In: *International conference on Computational Sciencw (Proceedings)*. Part II. Vol. 2. pp. 235-244. Springer. San Francisco. California. USA.
- Adiangga, D. 2015. *Perbandingan Multivariate Adaptive Regression Spline (MARS) dan Pohon Klasifikasi C5.0 Pada Data Tidak Simbang (Studi Kasus: Pekerja Anak di Jakarta)*. Tesis. IPB. Bogor
- Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. Wiley Interscience. New York.
- Arleina, O.D., & Otok, B.W. 2014. Bootstrap Aggregating Multivariate Adaptive regression Splines (bagging MARS) untuk Mengklasifikasikan Rumah Tangga Miskin di Kabupaten Jombang. *Jurnal Sains dan Seni Pomits*. Vol.3. No.2. pp. 91-96.
- Breiman, L. 1991. Discussion: Multivariate adaptive Regression Splines. *The Annals of Statistics*. Vol. 19. pp. 82-92.
- _____. 1996a. Bagging predictors. *Machine Learning*, Vol. 24 No. 2. pp. 123-140. <http://link.springer.com/10.1007/BF00058655>. Tanggal akses 16 September 2019.
- Buhlmann, P.L. 2003. *Bagging, Subbagging and Bragging for Improving Some Prediction Algorithms*. Research Report No. 113. ETH Zentrum. CH-8092 Zurich. Switzerland.
- Chawla, N.V. Bowyer, K.W. Hall, L.O. & Kegelmeyer, W.P. 2002. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*. Vol. 16. pp. 321-357.
- De Veaux, R.D. & Ungar, LH. 1994. *Multicollinearity: a Tale of two Nonparametric Regressions*. Princeton University. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.2345&rep=rep1&type=pdf>. Tanggal akses 16 September 2019.



Eubank, R. L. 1988. *Nonparametric Regression and Spline Smoothing* 2nd edition. Marcel Dekker, Inc. New York.

Friedman, J.H. 1991. Multivariate adaptive Regression Splines. *The Annals of Statistics*. Vol. 19. No. 1. pp. 1-14.

Han, J. & Kamber, M., 2006. *Data Mining Concepts and Techniques*. 2nd ed. US: Elsevier.

Hastie, T., Tibshirani, R. & Friedman, J.H. 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction* 2nd Edition. Springer Science Business Media. New York.

IRIS. 2005. *Note on Assessment and Improvement of Tool Accuracy*. Mimeograph. IRIS Center, University of Maryland

Kriner, M. 2007. *Survival Analysis with Multivariate adaptive Regression Splines*. Disertasi Fakultas Matematika, Informatika dan Statistika Universitas Munchen. http://edoc.ub.uni-muenchen.de/7369/1/Kriner_Monika.pdf. Tanggal akses 16 September 2019.

Li, F. 2000. *Multicollinearity*. Department of Statistic. Stockholm University. <http://people.su.se/~fli/teaching/econometrics/slides/F7-Multicollinearity.pdf>. Tanggal akses 16 September 2019.

Mina, C.D. 2008. *Unearthing Poverty with MARS: Application of Multivariate Adaptive Regression Spine (MARS) in Identifying Household Poverty Correlates in The Philippines*. The Faculty of school of Statistics University of The Philippines. Diliman. http://www.salford-systems.com/doc/christian_mina.pdf. Tanggal akses 16 September 2019.

Moisen, G.G. & Frescino, T.S. 2001. *Comparing Five Modelling Techniques for Predicting Forest Characteristics*. Ecol Modell. 157. pp. 209-225

Nash, M.S. & Bradford D.F. 2001. *Parametric and Non Parametric Logistic Regression for Prediction of Precense/ Absence of an Amphibian*. Las Vegas, Nevada.

Naufal, H.I. 2019. *Ensemble Adaboost pada Metode Classification and Regression Trees dalam Mengatasi Class Imbalance pada*



Data Kolektibilitas Kredit Bank. Skripsi. Universitas Brawijaya. Malang.

Steinberg, D. Colla, P.L. & Martin, K. 2001. *MARS™ User Guide*. Salford Sytems. San Diego, CA.

