



TUGAS AKHIR - SS 141501

**KLASIFIKASI PENDERITA RETARDASI MENTAL
DI RUMAH SAKIT JIWA MENUR PROVINSI JAWA TIMUR
MENGUNAKAN *SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE (SMOTE) - CLASSIFICATION AND
REGRESSION TREES (CART)***

**PUTRI MILAKHUL KHASANAH
NRP 1311100 116**

Dosen Pembimbing
Ir. Sri Pingit Wulandari, M.Si.

Co. Pembimbing
Santi Wulan Purnami, M.Si., Ph.D

Program Studi S1 Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015



FINAL PROJECT - SS 141501

**CLASSIFICATION OF MENTAL RETARDATION PATIENTS
IN MENUR MENTAL HOSPITAL EAST JAVA PROVINCE
USING SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE (SMOTE) - CLASSIFICATION AND
REGRESSION TREES (CART)**

PUTRI MILAKHUL KHASANAH
NRP 1311100 116

Supervisor
Ir. Sri Pingit Wulandari, M.Si.

Co. Supervisor
Santi Wulan Purnami, M.Si., Ph.D

Undergraduate Programme of Statistics
Faculty of Mathematics and Natural Sciences
Institut Teknologi Sepuluh Nopember
Surabaya 2015

LEMBAR PENGESAHAN

**KLASIFIKASI PENDERITA RETARDASI MENTAL DI
RUMAH SAKIT JIWA MENUR PROVINSI JAWA TIMUR
MENGUNAKAN SYNTHETIC MINORITY
OVERSAMPLING TECHNIQUE (SMOTE)-CLASSIFICATION
AND REGRESSION TREES (CART)**

TUGAS AKHIR

**Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember**

Oleh :

**PUTRI MILAKHUL KHASANAH
1311 100 116**

Disetujui oleh Pembimbing Tugas Akhir

**Ir. Sri Pingit Wulandari, M.Si.
NIP. 19620603 198701 2 001**

(*Sri Pingit Wulandari*)

**Santi Wulan Purnami, M.Si., Ph.D
NIP. 19720923 199803 2 001**

(*Santi Wulan Purnami*)

**Mengetahui
Ketua Jurusan Statistika FMIPA-ITS**

**Dr. Muhammad Mashuri, M.T.
NIP. 19620408 198701 1 001**

SURABAYA, JULI 2015

**KLASIFIKASI PENDERITA RETARDASI MENTAL DI
RUMAH SAKIT JIWA MENUR PROVINSI JAWA TIMUR
MENGUNAKAN *SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE (SMOTE) - CLASSIFICATION AND
REGRESSION TREES (CART)***

Nama Mahasiswa : Putri Milakhul Khasanah
NRP : 1311 100 116
Jurusan : Statistika FMIPA-ITS
Dosen Pembimbing : Ir. Sri Pingit Wulandari, M.Si.
Co. Pembimbing : Santi Wulan Purnami, M.Si.,Ph.D

Abstrak

Retardasi mental (RM) merupakan salah satu hambatan dalam pertumbuhan dan perkembangan yang dialami oleh anak. Pada setiap tahunnya, penderita retardasi mental di Indonesia mengalami peningkatan sebesar 3%. Retardasi mental telah menjadi masalah kesehatan dan sosial. Untuk itu diperlukan penanganan khusus untuk mengatasi masalah retardasi mental guna menekan angka penderita retardasi mental di Indonesia. Klasifikasi retardasi mental dapat meningkatkan penanganan pasien dalam memberikan perlakuan yang sesuai berdasarkan tipenya yakni retardasi mental ringan, sedang, berat, dan sangat berat. Penelitian ini berguna untuk mengetahui akurasi hasil klasifikasi yang sudah dilakukan oleh Rumah Sakit Jiwa Menur Surabaya provinsi Jawa Timur dan juga mengetahui variabel terpenting yang berguna untuk pengklasifikasian. Berdasarkan metode CART setelah dilakukan pre-processing dengan metode SMOTE didapatkan akurasi dengan menggunakan metode pemilihan pemilah indeks gini sebesar 72,7%. Sedangkan, akurasi dengan menggunakan metode pemilihan pemilah indeks twoing sebesar 71,4%. Faktor terpenting dalam klasifikasi penderita retardasi mental di Rumah Sakit Jiwa Menur Surabaya provinsi Jawa Timur adalah gejala psikiatrik yang ditunjukkan oleh pasien.

Kata Kunci : Akurasi, CART, Indeks Gini, Indeks Twoing, Retardasi Mental, SMOTE

“Halaman ini sengaja dikosongkan”

**CLASSIFICATION OF MENTAL RETARDATION PATIENTS
IN MENUR MENTAL HOSPITAL EAST JAVA PROVINCE
USING SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE (SMOTE) - CLASSIFICATION AND
REGRESSION TREES (CART)**

Name : Putri Milakhul Khasanah
NRP : 1311 100 116
Department : Statistika FMIPA-ITS
Supervisor : Ir. Sri Pingit Wulandari, M.Si.
Co. Supervisor : Santi Wulan Purnami, M.Si., Ph.D

Abstract

Mental retardation (MR) is one of the obstacle in the growth and development experienced by the child. Every year, people with mental retardation in Indonesia increased by 3%. Mental retardation became a health and social problem. So, it required special handling to overcome the problem of mental retardation in order to reduce the number of people with mental retardation in Indonesia. Classification of mental retardation can improve the treatment of patients in providing appropriate treatment based on the type of mental retardation i.e mild, moderate, severe, and profound. This research is useful to determine the accuracy of the results of the classification by Menur Mental Hospital Surabaya East Java province and to know the most important factors are useful for classification. Based on the method of CART after the pre-processing with SMOTE method, accuracy obtained by using the classifier selection method Gini index is 72,7%. On the other hand, accuracy by using the classifier selection method twoing index is 71,4%. The most important factor in the classification of patients with mental retardation in Menur Mental Hospitals Surabaya, East Java province is suffering psychiatric symptoms exhibited by the patients.

Keywords : Accuracy, CART, Gini Index, Mental Retardation, SMOTE, Twoing Index

“Halaman ini sengaja dikosongkan”

KATA PENGANTAR

Puji syukur kehadirat Allah SWT, atas segala rahmat dan karunia-Nya yang tak pernah henti diberikan, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul

**“Klasifikasi Penderita Retardasi Mental di Rumah Sakit Jiwa
Menur Provinsi Jawa Timur Menggunakan *Synthetic
Minority Oversampling Technique (SMOTE) - Classification
and Regression Trees (CART)*”**

dengan baik dan tepat pada waktunya. Penyusunan Tugas Akhir ini tak luput dari bantuan serta dukungan dari berbagai pihak. Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Muhammad Mashuri, M.T selaku Ketua Jurusan Statistika yang telah menyediakan fasilitas guna kelancaran pengerjaan Tugas Akhir ini.
2. Ibu Ir. Sri Pingit Wulandari, M.Si selaku dosen pembimbing yang telah sabar dan memberikan waktunya untuk membimbing dan memberikan masukan dalam penyelesaian Tugas Akhir ini.
3. Ibu Santi Wulan Purnami, M.Si., Ph.D selaku dosen co. pembimbing yang telah meluangkan waktunya untuk membimbing, memberikan bantuan berupa saran, dan juga memberikan informasi dalam penyelesaian Tugas Akhir ini.
4. Bapak Dr. Sutikno, M.Si dan Ibu Shofi Andari, M.Si selaku dosen penguji atas segala kritikan dan saran yang sangat membangun.
5. Ibu Dra. Lucia Aridinanti, M.T selaku Kaprodi S1 Jurusan Statistika ITS.
6. Ibu Dr. Irhamah, M.Si selaku dosen wali atas segala nasehat dan bimbingan yang berguna selama 8 semester pada saat perwalian.

7. Pihak Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur yang sangat ramah dan mau membagi ilmunya mengenai retardasi mental kepada penulis.
8. Ayahanda tercinta Musman Basor dan Ibunda tersayang Tuti Marlina atas doa, nasehat, dan kasih sayang yang sangat besar yang telah diberikan untuk penulis sehingga dapat menjadi motivasi dan penyemangat bagi penulis disaat menghadapi kesulitan.
9. Kakak-kakakku yang hebat Mas Ary, Mbak Ika, Mas Dwi, Mbak Neny, Mas Trias, dan Mbak Miming yang selalu memberikan nasehat dan dukungan kepada penulis untuk mengerjakan tugas akhir ini.
10. Ramadhana Dio Gradianta, seseorang yang selalu ada disaat susah maupun senang yang selalu memberikan semangat dan motivasi kepada penulis dalam menyelesaikan Tugas akhir ini.
11. Teman-teman seperjuangan Statistika ITS $\Sigma 22$ yang memberikan hangatnya sebuah kebersamaan dan rasa kekeluargaan selama 4 tahun ini.
12. Serta semua pihak yang telah memberikan dukungan baik moril maupun materiil yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari masih banyaknya kekurangan dalam pembuatan laporan Tugas Akhir ini, besar harapan bagi penulis untuk dapat menerima saran dan kritik yang bersifat membangun guna perbaikan di masa mendatang. Semoga laporan Tugas Akhir ini dapat bermanfaat bagi penelitian selanjutnya.