Suyanto. 2018. *Machine Learning Tingkat Dasar dan Lanjutan*. Informatika. Bandung

Yanmin, S., Wong, C. & Kamel, M., 2007. *Pattern Recognition. Cost-Sensitive Boosting for Classification of Imbalanced Data*, pp. 3358-3360.



LAMPIRAN

Lampiran 1. Syntax R untuk membangkitkan data tidak seimbang

```
library(caret)
library(car)
library(caTools)
library(readxl)

#impor data dari excel
Data.asli = read_excel("E:/Kuliah/SKRIPSI/Data Kesehatan Kredit Bank X telat 1 th.xlsx", col_types = c("numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric"))
Data.asli=as.data.frame(Data.asli)
head(Data.asli)

#matriks peubah bebas
x0=rep(1,nrow(Data.asli))
bebas=as.matrix(Data.asli[,-7])
X=as.matrix(cbind(x0,bebas))

#menduga parameter model regresi logistik
reglog=glm(COL~., data = Data.asli, family = binomial, control = list(maxit=1000))
summary(reglog)
vif(reglog) #in library(car)
beta=as.matrix(round(reglog$coefficients, digits = 2))
beta

#membangkitkan respon
y=X%*%beta
pel=exp(y) / (1+exp(y))
respon=rbinom(nrow(Data.asli),1,pel)

#data simulasi
data.simul=as.data.frame(cbind(X[,2:7],respon))

#mengambil contoh acak
s = 500 #jumlah data bangkitan (100,200,500 dan 1000)
p = 0.1 #proporsi kelas minoritas (0.1,0.15 dan 0.2)
```


Lanjutan Lampiran 1.

```
#mengambil data acak dari data dengan respon=1
dataminor=data.simul[which(data.simul$respon=="1"),
]
n.minor=nrow(dataminor)
sampelminor=round(p*s)
idminor=seq(1:n.minor)
sample.idminor=sample(idminor,sampelminor,replace =
FALSE)
sampeldata.minor=dataminor[sample.idminor,]
head(sampeldata.minor)

#mengambil data acak dari data dengan respon=0
datamayor=data.simul[which(data.simul$respon=="0"),
]
n.mayor=nrow(datamayor)
sampelmayor=round(s-(p*s))
idmayor=seq(1:n.mayor)

sample.idmayor=sample(idmayor,sampelmayor,replace =
FALSE)
sampeldata.mayor=datamayor[sample.idmayor,]
head(sampeldata.mayor)

#data hasil simulasi dan pengambilan contoh acak
sample.simul=rbind(sampeldata.minor,sampeldata.mayo
r)
sample.simul$GENDER=as.factor(sample.simul$GENDER)
#MALE=1, FEMALE=0
sample.simul$respon=as.factor(sample.simul$respon)
#MINOR=1, MAYOR=0
sample.simul

#split data latih dan data uji
label.split=sample.split(sample.simul$respon,SplitR
atio=0.8)
data.latih=subset(sample.simul, label.split == TRUE
)
data.uji=subset(sample.simul, label.split == FALSE)
```

Lampiran 2. Syntax perhitungan ketepatan klasifikasi pada metode MARS

```
#model MARS
library(earth)
MARS=earth(respon~.,data = data.latih,glm = list(family=binomial("logit"),control = list(maxit=1000)),
degree = 2) #in library(earth)
model.MARS=summary(MARS)

GCV.MARS[b]=model.MARS$gcv

#prediksi kelas pada data latih untuk model MARS
MARSprob.latih=predict(MARS,type = "response")
MARSpred.latih=ifelse(MARSprob.latih>0.5,1,0)
CMLatih.MARS=confusionMatrix(table(MARSpred.latih,data.latih$respon))
akurasi.MARS.latih[b]=CMLatih.MARS$overall[1]
sensitifitas.MARS.latih[b]=CMLatih.MARS$byClass[1]
spesitifitas.MARS.latih[b]=CMLatih.MARS$byClass[2]
APER.MARS.latih[b]=1-akurasi.MARS.latih[b]

#prediksi kelas pada data uji untuk model MARS
MARSprob.uji=predict(MARS,newdata = data.uji,type = "response")
MARSpred.uji=ifelse(MARSprob.uji>0.5,1,0)
CMuji.MARS=confusionMatrix(table(MARSpred.uji,data.uji$respon))
akurasi.MARS.uji[b]=CMuji.MARS$overall[1]
sensitifitas.MARS.uji[b]=CMuji.MARS$byClass[1]
spesitifitas.MARS.uji[b]=CMuji.MARS$byClass[2]
APER.MARS.uji[b]=1-akurasi.MARS.uji[b]
```



Lampiran 3. Syntax perhitungan ketepatan klasifikasi pada metode *Bagging*-MARS

```

#model Bagging MARS
r=25 #r=jumlah resampling
n=nrow(data.latih)
m=nrow(data.uji)
v=ncol(data.latih)
library(earth)
predik.latih=matrix(0,n,r)
predik.uji=matrix(0,m,r)
gcv=matrix(0,r,1)
for(j in 1:r){
+ resampling=matrix(0,n,v)
+ index=round(runif(n,1,n))
+ for(i in 1:n){
+ for(k in 1:v){
+ resampling[i,k]=data.latih[index[i],k]}
+ resampling[, (v-1)]=ifelse(resampling[, (v-1)]==1
,0,1)
+ resampling[,v]=ifelse(resampling[,v]==1,0,1)
+ resampling=as.data.frame(resampling)
+ colnames(resampling) <- c("TERM", "PMTAMT", "RATE
", "AOC", "AGE", "GENDER", "respon")
+ resampling$GENDER=as.factor(resampling$GENDER)
+ resampling$respon=as.factor(resampling$respon)
+ colnames(resampling) <- c("TERM", "PMTAMT", "RATE
", "AOC", "AGE", "GENDER", "respon")
+ Bagging.MARS=earth(respon~.,data=resampling, gl
m = list(family=binomial("logit"),control=list(maxi
t=1000)), degree = 2)
+ model.bagging.MARS=summary(Bagging.MARS)
+ gcv[j,]=model.bagging.MARS$gcv
+ BaggingMARS.problatih=predict(Bagging.MARS,newd
ata = data.latih,type = "response")
+ predik.latih[,j]=ifelse(BaggingMARS.problatih>0
.5,-1,1)
+ BaggingMARS.probuji=predict(Bagging.MARS,newdat
a = data.uji,type = "response")
+ predik.uji[,j]=ifelse(BaggingMARS.probuji>0.5,-
1,1)}
GCV.bagging.MARS[b]=mean(gcv)

```



Lanjutan Lampiran 3.

```
#prediksi kelas pada data latih untuk model bagging
MARS
BaggingMARS.prediklatih=rep(0,n)
for(l in 1:n){
+   BaggingMARS.prediklatih[l]=sum(predik.latih[l,])
}
+   BaggingMARS.prediklatih[l]=ifelse(BaggingMARS.p
rediklatih[l]>0,0,1) }
Cmlatih.baggingMARS=confusionMatrix(table(BaggingMA
RS.prediklatih,data.latih$respon))
akurasi.baggingMARS.latih[b]=Cmlatih.baggingMARS$ov
erall[1]
sensitifitas.baggingMARS.latih[b]=Cmlatih.baggingMA
RS$byClass[1]
spesitifitas.baggingMARS.latih[b]=Cmlatih.baggingMA
RS$byClass[2]
APER.baggingMARS.latih[b]=1-akurasi.baggingMARS.lat
ih[b]

#prediksi kelas pada data uji untuk model bagging.M
ARS
BaggingMARS.predikuji=matrix(0,m,1)
for(h in 1:m){
+   BaggingMARS.predikuji[h,]=sum(predik.uji[h,])
+   BaggingMARS.predikuji[h,]=ifelse(BaggingMARS.pr
edikuji[h,]>0,0,1) }
CMuji.baggingMARS=confusionMatrix(table(BaggingMARS
.predikuji,data.uji$respon))
akurasi.baggingMARS.uji[b]=CMuji.baggingMARS$overal
l[1]
sensitifitas.baggingMARS.uji[b]=CMuji.baggingMARS$b
yClass[1]
spesitifitas.baggingMARS.uji[b]=CMuji.baggingMARS$b
yClass[2]
APER.baggingMARS.uji[b]=1-akurasi.baggingMARS.uji[b]
]
```