Surabaya, Juli 2015

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
TITLE PAGE	iii
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xix
DAFTAR LAMPIRAN	xxi
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	6
BAB II TINJAUAN PUSTAKA	
2.1 Statistik Deskriptif	7
2.2 <i>Missing Value</i>	7
2.3 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	8
2.4 <i>Classification and Regression Trees (CART)</i>	13
2.4.1 Pembentukan Pohon Klasifikasi	15
2.4.2 Pemangkasan Pohon Klasifikasi	19
2.4.3 Penentuan Pohon Klasifikasi Optimal	21
2.4.4 Ilustrasi CART dengan Pemilihan Pemilah Indeks Gini	23
2.4.5 Ilustrasi CART dengan Pemilihan Pemilah Indeks Twoing	31
2.4.6 Ukuran Ketepatan Klasifikasi	39
2.5 Retardasi Mental	41
2.5.1 Klasifikasi Retardasi Mental	41
2.5.2 Faktor Etiologi (Faktor Penyebab)	43
2.5.3 Gejala Psikiatrik (Tingkah Laku)	43

2.6 Penelitian Sebelumnya.....	44
BAB III METODE PENELITIAN	
3.1 Sumber Data.....	47
3.2 Kerangka Konsep Penelitian.....	47
3.3 Variabel Penelitian.....	48
3.4 Langkah Analisis Penelitian.....	51
3.5 Diagram Alir Penelitian.....	53
BAB IV ANALISIS DAN PEMBAHASAN	
4.1 Karakteristik Deskriptif Penderita Retardasi Mental di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur.....	55
4.2 Analisis <i>Pre-Processing</i> dengan Metode SMOTE untuk <i>Imbalanced Data</i>	63
4.3 Analisis CART untuk Klasifikasi Penderita RM di RSJ Menur Provinsi Jawa Timur.....	66
4.3.1 Kemungkinan Pemilah untuk Setiap Variabel Prediktor.....	66
4.3.2 Analisis CART Penderita RM di RSJ Menur dengan Pemilahan Pemilah Indeks Gini.....	68
4.3.3 Analisis CART Penderita RM di RSJ Menur dengan Pemilahan Pemilah Indeks Twoing.....	81
4.3.4 Perbandingan Analisis CART Penderita RM di RSJ Menur dengan Pemilah Indeks Gini dan Indeks Twoing.....	93
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan.....	95
5.2 Saran.....	96
DAFTAR PUSTAKA	97
LAMPIRAN	101
BIODATA PENULIS	149

DAFTAR TABEL

	Halaman
Tabel 2.1	Data Simulasi untuk SMOTE9
Tabel 2.2	Distribusi Data dengan Menggunakan SMOTE ... 11
Tabel 2.3	Data Simulasi Setelah Menggunakan SMOTE 11
Tabel 2.4	Variabel Penelitian Ilustrasi CART23
Tabel 2.5	Data Sampel Ilustrasi CART24
Tabel 2.6	Proporsi Pengamatan pada Respon Y25
Tabel 2.7	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> $X_1=5,5$25
Tabel 2.8	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> sebesar $X_1=8$26
Tabel 2.9	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> sebesar $X_1=11$26
Tabel 2.10	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> sebesar $X_2=1:2$27
Tabel 2.11	Perbandingan <i>Goodness of Split</i> dari Pemilah 1 ...27
Tabel 2.12	Data Sampel Ilustrasi CART Pemilahan 2.....28
Tabel 2.13	Perbandingan <i>Goodness of Split</i> dari Pemilah 2 ...29
Tabel 2.14	Pemberian Label pada Simpul Terminal 1.....30
Tabel 2.15	Proporsi Pengamatan Simpul Kiri dan Simpul Kanan dengan <i>Threshold</i> $X_1=5,5$32
Tabel 2.16	Proporsi Pengamatan Simpul Kiri dan Simpul Kanan dengan <i>Threshold</i> $X_1=8$33
Tabel 2.17	Proporsi Pengamatan Simpul Kiri dan Simpul Kanan dengan <i>Threshold</i> $X_1=11$33
Tabel 2.18	Proporsi Pengamatan Simpul Kiri dan Simpul Kanan dengan <i>Threshold</i> $X_2=1:2$34
Tabel 2.19	Nilai Impurity Variabel X_1 dan X_2 Pada Simpul Kanan35
Tabel 2.20	Pemberian Label pada Simpul Terminal 2.....36
Tabel 2.21	Pasien RM Baru dengan Karakteristik Variabel Umur Penderita (X_1) dan Faktor Genetik (X_2).....38
Tabel 2.22	Hasil Klasifikasi Pasien RM Baru39

Tabel 2.23	<i>Crosstab</i> Ketepatan Klasifikasi	40
Tabel 3.1	Struktur Data	47
Tabel 3.2	Variabel Penelitian	49
Tabel 4.1	Deskriptif Karakteristik Umur Penderita Berdasarkan Masing-Masing Tipe.....	58
Tabel 4.2	Distribusi Data Masing-Masing Tipe RM Setelah Dilakukan SMOTE.....	64
Tabel 4.3	Banyaknya Kemungkinan Pemilah Variabel Prediktor.....	67
Tabel 4.4	Variabel Terpenting dari Pohon Klasifikasi Maksimal dengan Pemilah Indeks Gini	68
Tabel 4.5	Urutan Pembentukan Pohon Klasifikasi (<i>Tree Sequence</i>) 1	71
Tabel 4.6	Pelabelan Kelas Simpul Terminal dengan Pemilah Indeks Gini	73
Tabel 4.7	Karakteristik Kelas Penderita RM Berdasarkan Simpul Terminal dengan Persentase Terbesar 1	74
Tabel 4.8	Ilustrasi Data Rekam Medis Pasien Baru.....	75
Tabel 4.9	Hasil Klasifikasi Pasien RM Baru	77
Tabel 4.10	Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data <i>Learning</i> Sebelum SMOTE	77
Tabel 4.11	Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data <i>Testing</i> Sebelum SMOTE	78
Tabel 4.12	Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data <i>Learning</i> Setelah SMOTE	79
Tabel 4.13	Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data <i>Testing</i> Setelah SMOTE	80
Tabel 4.14	Variabel Terpenting dari Pohon Klasifikasi Maksimal dengan Pemilah Indeks Twoing.....	82
Tabel 4.15	Urutan Pembentukan Pohon Klasifikasi (<i>Tree Sequence</i>) 2	84

Tabel 4.16	Variabel Terpenting dalam Pembentukan Pohon Klasifikasi Optimal	85
Tabel 4.17	Pelabelan Kelas Simpul Terminal dengan Pemilah Indeks Twoing.....	87
Tabel 4.18	Karakteristik Kelas Penderita RM Berdasarkan Simpul Terminal dengan Persentase Terbesar 2	88
Tabel 4.19	Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Twoing Data <i>Learning</i> Sebelum SMOTE	89
Tabel 4.20	Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Twoing Data <i>Testing</i> Sebelum SMOTE	90
Tabel 4.21	Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Twoing Data <i>Learning</i> Setelah SMOTE	91
Tabel 4.22	Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Twoing Data <i>Testing</i> Setelah SMOTE	92
Tabel 4.23	Perbandingan Analisis CART dengan Pemilah Indeks Gini dan Indeks Twoing.....	93

“Halaman ini sengaja dikosongkan”

DAFTAR GAMBAR

	Halaman
Gambar 2.1	Ilustrasi SMOTE 9
Gambar 2.2	Persentase Setiap Kelas..... 10
Gambar 2.3	Plot Data Setelah Replikasi dengan SMOTE 12
Gambar 2.4	Ilustrasi Pohon Klasifikasi 15
Gambar 2.5	Ilustrasi Prosedur <i>Cross Validation 10-Fold</i> 22
Gambar 2.6	Pohon Klasifikasi Ilustrasi Hasil Pemilahan I 28
Gambar 2.7	Pohon Klasifikasi Hasil Pemilahan II..... 29
Gambar 2.8	Pohon Klasifikasi dengan Label Kelas..... 31
Gambar 2.9	Pohon Klasifikasi Ilustrasi 2 Hasil Pemilahan I.... 34
Gambar 2.10	Pohon Klasifikasi Pemilahan 2 Pada Simpul Kanan..... 35
Gambar 2.11	Pohon Klasifikasi dengan Label Kelas Ilustasi 2 ... 37
Gambar 3.1	Kerangka Konsep Penelitian Teori H.L. Blum 48
Gambar 3.2	Diagram Alir Penelitian 53
Gambar 4.1	Jumlah Penderita RM di RSJ Menur Berdasarkan Masing-Masing Tipe 56
Gambar 4.2	Perbandingan Urutan Kelahiran Anak Berdasarkan Masing-Masing Tipe 56
Gambar 4.3	Perbandingan Kondisi Dilahirkan Berdasarkan Masing-Masing Tipe 57
Gambar 4.4	Perbandingan Jenis Kelamin Berdasarkan Masing-Masing Tipe 59
Gambar 4.5	Perbandingan Faktor Genetik Berdasarkan Masing-Masing Tipe 60
Gambar 4.6	Perbandingan Faktor Etiologik Berdasarkan Masing-Masing Tipe 61
Gambar 4.7	Perbandingan Gejala Psikiatrik Berdasarkan Masing-Masing Tipe 62
Gambar 4.8	Persentase Banyaknya Variabel Respon Berdasarkan Tipe RM 63
Gambar 4.9	Persentase Banyaknya Variabel Respon Setelah Menggunakan SMOTE..... 65