Lampiran 4. Syntax perhitungan ketepatan klasifikasi pada metode SMOTE-MARS

```
#pemodelan SMOTE MARS
library(DMwR)
library(earth)

#membuat data sintetis
data.buatan=SMOTE(respon~.,data=data.latih,900,5,60) #in library(DMwR)

#model SMOTE.MARS
SMOTE.MARS=earth(respon~.,data=data.buatan, glm = 1
ist(family=binomial("logit"),control = list(maxit=1
000)), degree = 2) #in library(earth)
model.SMOTE.MARS=summary(SMOTE.MARS)

GCV.SMOTE.MARS[b]=model.SMOTE.MARS$gcv

#prediksi kelas pada data latih untuk model SMOTE.M
ARS
SMOTEMARS.problatih=predict(SMOTE.MARS,newdata = da
ta.latih,type = "response")
SMOTEMARS.predlatih=ifelse(SMOTEMARS.problatih>0.5,
1,0)
head(SMOTEMARS.predlatih)
Cmlatih.SMOTEMARS=confusionMatrix(table(SMOTEMARS.p
redlatih,data.latih$respon))
akurasi.SMOTEMARS.latih[b]=Cmlatih.SMOTEMARS$overal
l[1]
sensitifitas.SMOTEMARS.latih[b]=Cmlatih.SMOTEMARS$b
yClass[1]
spesitifitas.SMOTEMARS.latih[b]=Cmlatih.SMOTEMARS$b
yClass[2]
APER.SMOTEMARS.latih[b]=1-akurasi.SMOTEMARS.latih[b
]

#prediksi kelas pada data uji untuk model SMOTE.MAR
S
SMOTEMARS.probuji=predict(SMOTE.MARS,newdata = data
.uji,type = "response")
SMOTEMARS.preduji=ifelse(SMOTEMARS.probuji>0.5,1,0)
CMuji.SMOTEMARS=confusionMatrix(table(SMOTEMARS.pre
duji,data.uji$respon))
```



Lanjutan Lampiran 4.

```

akurasi.SMOTEMARS.uji [b]=CMuji.SMOTEMARS$overall [1]
sensitifitas.SMOTEMARS.uji [b]=CMuji.SMOTEMARS$byCla
ss [1]
spesitifitas.SMOTEMARS.uji [b]=CMuji.SMOTEMARS$byCla
ss [2]
APER.SMOTEMARS.uji [b]=1-akurasi.SMOTEMARS.uji [b]

```



Lampiran 5. Contoh *output* perhitungan ketepatan klasifikasi metode MARS

```
> MARS=earth(respon~.,data = data.latih,glm = list(
family=binomial("logit"),control = list(maxit=1000)
),degree = 2) #in library(earth)
> summary(MARS)
```

```
Call:earth(formula=respon~.,data=data.latih,glm=lis
t(family=binomial("logit"),control=list(maxit=1...
),degree=2)
```

GLM coefficients

```

1
(Intercept)      0.02169779
h(13-AOC)        -0.27718684
h(AOC-13)        0.09424587
h(120-TERM) * RATE 0.71582343
```

GLM (family binomial, link logit):

```

nulldev df      dev df      devratio      AIC iters
converged
135.267 159   99.9932 156         0.261    108     5
1
```

Earth selected 4 of 17 terms, and 3 of 6 predictors

Termination condition: Reached nk 21

Importance: AOC, TERM, RATE, PMTAMT-unused, AGE-unused, GENDER1-unused

Number of terms at each degree of interaction: 1 2

```

1
Earth GCV 0.1025697      RSS 14.71378      GRSq 0.20555
66      RSq 0.2787364
```

#prediksi kelas pada data latih untuk model MARS

```
> MARSprob.latih=predict(MARS,type = "response")
> MARSpred.latih=ifelse(MARSprob.latih>0.5,1,0)
> confusionMatrix(table(MARSpred.latih,data.latih$
respon))
```

Confusion Matrix and Statistics

```

MARSpred.latih  0   1
                0 130  14
                1   6  10
```