Gambar 4.10	Topology Pohon Klasifikasi Maksimal dengan Pemilahan Pemilah Indeks Gini	69
Gambar 4.11	Plot <i>Relative Cost</i> dalam Klasifikasi Penderita Retardasi Mental dengan pemilah Indeks Gini	70
Gambar 4.12	Topology Pohon Klasifikasi Optimal dengan Pemilahan Pemilah Indeks Gini	72
Gambar 4.13	Topology Pohon Klasifikasi Maksimal dengan Pemilahan Pemilah Indeks Twoing.....	83
Gambar 4.14	Plot <i>Relative Cost</i> dalam Klasifikasi Penderita Retardasi Mental dengan pemilah Indeks Twoing.....	84
Gambar 4.15	Topology Pohon Klasifikasi Optimal dengan Pemilahan Pemilah Indeks Twoing	86

DAFTAR PUSTAKA

- American Association on Mental Retardation. (2002). *Mental Retardation: Definition, Classification, and Systems of Supports*. 10th ed. Washington, DC: American Association on Mental Retardation.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1993). *Classification and Regression Trees*. New York : Chapman Hall.
- California Statistical Software Inc. (2000). *CART References Guide*. San Diego, California : Salford System.
- Chawla, V. N. (2003, Aug 21). *C4.5 and Imbalance Data sets: Investigating the Effect of Sampling Method, Probabilistic Estimate, and Decision Tree Structure*.
- Depsos (2008). *Penyelenggaraan Kesejahteraan Sosial di 50 Kabupaten Daerah Tertinggal*, Jakarta: Depsos RI.
- Hair J.F., Rolph E. Anderson, Ronald L. Tatham, William C. Black. (2010). *Multivariate Data Analysis*. Sixth Edition, Pearson Education Prentice Hall, Inc.
- Hendriani, W., Handariyati, R., & Sakti, T. M. (2006). *Penerimaan Keluarga Terhadap Individu yang Mengalami Keterbelakangan Mental*. Surabaya: Insan
- Inayah, R. (2014). *Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang Berdasarkan Faktor-Faktor yang Mempengaruhi dengan Pendekatan CART*. Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember, Surabaya.
- Lewis, R. J. (2000). *An Introduction to Classification and Regression Trees (CART) Analysis*. Annual Meeting of the Society for Academic Emergency Medicine. California, UCLA Medical Center.
- Lumbantobing, S. (2006). *Anak dengan Mental Terbelakang*. Jakarta: Balai Penerbit FKUI.
- Moeschler, J. B. & Shevell, M. (2013). *Clinical Genetic Evaluation of the Child with Mental Retardation or Developmental Delays*. American Academy of Pediatrics.

- Napolion, K. (2010). *Pengalaman Keluarga Dalam Merawat Anak Tunagrahita di Kelurahan Balumbang Jaya Kecamatan Bogor Barat Kota Bogor 2010: Studi Fenomenologi*. Tesis Ilmu Keperawatan Universitas Indonesia, Depok.
- Notoadmodjo, S. (2010). *Kesehatan Masyarakat Ilmu dan Seni*. Jakarta: Rineka Cipta.
- Prasadio, T. (1976). *Gangguan Psikiatrik Pada Anak-Anak dengan Retardasi Mental*. Surabaya: Universitas Airlangga.
- Rahmawati, I. (2015). *Klasifikasi Faktor-Faktor yang Mempengaruhi Korban Kecelakaan Lalu Lintas di Surabaya Dengan Pendekatan Regresi Logistik Multinomial dan Random Forests*. Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember, Surabaya.
- Retnaningtyas, A. & Setyaningsih, R. (2009). *Perilaku Seksual Remaja Retardasi Mental, Jurnal psikologi Proyeksi, Vol. 4 No.2*.
- Seftiana, D. (2014). *Klasifikasi Rumah Tangga Sangat Miskin Di Kabupaten Jombang Menurut Paket Bantuan Rumah Tangga Yang Diharapkan Dengan Pendekatan RF-Cart (Random Forest Classification And Regression Treess)*. Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember, Surabaya.
- Shevell M., Ashwal S., & Donley D. (2003). *Practice Parameter: Evaluation of The Child with Global Developmental Delay Report of The Quality Standards Subcommittee of The American Academy of Neurology and The Practice Committee of The Child Neurology Society*. *Neurology* 60, 367– 380.
- Stromme, P. (2000). *Aetiology in Severe and Mild Mental Retardation: a Population-Based Study of Norwegian Children*. *Dev Med Child Neurology* 42, 76– 86.
- Sularyo, T.S. & Kadim, M. (2000). *Topik Khusus : Retardasi Mental*. *Jurnal Sari Pediatrik. Vol. 2, No. 3, Hal. 2-4*.
- Sutton, C. D. (2005). *Classification and Regression Trees, Bagging, and Boosting*. *Handbook of Statistics*, 24,303-329.

- Tomb, A. (2004). *Buku Saku Psikiatri*. Jakarta: EGC.
- Trapsilasiwi, R. K. (2014). *Klasifikasi Multiclass untuk Imbalanced Data Menggunakan SMOTE Least Square Support Vector Machine*. Thesis Statistika Institut Teknologi Sepuluh Nopember, Surabaya.
- Walpole, Ronald E. (1995). *Pengantar Statistika*. Jakarta: Gramedia Pustaka Utama.
- Witten, Ian. H. (2011). *Data Mining Practical Machine Learning Tools and Technique*. Third Edition. New York: Morgan Kaufman.
- WHO. (1998). *Primary Prevention of Mental Neurological and Psychological Disorders*. 8-53.
- Zheng, H., Chen, L., Han, X., Zhao, X. & Ma, Yan. (2009). *Classification and Regression Trees (CART) for Analysis of Soybean Yield Variability Among Fields in Northeast China : The Importance of Phosporus Application Rates under Drought Conditions, Agriculture, Ecosystems and Environment 132, 98-105*.

“Halaman ini sengaja dikosongkan”

BIODATA PENULIS



Penulis bernama Putri Milakhul Khasanah yang akrab disapa dengan Putri, merupakan anak dari pasangan Musman Basor dan Tuti Marlina. Penulis merupakan anak bungsu dari 4 bersaudara yang dilahirkan di kota Surabaya pada tanggal 11 Mei 1993. Penulis memulai jenjang pendidikan yang pertama di Taman kanak-kanak Arisska Sidoarjo (1998-1999), kemudian melanjutkan di SDN 1 Pabean (1999-2005). Setelah menamatkan pendidikan sekolah dasar penulis melanjutkan pendidikan di SMPN 1 Waru (2005-2008) dan di SMUN 1 Waru (2008-2011). Pada tahun 2011, penulis melanjutkan pendidikan di Perguruan Tinggi Negeri ITS dan diterima sebagai mahasiswi Jurusan Statistika dengan NRP 1311100116. Selama kuliah, penulis juga aktif dalam kegiatan organisasi dan pernah menjadi staff Divisi PERS HIMASTA-ITS pada tahun 2011, kemudian pada tahun 2012 penulis menjabat sebagai Redaktur di Divisi PERS HIMASTA-ITS. Selain itu penulis juga aktif dalam kepanitiaan seperti panitia IHMSI (Ikatan Himpunan Mahasiswa Statistika Indonesia) dan PRS (Pekan Raya Statistika). Pengalaman magang, penulis memiliki pengalaman magang di BPS (Badan Pusat Statistik) Provinsi Jawa Timur dan juga di PT. Angkasa Pura I (Persero).

Penulis menerima segala kritikan, masukan, dan saran yang bersifat membangun demi meningkatkan manfaat Tugas Akhir ini. Oleh karena itu, penulis siap membantu terkait dengan metode dan ilmu statistik, jika dibutuhkan pembaca dapat menghubungi penulis melalui email : putrimilakhul@gmail.com.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu hambatan dalam pertumbuhan dan perkembangan yang dialami oleh seorang anak adalah adanya retardasi mental (RM) yang diderita oleh anak tersebut. Menurut *American Association of Mental Retardation* (2002) retardasi mental adalah kelainan pada fungsi intelektual dan terjadi disabilitas atau ketidakmampuan dalam hal menyesuaikan diri (perilaku adaptif) yang ditandai dengan fungsi intelektual di bawah rata-rata dan mengakibatkan pertumbuhan dan perkembangan anak yang tidak sempurna. Ketidakmampuan ini muncul sebelum anak berusia 18 tahun. Selain itu, retardasi mental dapat memicu atau merupakan salah satu gejala dari gangguan atau penyakit lain.

Menurut WHO (1998), kurang lebih sebanyak 2% sampai dengan 3% dari jumlah populasi dunia mengalami retardasi mental dan kejadian retardasi mental pada anak-anak di bawah umur 18 tahun di negara maju diperkirakan berkisar 0,5% sampai 2,5% sedangkan di negara berkembang mencapai 4,6%. Banyak penelitian melaporkan angka kejadian retardasi mental lebih banyak terjadi pada anak laki-laki dibandingkan anak perempuan (WHO, 1998). Sedangkan menurut Retnaningtias & Setyaningsih (2009) menyebutkan bahwa adanya perbedaan perilaku seksual remaja retardasi mental perempuan dengan remaja retardasi mental laki-laki. Terjadinya retardasi mental tidak dapat dipisahkan dari tumbuh kembang seorang anak, faktor penentu tumbuh kembang seorang anak pada garis besarnya adalah faktor genetik yang menentukan sifat bawaan anak tersebut dan faktor lingkungan (Lumbantobing, 2006).

Seperti yang diketahui, retardasi mental sudah menjadi masalah kesehatan dan kesejahteraan sosial bagi penderita, keluarga, dan masyarakat di Indonesia. Menurut Hendriani, Handariyati, dan Sakti (2006) berdasarkan data pokok sekolah luar biasa (SLB) di seluruh Indonesia, jika dilihat dari kelompok usia

sekolah, jumlah penduduk di Indonesia yang menyandang keterbelakangan mental adalah sebanyak 62.011 anak, dengan perbandingan 60% diderita oleh anak laki-laki dan 40% diderita oleh anak perempuan. Berdasarkan jumlah tersebut sebanyak 11,4% atau sebanyak 7.069 anak mengalami retardasi mental dengan proporsi untuk masing-masing tipe yakni penderita yang mengalami retardasi mental sangat berat adalah sebanyak 2,5% atau sebanyak 1.550 anak, penderita retardasi mental berat adalah sebanyak 2,8% atau sebanyak 1.736 anak, penderita retardasi mental sedang adalah sebanyak 2,6% atau sebanyak 1.612 anak, dan penderita dengan retardasi mental ringan atau lemah pikiran adalah sebanyak 3,5% atau sebanyak 2.170 anak.

Informasi juga diberikan oleh pusat data dan informasi (Pusdatin) Kesejahteraan Sosial Departemen Sosial RI tahun 2006 jumlah penyandang cacat di Indonesia adalah sebesar 2.364.000 jiwa termasuk penyandang cacat mental. *Survey* yang dilakukan oleh Depsos RI pada tahun 2007 menunjukkan bahwa jumlah penduduk di Indonesia yang tercatat sebanyak 220 juta jiwa, terdapat sebanyak 3,11% dari total penduduk di Indonesia adalah penyandang cacat atau sekitar 7,8 juta jiwa. Jumlah penyandang cacat meningkat jika dibandingkan pada tahun 2002 dengan jumlah penduduk sebesar 212.020.759 jiwa, maka jumlah penyandang cacat adalah sekitar 6.593.846 orang, sedangkan jumlah penyandang cacat mental adalah sebesar 848.083 orang (Depsos, 2008).

WHO telah mengklasifikasikan retardasi mental menjadi 4 tipe yakni *profound* (sangat berat), *severe* (berat), *moderate* (sedang), dan *mild* (ringan). Klasifikasi tersebut dilakukan untuk memudahkan proses perawatan dan pemberian *treatment* kepada masing-masing tipe penderita retardasi mental. Berdasarkan klasifikasi yang sudah ada, penelitian ini bertujuan untuk mendapatkan ketepatan klasifikasi dan juga untuk menentukan faktor-faktor terpenting yang berpengaruh terhadap pengklasifikasian. Faktor-faktor yang mempengaruhi klasifikasi penderita retardasi mental sangat beragam dan kompleks, beberapa faktor secara umum seperti faktor genetik (keturunan), faktor prenatal, faktor perinatal,

faktor posnatal, faktor lingkungan, faktor etiologi yang menjadi penyebab utama penderita mengalami retardasi mental, dan gejala psikiatrik yang ditimbulkan oleh penderita (Moeschler, 2013).

Pemberian *treatment* atau pengobatan bagi penderita retardasi mental dilakukan berdasarkan masing-masing tipe yang diderita oleh pasien. Artinya, perlakuan untuk masing-masing tipe penderita RM itu berbeda. Oleh karena itu, ketepatan dalam pemberian *treatment* sangat dipengaruhi oleh ketepatan pengklasifikasian yang telah dilakukan. Oleh sebabnya, dibutuhkan suatu metode pengklasifikasian yang bersifat nonparametrik yakni menggunakan metode *classification and regression trees* (CART). Metode CART memiliki beberapa keunggulan diantaranya mampu bekerja pada dimensi data yang besar dan struktur data yang kompleks, tidak terikat oleh asumsi kenormalan maupun variansi homogen, dapat mengetahui interaksi antar variabel prediktor dan hasil klasifikasi yang diperoleh lebih mudah dipahami serta diinterpretasikan (Lewis, 2000). Metode CART juga mampu menunjukkan karakteristik yang ditimbulkan oleh masing-masing tipe pada variabel respon sehingga memudahkan dalam hal pengklasifikasian. Beberapa penelitian menunjukkan dengan menggunakan metode CART, mampu menghasilkan akurasi yang lebih tinggi dibandingkn dengan metode konvensional lainnya seperti regresi logistik, *multivariate adaptive regression spline* (MARS), dan *support vector machine* (SVM). Namun, metode CART juga memiliki kelemahan yaitu menghasilkan pohon yang kurang stabil dimana perubahan kecil pada data *learning* mampu menyebabkan perubahan yang signifikan terhadap pohon yang terbentuk (Sutton, 2005).

Terdapat beberapa cara yang digunakan untuk meningkatkan ketepatan klasifikasi yakni dengan melakukan analisis *pre-processing* terlebih dahulu dengan melakukan analisis *missing value* yaitu tidak tersedianya informasi pada sebuah subyek (kasus) yang diakibatkan informasi tersebut sengaja tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada. Adanya kasus *missing value* dapat mengakibatkan hilangnya informasi penting

dan kurang akuratnya proses pengklasifikasian yang telah dilakukan. Selain itu, untuk mengatasi adanya kasus *imbalanced data* dibutuhkan analisis *pre-processing* untuk menyeimbangkan jumlah data pada masing-masing kelas. *Imbalanced data* adalah kondisi data yang tidak seimbang, dimana jumlah anggota kelas yang satu melebihi jumlah anggota kelas yang lain. *Imbalanced data* dapat menjadi suatu masalah karena suatu *machine learning* akan menghasilkan akurasi atau ketepatan klasifikasi yang baik terhadap kelas data *learning* dengan jumlah anggota yang banyak, sedangkan kelas dengan jumlah anggota sedikit akan memiliki akurasi yang kurang baik. Untuk mengatasi hal tersebut, dapat menggunakan metode *synthetic minority oversampling technique* (SMOTE) yakni salah satu metode *oversampling* yang ditemukan oleh Chawla (2003) berupa teknik penambahan jumlah sampel pada kelas negatif dengan melakukan replikasi data pada kelas negatif secara acak sehingga menghasilkan jumlah data yang sama dengan data pada kelas positif atau data mayor.

Untuk mengetahui ketepatan klasifikasi yang dihasilkan dan faktor-faktor yang mempengaruhi pengklasifikasian pasien penderita retardasi mental dalam penelitian ini dibutuhkan data yang akurat. Seperti yang diketahui, di Jawa Timur tepatnya di Surabaya terdapat rumah sakit jiwa yang khusus menangani penderita keterbelakangan mental dengan peningkatan jumlah pasien yang cukup banyak pada setiap tahunnya yakni di Rumah Sakit Jiwa Menur Provinsi Jawa Timur dengan menuju pada pada poli tumbuh kembang yang khusus menangani kasus pertumbuhan dan perkembangan anak diharapkan terdapat data yang sesuai sehingga dapat digunakan untuk melanjutkan penelitian pada tahap selanjutnya.

1.2 Rumusan Masalah

Retardasi mental (RM) telah menjadi masalah sosial dan kesehatan di Indonesia. Jumlah penderita retardasi mental meningkat sebesar 3% untuk setiap tahunnya. WHO telah mengklasifikasikan penderita retardasi mental menjadi 4 tipe yakni RM ringan (*mild*), RM sedang (*moderate*), RM berat (*severe*), dan RM sangat berat (*profound*). Pengklasifikasian penderita RM

berguna dalam pemberian *treatment* dan pengobatan secara tepat. Oleh karena itu, dibutuhkan suatu metode pengklasifikasian yakni dengan menggunakan metode CART yang menghasilkan pohon keputusan. Sehingga, berdasarkan penelusuran pohon keputusan tersebut pengklasifikasian penderita RM semakin mudah dilakukan. Berdasarkan uraian tersebut, maka permasalahan yang dibahas dalam penelitian ini adalah mengenai jumlah data yang tidak seimbang berdasarkan kategori pada variabel respon dengan metode *pre-processing* menggunakan SMOTE. Selain itu, pengklasifikasian penderita RM di Rumah Sakit Jiwa Menur Provinsi Jawa Timur menggunakan metode *classification and regression trees* (CART) secara efisien dan akurat dengan metode pemilah indeks gini dan indeks twoing.