Lanjutan Lampiran 5.

```

Accuracy : 0.875
95% CI : (0.8136, 0.9219)
No Information Rate : 0.85
P-Value [Acc > NIR] : 0.2223
Kappa : 0.4318
Mcnemar's Test P-Value : 0.1175
Sensitivity : 0.9559
Specificity : 0.4167
Pos Pred Value : 0.9028
Neg Pred Value : 0.6250
Prevalence : 0.8500
Detection Rate : 0.8125
Detection Prevalence : 0.9000
Balanced Accuracy : 0.6863
'Positive' Class : 0

```

```

#prediksi kelas pada data uji untuk model MARS
> MARSprob.uji=predict(MARS,newdata = data.uji,type
= "response")
> MARSpred.uji=ifelse(MARSprob.uji>0.5,1,0)
> confusionMatrix(table(MARSpred.uji,data.uji$respo
n))

```

Confusion Matrix and Statistics

MARSpred.uji	0	1
0	32	4
1	2	2

```

Accuracy : 0.85
95% CI : (0.7016, 0.9429)
No Information Rate : 0.85
P-Value [Acc > NIR] : 0.6067
Kappa : 0.3182
Mcnemar's Test P-Value : 0.6831
Sensitivity : 0.9412
Specificity : 0.3333
Pos Pred Value : 0.8889
Neg Pred Value : 0.5000
Prevalence : 0.8500
Detection Rate : 0.8000
Detection Prevalence : 0.9000
Balanced Accuracy : 0.6373
'Positive' Class : 0

```



Lampiran 6. Contoh *output* perhitungan ketepatan klasifikasi metode *Bagging-MARS*

```
> confusionMatrix(table(BaggingMARS.prediklatih, data.latih$respon))
```

Confusion Matrix and Statistics

BaggingMARS.prediklatih	0	1
0	134	10
1	2	14

Accuracy : 0.925

95% CI : (0.8727, 0.9606)

No Information Rate : 0.85

P-Value [Acc > NIR] : 0.003131

Kappa : 0.6591

McNemar's Test P-Value : 0.043308

Sensitivity : 0.9853

Specificity : 0.5833

Pos Pred Value : 0.9306

Neg Pred Value : 0.8750

Prevalence : 0.8500

Detection Rate : 0.8375

Detection Prevalence : 0.9000

Balanced Accuracy : 0.7843

'Positive' Class : 0

```
#prediksi kelas pada data uji untuk model bagging.MARS
```

```
> BaggingMARS.predikuji=matrix(0,m,1)
```

```
> for(h in 1:m){
```

```
+ BaggingMARS.predikuji[h,]=sum(predik.uji[h,])
```

```
+ BaggingMARS.predikuji[h,]=ifelse(BaggingMARS.predikuji[h,]>0,0,1)}
```

```
> confusionMatrix(table(BaggingMARS.predikuji,data.uji$respon))
```

Confusion Matrix and Statistics

BaggingMARS.predikuji	0	1
0	32	5
1	2	1

Accuracy : 0.825

Lanjutan Lampiran 6.

95% CI : (0.6722, 0.9266)

No Information Rate : 0.85

P-Value [Acc > NIR] : 0.7559

Kappa : 0.1358

McNemar's Test P-Value : 0.4497

Sensitivity : 0.9412

Specificity : 0.1667

Pos Pred Value : 0.8649

Neg Pred Value : 0.3333

Prevalence : 0.8500

Detection Rate : 0.8000

Detection Prevalence : 0.9250

Balanced Accuracy : 0.5539

'Positive' Class : 0



Lampiran 7. Contoh *output* perhitungan ketepatan klasifikasi metode SMOTE-MARS

```
> data.buatan=SMOTE(respon~.,data=data.latih,900,5,
200) #in library(DMwR)
#model SMOTE.MARS
> SMOTE.MARS=earth(respon~.,data=data.buatan,glm =
list(family=binomial("logit"),control = list(maxit=
1000)),degree = 2) #in library(earth)
> summary(SMOTE.MARS)
```

```
Call: earth(formula=respon~., data=data.buatan,glm=
list(family=binomial("logit"),control=list(maxit=1.
..),degree=2)
```