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mengatasi jumlah data yang tidak seimbang untuk masing-masing kelas berdasarkan kategori pada variabel respon dengan melakukan analisis *pre-processing* menggunakan metode SMOTE.
2. Mendapatkan metode pemilihan pemilah terbaik berdasarkan perbandingan ketepatan klasifikasi yang dihasilkan oleh metode pemilihan pemilah indeks gini dan indeks twoing untuk klasifikasi penderita retardasi mental di Rumah Sakit Jiwa Menur Provinsi Jawa Timur.
3. Mengetahui karakteristik masing-masing tipe penderita retardasi mental di Rumah Sakit Jiwa Menur Provinsi Jawa Timur yang dihasilkan berdasarkan analisis CART.

1.4 Manfaat Penelitian

Ada beberapa manfaat yang diharapkan dari pelaksanaan penelitian ini adalah sebagai berikut.

1. Bagi bidang kesehatan penelitian ini berguna untuk menambah ketepatan dalam memberikan *treatment* atau perlakuan terhadap penderita retardasi mental di masa yang

akan datang berdasarkan variabel apa saja yang mempengaruhi. Selain itu, dengan menggunakan metode *classification and regression trees* (CART) diharapkan dapat memberikan kemudahan dalam proses pengklasifikasian penderita retardasi mental yang dilakukan dengan melihat karakteristik masing-masing tipe penderita retardasi mental. Kemudian ketepatan klasifikasi yang dihasilkan dapat dijadikan sebagai informasi tambahan bagi Dinas Kesehatan dan rumah sakit setempat dalam membuat kebijakan guna menekan angka penderita retardasi mental di Indonesia.

2. Bagi bidang pendidikan, penelitian ini diharapkan dapat berguna untuk menambah wawasan dan pengetahuan pada penelitian selanjutnya dalam kategori klasifikasi untuk meningkatkan kesejahteraan masyarakat.

1.5 Batasan Masalah

Unit penelitian yang digunakan dalam penelitian ini hanya menggunakan 3 kategori dalam variabel respon yakni pasien penderita retardasi mental dengan tipe ringan, sedang, dan berat. Selain itu batasan masalah yang digunakan dalam penelitian ini adalah pasien yang diteliti merupakan pasien yang sedang menjalani proses rawat jalan di Rumah Sakit Jiwa Menur Provinsi Jawa Timur pada tahun 2013-2014 dan berusia dibawah 18 tahun dikarenakan masa pertumbuhan dan perkembangan seorang anak berada pada usia dibawah 18 tahun.

BAB II

TINJAUAN PUSTAKA

2.1 Statistik Deskriptif

Statistik deskriptif merupakan suatu metode statistik yang digunakan untuk mengumpulkan data, meringkas, menyajikan, dan mendeskripsikan data dalam bentuk yang mudah dibaca sehingga memberikan kemudahan dalam memberikan informasi.

Statistik deskriptif berguna untuk menggambarkan karakteristik unit atau obyek yang diteliti. Selain itu statistika deskriptif digunakan untuk mengetahui hubungan antara variabel respon dan variabel prediktor. Statistika deskriptif yang sering digunakan adalah mencari nilai rata-rata, deviansi standar, variansi, nilai minimum dan nilai maksimum (Walpole, 1995).

2.2 *Missing Value*

Missing value adalah hilangnya suatu nilai pada obyek (kasus) yang terjadi karena informasi pada objek tersebut sengaja tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada. Tipe dari *missing value* diantaranya adalah sebagai berikut.

1. *Missing Completely at Random* (MCAR)
Missing data terjadi secara random/acak dari sampel yang lengkap.
2. *Missing Not at Random* (MNAR)
Probabilitas dari sebuah observasi yang hilang tidak berkaitan dengan hasil observasi lain. Sehingga nilainya akan berkaitan dengan dirinya sendiri.
3. *Missing at Random* (MAR)
Probabilitas sebuah observasi dari data yang hilang biasanya berkaitan dengan informasi yang diberikan responden dengan suatu alasan untuk tidak memberikan informasi.

Dalam menangani kasus *missing value*, jika data yang hilang bersifat kategori maka dapat ditangani dengan melakukan inputasi menggunakan nilai modus. Sedangkan apabila data yang hilang bersifat kontinyu maka dapat ditangani dengan melakukan inputasi

menggunakan nilai *mean*. Jika variabel atau kasus mengandung *missing value* kurang dari 30% maka data tersebut dapat dieliminasi dan tidak diikuti dalam proses analisis (Hair, 2010).

2.3 *Synthetic Minority Oversampling Technique (SMOTE)*

Untuk mengatasi masalah *imbalanced data* dapat menggunakan metode *synthetic minority oversampling technique (SMOTE)* yakni salah satu metode *oversampling* yang ditemukan oleh Chawla (2003) berupa teknik penambahan jumlah sampel pada kelas negatif dengan melakukan replikasi data pada kelas negatif secara acak sehingga menghasilkan jumlah data yang sama dengan data pada kelas positif atau data mayor. Data yang direplikasi merupakan data yang berasal dari kelas minor. Metode yang digunakan pada algoritma SMOTE adalah *k-nearest neighbors* (ketetanggaan data) yang termasuk dalam kelompok metode statistik nonparametrik. Metode ini bekerja dengan mengelompokkan data terdekat yang dipilih berdasarkan jarak Euclidean antara kedua data. Penentuan jumlah replikasi yang dilakukan disesuaikan dengan jumlah anggota pada kelas mayor. Jumlah replikasi harus sesuai dengan jumlah k pada *nearest neighbour*, jika jumlah replikasi sebanyak n maka jumlah k sebanyak $n-1$.

Misalkan terdapat dua struktur data dengan p dimensi yaitu $x^T = [x_1, x_2, \dots, x_p]$ dan $y^T = [y_1, y_2, \dots, y_p]$ maka jarak Euclidean $d(x, y)$ yang dihasilkan antara kedua data ditunjukkan pada persamaan sebagai berikut.

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (2.1)$$

“*Synthetic*” atau replikasi data dilakukan dengan menggunakan persamaan sebagai berikut.

$$x_{syn} = x_i + (x_{knn} - x_i) \times \tau \quad (2.2)$$

dengan

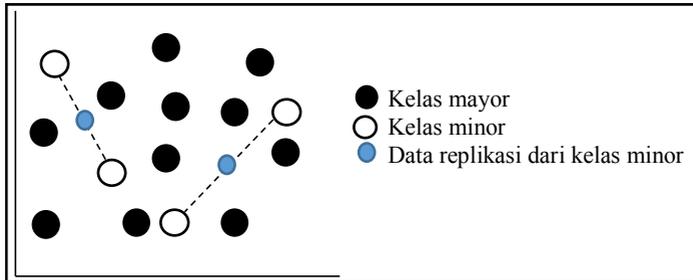
x_{syn} = data hasil replikasi.

x_i = data yang akan direplikasi

x_{knn} = data yang memiliki jarak terdekat dari data yang akan direplikasi.

τ = bilangan random 0 sampai 1

Ilustrasi mengenai algoritma *synthetic minority oversampling technique* (SMOTE) digambarkan pada Gambar 2.1 sebagai berikut.



Gambar 2. 1 Ilustrasi SMOTE

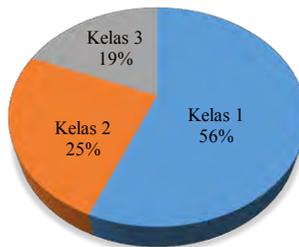
(Sumber : Chawla, 2003)

Sebagai contoh, akan dijelaskan ilustrasi mengenai metode SMOTE dalam menangani kasus *imbalanced data*. Berikut ini akan digunakan 16 data dengan jumlah variabel prediktor sebanyak 2 yaitu X_1 dan X_2 dengan 3 kelas pada variabel respon (Y) yakni kelas 1 sebanyak 9 data, kelas 2 sebanyak 4 data, dan kelas 3 sebanyak 3. Sehingga, kelas mayor dengan jumlah data terbanyak dimiliki oleh variabel respon kelas 1. Sedangkan, variabel respon kelas 2 dan kelas 3 merupakan kelas minor. Data simulasi SMOTE akan ditampilkan pada Tabel 2.1 sebagai berikut.

Tabel 2.1 Data Simulasi untuk SMOTE

No.	X_1	X_2	Y	No.	X_1	X_2	Y
1.	2	3	1	9.	2	4	1
2.	1	2	1	10.	4	6	2
3.	2	4	1	11.	5	1	2
4.	2	6	1	12.	1	3	2
5.	3	4	1	13.	2	5	2
6.	4	3	1	14.	3	6	3
7.	5	5	1	15.	4	1	3
8.	1	2	1	16.	4	7	3

Berdasarkan Tabel 2.1 maka persentase banyaknya data pada masing-masing kelas pada variabel respon akan ditunjukkan pada Gambar 2.2. Dapat diketahui bahwa persentase masing-masing kelas tidak seimbang dibuktikan dari nilai persentase yang dihasilkan, sehingga kelas mayor adalah data yang terdapat pada kelas 1 dibuktikan dari jumlah persentase terbesar yakni sebesar 53%, sedangkan persentase jumlah anggota untuk kelas 2 adalah sebesar 25%, dan persentase untuk jumlah anggota kelas 3 adalah sebesar 19%.