GLM coefficients

	1
(Intercept)	2.1210
h(1.33313-PMTAMT)	-13.7840
h(PMTAMT-1.33313)	3.3103
h(0.103812-RATE)	-64.2941
h(RATE-0.103812)	-152.1537
h(5.27037-AOC)	-7.1315
h(AOC-5.27037)	1.1749
h(AGE-31)	0.0913
h(120-TERM) * h(RATE-0.103812)	5930.0885
h(TERM-120) * h(RATE-0.103812)	167.5403
h(TERM-122.627) * h(RATE-0.103812)	-175.2260
h(TERM-180) * h(RATE-0.103812)	7.3575
h(PMTAMT-1.33313) * h(AOC-3)	-0.9538
h(PMTAMT-1.33313) * h(3-AOC)	0.5544
h(6.1288-PMTAMT) * h(5.27037-AOC)	1.2340
h(PMTAMT-6.1288) * h(5.27037-AOC)	-105.8824
h(1.33313-PMTAMT) * h(47-AGE)	0.5105

```
GLM (family binomial, link logit):
```

nulldev	df	dev	df	devratio	AIC	iters
converged						
875.961	671	265.096	655	0.697	299.1	25

```
Earth selected 17 of 19 terms, and 5 of 6 predictor
s
```

```
Termination condition: Reached nk 21
```



Lanjutan Lampiran 7.

```
Importance: AOC, TERM, RATE, PMTAMT, AGE, GENDER1-  
used  
Number of terms at each degree of interaction: 1 7  
9  
Earth GCV 0.09705543      RSS 57.50548      GRSq 0.5785  
269      RSq 0.6272793
```

#prediksi kelas pada data latih untuk model SMOTE.MARS

```
> SMOTEMARS.problatih=predict(SMOTE.MARS,newdata =  
data.latih,type = "response")  
> SMOTEMARS.predlatih=ifelse(SMOTEMARS.problatih>0.  
5,1,0)  
> confusionMatrix(table(SMOTEMARS.predlatih,data.la  
atih$respon))
```

Confusion Matrix and Statistics

SMOTEMARS.predlatih	0	1
	0 123	9
	1 13	15

Accuracy : 0.8625
95% CI : (0.7993, 0.9118)

No Information Rate : 0.85

Lajutan Lampiran 7.

```
P-Value [Acc > NIR] : 0.3788  
Kappa : 0.4954  
Mcnemar's Test P-Value : 0.5224  
Sensitivity : 0.9044  
Specificity : 0.6250  
Pos Pred Value : 0.9318  
Neg Pred Value : 0.5357  
Prevalence : 0.8500  
Detection Rate : 0.7688  
Detection Prevalence : 0.8250  
Balanced Accuracy : 0.7647  
'Positive' Class : 0
```

#prediksi kelas pada data uji untuk model SMOTE.MARS

```
> SMOTEMARS.probuji=predict(SMOTE.MARS,newdata = da  
ta.uji,type = "response")
```



Lanjutan Lampiran 7.

```
> SMOTEMARS.preduji=ifelse(SMOTEMARS.probuji>0.5,1,
0)
> confusionMatrix(table(SMOTEMARS.preduji,data.uji$
respon))
```

Confusion Matrix and Statistics

SMOTEMARS.preduji	0	1
0	31	5
1	3	1

Accuracy : 0.8
 95% CI : (0.6435, 0.9095)

No Information Rate : 0.85
 P-Value [Acc > NIR] : 0.8646
 Kappa : 0.0909

McNemar's Test P-Value : 0.7237

Sensitivity : 0.9118
 Specificity : 0.1667

Pos Pred Value : 0.8611
 Neg Pred Value : 0.2500

Prevalence : 0.8500
 Detection Rate : 0.7750

Detection Prevalence : 0.9000
 Balanced Accuracy : 0.5392

'Positive' Class : 0



Lampiran 8. Rata-rata ukuran ketepatan klasifikasi dari 200 kali ulangan

Rata-Rata Akurasi Data Latih

Meto-de		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	10	0,95	0,91	0,87	0,97	0,95	0,93	0,97	0,95	0,93
	0	30	69	90	36	71	70	47	44	18
	20	0,93	0,90	0,87	0,95	0,93	0,91	0,93	0,91	0,88
	0	79	61	49	43	24	21	94	08	34
	50	0,92	0,89	0,85	0,93	0,90	0,87	0,91	0,88	0,85
	0	71	12	95	44	31	52	43	36	27
	10	0,91	0,88	0,85	0,92	0,89	0,86	0,90	0,87	0,84
00	99	66	45	42	21	15	69	73	37	