Gambar 2.2 Persentase Setiap Kelas

Sehingga, diperlukan penanganan untuk kasus *imbalanced data* menggunakan metode SMOTE dengan langkah-langkah sebagai berikut.

1. Setiap data pada kelas minor akan direplikasikan dengan mencari tetangga terdekat (x_{knn}) dengan menggunakan jarak Euclidean. Pada ilustrasi ini kelas minor terdapat pada kelas 2 dan kelas 3. Kemudian ingin dilakukan replikasi (x_i) pada kelas 3 dengan koordinat data ilustrasi (3,6) dengan (4,1) dan data koordinat (3,6) dengan (4,7). Perhitungannya berdasarkan persamaan (2.1) adalah sebagai berikut.

$$d\left[\begin{bmatrix} 3 \\ 6 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix}\right] = \sqrt{(3-4)^2 + (6-1)^2} = \sqrt{26}$$

$$d\left[\begin{bmatrix} 3 \\ 6 \end{bmatrix}, \begin{bmatrix} 4 \\ 7 \end{bmatrix}\right] = \sqrt{(3-4)^2 + (6-7)^2} = \sqrt{2}$$

Berdasarkan perhitungan dapat diketahui jarak terpendek adalah jarak yang dihasilkan oleh koordinat (4,7) terhadap

koordinat (3,6), sehingga x_{knn} yang terbentuk dari x_i adalah data dari koordinat (4,7).

2. Menghitung *synthetic* data pada kelas 3 sebagai berikut.

$$x_{syn} = [3, 6] + ([4, 7] - [3, 6]) \times 0,2 = [3, 6] + [0, 2; 0, 2] = [3, 2; 6, 2]$$

Data yang dihasilkan adalah data dengan koordinat [3,2;6,2] yang berada diantara x_{knn} dan x_i . Sehingga distribusi data yang dihasilkan pada kelas mayor dan minor menggunakan SMOTE adalah sebagai berikut.

Tabel 2.2 Distribusi Data dengan Menggunakan SMOTE

Sebelum direplikasi		Setelah direplikasi		Jumlah replikasi
Mayor	Minor	Mayor	Minor	
9 (56%)	4 (25%)	9 (34%)	8 (32%)	1 kali
	3(19%)		9 (34%)	2 kali

Berdasarkan Tabel 2.2, dapat diketahui jumlah data setelah menggunakan SMOTE jumlahnya bertambah yang semula berjumlah 16 data menjadi 26 data. Sedangkan jika dilihat dari persentase pada masing-masing kelas besarnya seimbang. Jumlah anggota pada kelas mayor yakni kelas 1 adalah sebanyak 9 data, sedangkan jumlah anggota pada kelas 2 sebanyak 4 data sehingga perlu dilakukan replikasi sebanyak 1 kali sehingga jumlah tetangga terdekatnya sama dengan nol. Sedangkan jumlah data pada kelas 3 sebanyak 3 data sehingga perlu dilakukan replikasi sebanyak 2 kali dan jumlah tetangga terdekatnya sama dengan satu untuk masing-masing data. Sehingga data baru setelah dilakukan replikasi menggunakan SMOTE ditunjukkan pada Tabel 2.3 sebagai berikut.

Tabel 2.3 Data Simulasi Setelah Menggunakan SMOTE

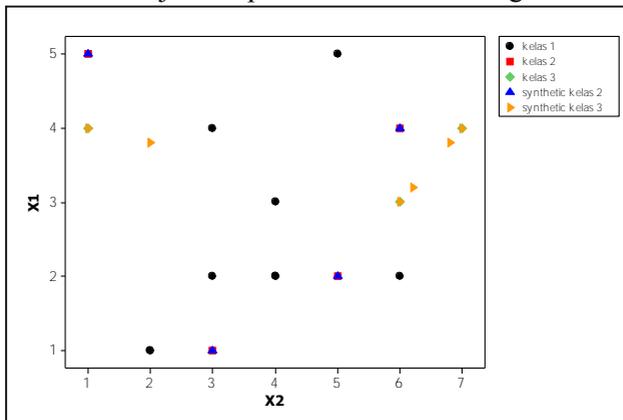
No.	X_1	X_2	Y	No.	X_1	X_2	Y
1.	2	3	1	14.	3	6	3
2.	1	2	1	15.	4	1	3
3.	2	4	1	16.	4	7	3
4.	2	6	1	17.*	4	6	2
5.	3	4	1	18.*	5	1	2
6.	4	3	1	19.*	1	3	2
7.	5	5	1	20.*	2	5	2

Tabel 2.3 Data Simulasi Setelah Menggunakan SMOTE (Lanjutan)

No.	X_1	X_2	Y	No.	X_1	X_2	Y
8.	1	2	1	21.*	3	6	3
9.	2	4	1	22.*	3,2	6,2	3
10.	4	6	2	23.*	4	1	3
11.	5	1	2	24.*	3,8	2	3
12.	1	3	2	25.*	4	7	3
13.	2	5	2	26.*	3,8	6,8	3

* *synthetic data*

Setelah dilakukan replikasi terlihat jumlah anggota kelas minor telah seimbang dengan jumlah anggota pada kelas mayor. Jumlah anggota pada kelas 2 setelah dilakukan replikasi sebanyak satu kali bertambah menjadi 8 dimana nilai x_{syn} memiliki titik koordinat yang sama dengan x_i . Untuk kelas 3, setelah dilakukan replikasi sebanyak dua kali jumlahnya menjadi 9, dimana 3 data memiliki nilai x_{syn} yang titik koordinatnya berbeda dengan titik koordinat x_i , sedangkan 3 nilai x_{syn} lainnya memiliki titik koordinat yang sama dengan titik koordinat x_i . Berdasarkan Gambar 2.3 dapat diketahui, sebaran data hasil replikasi kelas 2 yang ditunjukkan dengan pola yang berwarna biru, sedangkan data replikasi kelas 3 ditunjukkan dengan pola yang berwarna kuning. Titik koordinat anggota masing-masing kelas setelah dilakukan replikasi akan ditunjukkan pada Gambar 2.3 sebagai berikut.

**Gambar 2.3** Plot Data Setelah Replikasi dengan SMOTE

2.4 *Classification and Regression Trees (CART)*

CART adalah suatu metode pengklasifikasian yang dilakukan dengan teknik pohon keputusan (*decision trees*). Menurut Breiman, Friedman, Olshen, dan Stone (1993), CART akan menghasilkan pohon klasifikasi jika variabel respon mempunyai skala kategorik dan akan menghasilkan pohon regresi jika variabel respon berupa data kontinyu. CART memiliki istilah sifat “*binary recursive partitioning*”.

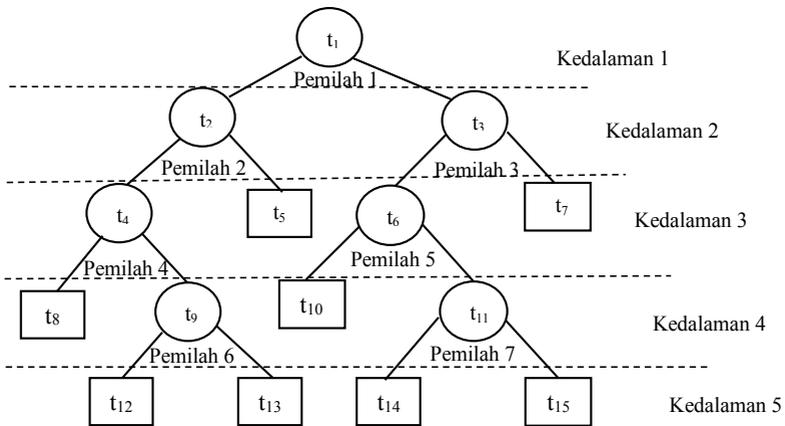
Istilah “*binary*” mengartikan suatu simpul induk (*parent node*) yang merupakan suatu variabel prediktor terpenting yang digunakan untuk memilah akan menghasilkan dua simpul anak baru (*child nodes*). Istilah “*recursive*” menjelaskan dari proses yang terjadi dalam prosedur penyekatan secara biner tersebut akan dilakukan secara berulang-ulang, artinya dari simpul anak yang sudah dihasilkan tersebut akan menjadi dua simpul induk baru dan akan dilakukan pemilahan lagi dan menghasilkan dua simpul anak yang baru. Begitu seterusnya sampai memenuhi kriteria tertentu sehingga sudah tidak dapat dipilah lagi. Sedangkan, istilah “*partitioning*” menjelaskan bahwa proses klasifikasi dilakukan dengan cara memilah suatu himpunan data menjadi bagian-bagian atau partisi-partisi (Lewis, 2000). Menurut Zheng, Chen, Han, Zhao, Ma, dan Yan (2009) data dependent tergantung dari partisi serangkaian *node* yang bercabang ke kanan dan ke kiri dapat disebut simpul anak (*child nodes*) yang berasal dari simpul utama (*parent node*). Setelah partisi telah berhenti, *child nodes* disebut sebagai *terminal nodes*.

Metode pengklasifikasian CART memiliki beberapa kelebihan diantaranya yang pertama adalah CART merupakan metode nonparametrik sehingga tidak ada asumsi distribusi variabel prediktor yang perlu dipenuhi (seperti asumsi multivariat normal dan varians homogen). Kedua, CART mampu mempertimbangkan interaksi antar variabel. Ketiga, metode *machine learning* ini memudahkan dalam hal eksplorasi dan pengambilan keputusan pada struktur data yang kompleks dan multi variabel karena struktur data dapat dilihat secara visual. Keuntungan lain adalah

hasil klasifikasi akhir berbentuk sederhana dan dapat mengklasifikasikan data baru secara lebih efisien serta mudah diinterpretasikan (Lewis, 2000). Selain beberapa kelebihan yang dimiliki, metode CART juga memiliki kelemahan yaitu pohon yang terbentuk tidak stabil artinya ketika terjadi sedikit perubahan pada data *learning* maka hasil prediksi pohon yang diperoleh dapat mengalami perbedaan yang cukup besar (Sutton, 2005).

Secara rinci struktur CART diawali dengan pembentukan simpul utama (*root node*) sebagai pemilah terpenting dalam menentukan klasifikasi kelas pengamatan yang menghimpun seluruh pengamatan dari data *learning* yang digunakan. Simpul utama tersebut sekaligus menjadi simpul induk (*parent node*) pertama yang hanya akan bisa dipilah menjadi dua simpul baru atau disebut simpul anak (*child node*) kiri dan kanan. Sehingga, setiap simpul hanya bisa dipilah menjadi dua simpul baru atau simpul anak yang mana simpul awal tersebut disebut sebagai simpul induk dari simpul anak kiri dan kanan yang baru. Proses pemilahan secara biner tersebut bisa diulang-ulang seterusnya yaitu pemilah simpul induk menjadi dua simpul anak kiri dan kanan. Kemudian masing-masing anak tersebut akan berperan sebagai simpul induk baru yang dipilah lagi menjadi dua simpul anak baru demikian seterusnya sampai terbentuk terminal node yakni simpul akhir yang sudah tidak bisa dipilah lagi, artinya simpul anak yang dihasilkan sudah homogen (Lewis, 2000).

Simpul awal yang merupakan variabel terpenting dalam menduga kelas amatan disebut sebagai *parent node* dinotasikan dengan t_1 , simpul dalam atau *internal nodes* dinotasikan dengan t_2, t_3, t_4, t_6, t_9 , dan t_{11} serta simpul akhir yang disebut sebagai *terminal nodes* dinotasikan dengan $t_5, t_7, t_8, t_{10}, t_{12}, t_{13}, t_{14}$, dan t_{15} dimana setelahnya tidak ada lagi pemilahan. Setiap simpul berada pada kedalaman (*depth*) tertentu dimana t_1 berada pada kedalaman 1, t_2 dan t_3 berada pada kedalaman 2, t_4, t_5, t_6 , dan t_7 berada pada kedalaman 3, dan begitu seterusnya hingga didapatkan simpul t_{12}, t_{13}, t_{14} , dan t_{15} yang berada pada kedalaman 5. Ilustrasi pohon klasifikasi ditunjukkan pada Gambar 2.4 sebagai berikut.



Gambar 2.4 Ilustrasi Pohon Klasifikasi

(Sumber : Breiman, Friedman, Olshen, dan Stone, 1993)

Algoritma CART secara umum melalui tiga tahapan yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi, dan penentuan pohon klasifikasi optimal.

2.4.1 Pembentukan Pohon Klasifikasi

Dalam pembentukan pohon klasifikasi dapat diawali dengan menentukan variabel dan nilai dari variabel tersebut yang disebut *threshold* untuk dijadikan pemilah bagi setiap simpul. Dalam prosesnya diperlukan data *learning* L yang terdiri atas N pengamatan dengan data X_n, J_n , dimana $x_n \in X$ dan $j_n \in \{1, \dots, J\}$, $n=1, 2, \dots, N$, dengan variabel x merupakan variabel prediktor, X merupakan ruang sampel dari x , dan j merupakan label kelas. Proses pembentukan pohon klasifikasi melalui tiga tahap analisis yakni sebagai berikut.

a. Pemilihan Pemilah

Dalam tahap pemilihan pemilah bertujuan untuk mengurangi tingkat keheterogenan pada simpul induk dan mendapatkan simpul anak dengan tingkat homogenitas yang tinggi. Data yang digunakan adalah data *learning* L yang kemudian dipilah berdasarkan aturan pemilahan menggunakan kriteria *goodness of split*.

Himpunan bagian yang dihasilkan dari proses pemilahan harus lebih homogen dibandingkan dengan simpul induknya. Hal ini dapat dilakukan dengan menganalisis fungsi keheterogenan simpul (*impurity* atau $i(t)$). Aturan pemilahan simpul induk menjadi 2 simpul anak menurut Breiman, Friedman, Olshen, dan Stone (1993) yaitu setiap pemilah hanya bergantung pada nilai yang berasal dari satu variabel prediktor. Berikut ini merupakan ketentuan jumlah pemilah untuk masing-masing jenis skala variabel.

1. Jika X_j merupakan variabel kontinyu, maka pemilahan yang diperbolehkan adalah $x_j \leq c_i$ dan $x_j > c_i$, dengan $i=1, 2, \dots, n-1$ dan c_i adalah nilai tengah atau *median* dari dua nilai amatan sampel yang berurutan yang berbeda dari variabel X_j . Jika suatu ruang sampel berukuran n dan terdapat sejumlah n pengamatan sampel yang berbeda pada variabel X_j , maka terdapat $n-1$ kemungkinan pemilahan yang berbeda.
2. Jika X_j merupakan variabel kategorik, maka pemilahan berasal dari semua kemungkinan pemilahan berdasarkan terbentuknya dua simpul yang saling lepas (*disjoint*). Bila kategori berskala nominal bertaraf L , maka akan diperoleh sebanyak $2^{L-1}-1$ pemilahan yang mungkin terjadi. Bila kategori berskala ordinal bertaraf L , maka akan diperoleh sebanyak $L-1$ pemilahan.

Menurut Breiman, Friedman, Olshen, dan Stone (1993) fungsi heterogenitas yang umum digunakan adalah indeks gini. Kelebihan dari indeks gini adalah metode ini melakukan proses perhitungan dengan sederhana, relatif cepat, dan sesuai untuk diterapkan dalam berbagai kasus. Pemilah terbaik dipilih dari semua kemungkinan pemilahan pada setiap variabel prediktor berdasarkan pada nilai penurunan keheterogenan tertinggi. Fungsi indeks gini dituliskan dalam persamaan sebagai berikut.

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j \quad (2.3)$$

dengan

$i(t)$ = *impurity* atau fungsi keheterogenan simpul t

$p(j|t)$ = proporsi kelas j pada simpul t

$p(i|t)$ = proporsi kelas i pada simpul t

Selain indeks gini terdapat ukuran lain yang dapat digunakan untuk mendefinisikan fungsi keheterogenan diantaranya adalah indeks informasi, indeks entropy, dan indeks twoing. Namun menurut Breiman, Friedman, Olshen, dan Stone (1993) fungsi keheterogenan yang paling mudah untuk diterapkan selain indeks gini adalah indeks twoing, dimana formula yang digunakan untuk indeks twoing adalah sebagai berikut.

$$i(t) = \frac{P_L P_R}{4} [\sum p(j|t_L) - p(j|t_R)]^2 \quad (2.4)$$

dengan

$i(t)$ = *impurity* atau fungsi keheterogenan simpul t

P_L = proporsi pengamatan pada simpul kiri

P_R = proporsi pengamatan pada simpul kanan

$p(j|t_L)$ = proporsi pengamatan dari simpul t menuju simpul kiri dengan kelas j

$p(j|t_R)$ = proporsi pengamatan dari simpul t menuju simpul kanan dengan kelas j

Kemudian pemilihan pemilah terbaik dilakukan berdasarkan kriteria *goodness of split* yang merupakan suatu nilai yang dijadikan patokan pemilahan oleh suatu pemilah s pada suatu simpul t . *Goodness of split* menunjukkan ukuran penurunan keheterogenan suatu kelas dengan formula sebagai berikut.

$$\phi(s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.5)$$

dengan

$\phi(s, t)$ = nilai *goodness of split*

$i(t)$ = fungsi heterogenitas pada simpul t

P_L = proporsi pengamatan simpul kiri

P_R = proporsi pengamatan simpul kanan

$i(t_L)$ = fungsi heterogenitas pada simpul anak kiri

$i(t_R)$ = fungsi heterogenitas pada simpul anak kanan.

$$\Delta i(s^*, t_1) = \underset{s \in S}{maks} \Delta i(s, t) \quad (2.6)$$

Pemilah yang menghasilkan nilai *goodness of split* ($\phi(s, t)$) yang tertinggi merupakan pemilah terbaik karena mampu mereduksi heterogenitas lebih tinggi. Pemilah terbaik tersebut dapat menunjukkan *variable important* yang akan menghasilkan

skor variabel untuk menunjukkan seberapa besar variabel tersebut memberikan kontribusi dalam proses pembentukan pohon. Berikut ini persamaan untuk menentukan skor pada *variable important*.

$$\text{skor} = \sum_{i=1}^n \phi(s, t_i) \quad (2.7)$$

$\phi(s, t_i)$ = nilai *goodness of split* pada setiap simpul (*improvement*)

Nilai skor diperoleh dengan menjumlahkan *goodnes of split (improvement)* dari masing-masing variabel yang berperan sebagai *surrogates* untuk setiap simpul (CART *References Guide*, 2000).

b. Penentuan *Terminal Node*

Tahap yang kedua adalah penentuan terminal node dimana simpul t akan diputuskan untuk dipilah lagi atau dapat dijadikan *terminal node* jika tidak terdapat penurunan heterogenitas yang berarti. Berikut ini merupakan kriteria untuk memutuskan suatu simpul tidak akan dipilah lagi atau menjadi simpul terminal.