Rata-Rata Akurasi Data Uji

Met-ode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
N	10	0,85	0,83	0,79	0,87	0,84	0,80	0,79	0,78	0,76
	0	60	70	38	75	70	07	70	45	40
	20	0,88	0,84	0,80	0,88	0,85	0,82	0,85	0,83	0,79
	0	14	96	71	75	99	28	14	38	99
	50	0,89	0,85	0,82	0,90	0,86	0,83	0,88	0,84	0,81
	0	55	90	32	32	81	03	30	99	43
	10	0,90	0,86	0,83	0,91	0,87	0,83	0,89	0,86	0,82
00	60	76	38	04	47	96	21	02	57	

Rata-Rata Sensitifitas Data Latih

Meto-de		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	10	0,99	0,98	0,96	0,99	0,99	0,99	0,98	0,98	0,99
	0	08	21	80	92	74	47	91	73	05
	20	0,98	0,97	0,96	0,99	0,99	0,98	0,97	0,98	0,98
	0	83	76	43	87	34	61	72	02	11
	50	0,98	0,97	0,96	0,99	0,98	0,97	0,97	0,97	0,97
	0	81	58	20	54	67	40	31	48	53
	10	0,98	0,97	0,95	0,99	0,98	0,96	0,97	0,97	0,97
00	77	68	98	26	34	86	15	38	41	



Rata-Rata Sensitifitas Data Uji

Meto-de		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	10	0,91	0,92	0,90	0,93	0,93	0,91	0,84	0,85	0,84
	0	36	15	78	83	34	78	22	29	94
	20	0,94	0,94	0,91	0,96	0,95	0,93	0,91	0,91	0,91
	0	83	10	95	19	38	47	31	85	50
	50	0,96	0,95	0,93	0,97	0,96	0,94	0,95	0,95	0,94
	0	99	73	97	86	76	75	25	07	67
	100	0,97	0,96	0,94	0,98	0,97	0,95	0,96	0,96	0,96
		89	59	62	51	35	52	14	22	11

Rata-Rata APER Data Latih

Meto-de		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	10	0,04	0,08	0,12	0,02	0,04	0,06	0,02	0,04	0,06
	0	70	31	09	64	29	30	53	56	82
	20	0,06	0,09	0,12	0,04	0,06	0,08	0,06	0,08	0,11
	0	21	39	51	57	76	79	06	93	66
	50	0,07	0,10	0,14	0,06	0,09	0,12	0,08	0,11	0,14
	0	30	88	05	56	69	48	57	64	73
	100	0,08	0,11	0,14	0,07	0,10	0,13	0,09	0,12	0,15
		01	34	55	58	79	85	31	27	63

Rata-Rata APER Data Uji

Meto-de		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	10	0,14	0,16	0,20	0,12	0,15	0,19	0,20	0,21	0,23
	0	40	30	63	25	38	93	30	55	60
	20	0,11	0,15	0,19	0,11	0,14	0,17	0,14	0,16	0,20
	0	86	04	29	25	01	73	86	63	01
	50	0,10	0,14	0,17	0,09	0,13	0,16	0,11	0,15	0,18
	0	46	10	68	68	19	98	71	01	58
	100	0,09	0,13	0,16	0,08	0,12	0,16	0,10	0,13	0,17
		40	25	62	97	53	05	80	98	43



Lampiran 9. Standar deviasi ukuran ketepatan klasifikasi dari 200 kali ulangan

Standar Deviasi Akurasi Data Latih

Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,02	0,03	0,03	0,01	0,02	0,03	0,01	0,02	0,03
		60	38	76	70	76	08	99	81	13
	200	0,01	0,02	0,03	0,01	0,01	0,02	0,01	0,02	0,02
		84	36	04	61	87	22	80	14	43
	500	0,01	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,01
		08	42	71	95	27	54	29	43	82
	100	0,00	0,01	0,01	0,00	0,01	0,01	0,00	0,01	0,01
		0	70	11	34	64	06	26	91	16

Standar Deviasi Akurasi Data Uji

Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,07	0,07	0,07	0,06	0,07	0,07	0,08	0,08	0,08
		63	25	43	15	06	01	76	85	57
	200	0,04	0,04	0,04	0,03	0,04	0,04	0,05	0,05	0,05
		00	80	88	74	65	53	08	48	11
	500	0,02	0,02	0,03	0,02	0,02	0,03	0,02	0,02	0,02
		33	80	11	08	62	07	86	97	78
	100	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,02	0,01
		0	39	86	46	23	72	12	82	08

Standar Deviasi Sensitifitas Data Latih

Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,01	0,01	0,01	0,00	0,01	0,00	0,01	0,01	0,01
		03	33	86	32	32	93	41	86	17
	200	0,00	0,01	0,01	0,00	0,00	0,01	0,01	0,01	0,01
		88	13	52	28	71	04	20	15	28
	500	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00
		56	89	19	34	62	94	91	85	93
	100	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
		0	44	75	90	35	59	77	69	70



Standar Deviasi Sensitifitas Data Uji

Meto-de		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,06	0,06	0,08	0,04	0,05	0,06	0,09	0,09	0,09
		95	68	02	69	61	65	19	02	11
	200	0,03	0,04	0,06	0,03	0,03	0,04	0,04	0,05	0,05
		60	49	01	09	95	57	96	41	03
	500	0,02	0,02	0,03	0,01	0,02	0,02	0,02	0,02	0,02
		18	58	10	73	14	90	66	81	57
	1000	0,01	0,01	0,02	0,00	0,01	0,01	0,01	0,01	0,01
		18	54	10	97	37	86	68	82	62

Standar Deviasi APER Data Latih

Meto-de		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,02	0,03	0,03	0,01	0,02	0,03	0,01	0,02	0,03
		60	38	76	70	76	08	99	81	13
	200	0,01	0,02	0,03	0,01	0,01	0,02	0,01	0,02	0,02
		84	36	04	61	87	22	80	14	43
	500	0,01	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,01
		08	42	71	95	27	54	29	43	82
	1000	0,00	0,01	0,01	0,00	0,01	0,01	0,00	0,01	0,01
		70	11	34	64	06	26	91	16	39

Standar Deviasi APER Data Uji

Meto-de		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,07	0,07	0,07	0,06	0,07	0,07	0,08	0,08	0,08
		63	25	45	15	00	01	76	85	57
	200	0,04	0,04	0,04	0,03	0,04	0,04	0,05	0,05	0,05
		00	80	88	74	65	53	08	48	11
	500	0,02	0,02	0,03	0,02	0,02	0,03	0,02	0,02	0,02
		33	80	11	08	62	07	86	97	78
	1000	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,02	0,01
		39	86	46	23	72	12	82	08	89



Lampiran 10. Rata-rata dan standar deviasi GCV dari 200 kali ulangan

Rata-Rata GCV

Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,06 24	0,09 63	0,12 28	0,03 48	0,06 51	0,08 42	0,04 18	0,04 95	0,05 35
	200	0,06 71	0,09 53	0,12 01	0,04 77	0,07 19	0,09 39	0,06 24	0,06 44	0,06 64
	500	0,06 85	0,09 57	0,11 87	0,05 84	0,08 40	0,10 57	0,07 36	0,07 80	0,08 07
	100	0,07 02	0,09 52	0,11 66	0,06 44	0,08 86	0,11 00	0,08 28	0,08 46	0,08 61
	0									

Standar Deviasi GCV

Metode		MARS			Bagging-MARS			SMOTE-MARS		
p		10%	15%	20%	10%	15%	20%	10%	15%	20%
n	100	0,01 51	0,01 64	0,01 98	0,01 08	0,06 71	0,01 75	0,01 31	0,01 28	0,01 27
	200	0,00 98	0,01 28	0,01 46	0,00 90	0,01 10	0,01 38	0,01 16	0,01 19	0,00 99
	500	0,00 65	0,00 83	0,00 91	0,00 62	0,00 79	0,00 85	0,00 93	0,00 86	0,00 72
	100	0,00 46	0,00 72	0,00 70	0,00 45	0,00 68	0,00 68	0,00 71	0,00 75	0,00 59
	0									