1. Hanya terdapat satu pengamatan ($n=1$) pada setiap simpul anak atau adanya batasan minimum n . Menurut Breiman, Friedman, Olshen, dan Stone (1993) pengembangan pohon akan berhenti bila pada simpul terdapat pengamatan berjumlah kurang dari 5 ($n_i < 5$).
2. Proses pembentukan pohon juga dapat berhenti bila sudah mencapai batasan jumlah level yang telah ditentukan atau tingkat kedalaman (*depth*) dalam pohon maksimal.

c. Penandaan Label Kelas

Dalam penandaan label kelas pada *terminal nodes* dperlu diberikan label kelas sehingga nantinya dapat diketahui karakteristik dari klasifikasi pengamatan untuk setiap kelas variabel respon yang terbentuk. Pemberian label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yaitu apabila,

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.8)$$

dengan

$p(j|t)$ = proporsi kelas j pada simpul t

$N_j(t)$ = banyaknya amatan kelas j pada *terminal node* t

$N(t)$ = jumlah total pengamatan dalam *terminal node* t
 maka label kelas untuk *terminal node* t adalah j_0 . Asumsi yang digunakan adalah biaya kesalahan klasifikasi untuk setiap kelas besarnya sama, pemberian label kelas j_0 tersebut memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t paling kecil yaitu sebesar $r(t) = 1 - \max_j p(j|t)$.

2.4.2 Pemangkasan Pohon Klasifikasi

Pohon klasifikasi yang terbentuk pertama kali adalah pohon klasifikasi maksimal yakni dapat diindikasikan pohon yang terbentuk dari hasil pemilahan bisa berukuran sangat besar, karena aturan penghentian pohon hanya berdasarkan pada jumlah pengamatan pada simpul akhir yang biasanya ditetapkan sebanyak satu pengamatan saja. Pembentukan pohon klasifikasi yang besar dapat menimbulkan dugaan adanya kasus *overfitting* yaitu nilai prediksi melebihi nilai yang sebenarnya. Sehingga, banyaknya pemilahan yang dilakukan tersebut dapat mengakibatkan semakin kecilnya tingkat kesalahan dalam prediksi yang dilakukan, sebab hal ini memungkinkan simpul terminal hanya beranggotakan satu pengamatan.

Sedangkan menurut Breiman, Friedman, Olshen, dan Stone (1993) bila dalam proses pemilahan tersebut diberikan batasan, padahal pemilahan masih layak untuk dilakukan maka akan terjadi kasus *underfitting*. Oleh karena itu, untuk mendapatkan pohon yang layak, perlu dilakukan pemangkasan pohon (*pruning*) yaitu suatu penilaian ukuran pohon tanpa mengorbankan akurasi yang berarti yang dilakukan melalui pengurangan simpul pohon sehingga dicapai ukuran pohon yang layak dan tidak terlalu melebar. Ukuran pohon yang layak dapat dilakukan pemangkasan pohon dengan ukuran *cost complexity*. Metode ini mengandalkan parameter *complexity* yang dinotasikan dengan α , yang secara perlahan nilainya meningkat selama proses pemangkasan.

Suatu simpul anak akan dipangkas jika hasil perubahan nilai kesalahan klasifikasi prediksi lebih kecil daripada α kali perubahan kompleksitas pohon. Pada sembarang pohon T yang merupakan

sub pohon dari pohon klasifikasi maksimal (T_{max}) sehingga $T < T_{max}$, dengan nilai $\alpha \geq 0$, maka persamaan fungsi *cost complexity* adalah sebagai berikut.

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \quad (2.9)$$

dengan

$R_\alpha(T)$ = *cost complexity measure* (ukuran kompleksitas suatu pohon T pada kompleksitas α)

$R(T)$ = penduga pengganti (*resubstitution estimate*) pohon atau ukuran kesalahan klasifikasi pohon T

α = *complexity parameter* (*cost* bagi penambahan satu simpul akhir pada pohon T)

$|\tilde{T}|$ = ukuran banyaknya *terminal node* pada pohon T .

$R_\alpha(T)$ merupakan kombinasi linear dari harga dan nilai kompleksitas pohon yang dibentuk. *Cost complexity pruning* dapat menentukan suatu pohon bagian $T(\alpha)$ yang meminimumkan $R_\alpha(T)$ pada seluruh pohon bagian atau untuk setiap nilai α . Selanjutnya, dilakukan pencarian pohon bagian $T(\alpha) < T_{max}$ yang dapat meminimumkan $R_\alpha(T)$ dengan persamaan sebagai berikut.

$$R_\alpha(T(\alpha)) = \min_{T < T_{max}} R_\alpha(T) \quad (2.10)$$

Tahapan dalam pemangkasan pohon adalah dimulai dengan mengambil t_R dan t_L dari T_{max} yang dihasilkan dari simpul induk t . Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan $R(t) = R(t_R) + R(t_L)$, maka simpul anak t_R dan t_L dipangkas. Sehingga, diperoleh pohon T_1 yang memenuhi kriteria $R(T_1) = R(T_{max})$. Proses ini diulang hingga tidak mungkin lagi dilakukan pemangkasan.

Hasil yang diperoleh adalah suatu barisan menurun dan tersarang dari pohon bagian yaitu $T_1 > T_2 > \dots > \{t_1\}$ dengan $T_1 < T_{max}$ dan suatu deret yang semakin meningkat dari α_k yaitu $\alpha_1 < \alpha_2 < \dots < \alpha_k$ dimana $\alpha_1 = 0$ dan $T(\alpha) = T(\alpha_k) = T_k$. Jika digunakan $R(T)$ sebagai kriteria penentuan pohon optimal, maka akan cenderung dipilih pohon terbesar T_1 , sebab semakin besar pohon maka nilai $R(T)$ semakin kecil.

2.4.3 Penentuan Pohon Klasifikasi Optimal

Dalam penentuan pohon klasifikasi yang optimal ukuran pohon yang terlalu besar akan menyebabkan nilai *cost complexity* yang tinggi karena struktur data yang digambarkan cenderung kompleks sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti yang cukup kecil. Bila $R(T)$ dipilih sebagai penduga terbaik, maka akan cenderung dipilih pohon yang besar, sebab pohon yang semakin besar akan membuat nilai $R(T)$ semakin kecil. Terdapat dua macam penduga untuk mendapatkan pohon klasifikasi optimal yaitu *test sample estimate* dan *cross validation V-fold estimate*.

a. *Test Sample Estimate*

Penduga sampel uji digunakan ketika data berukuran besar. Prosedur *test sample estimate* diawali dengan membagi data menjadi 2 bagian yakni data *learning* (L_1) dan data *testing* (L_2). Pengamatan dalam L_1 digunakan untuk membentuk pohon T , sedangkan pengamatan dalam L_2 digunakan untuk menduga $R^s(T)$. Persamaan *test sample estimate* adalah sebagai berikut.

$$R^s(T_i) = \frac{1}{N_2} \sum_{(x_n, j_n) \in L_2} X(d(x_n) \neq j_n) \quad (2.11)$$

dengan

$R^{ts}(T_i)$ = total proporsi kesalahan *test sample estimate*

N_2 = jumlah pengamatan dari data *learning* (L_2)

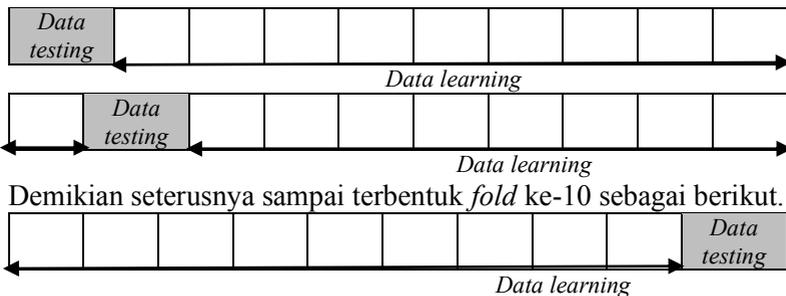
Sedangkan $X(\cdot)$ merupakan pertanyaan yang akan bernilai 0 jika pertanyaan dalam tanda kurung salah dan bernilai 1 jika pertanyaan dalam tanda kurung benar. Dikarenakan dalam hal ini ingin menduga proporsi kesalahan yang dihasilkan dari proses pembentukan pohon klasifikasi sehingga, pohon klasifikasi optimal yang dipilih adalah T_i dengan $R^{ts}(T_i) = \min_i R^{ts}(T_i)$.

b. *Cross Validation V-Fold Estimate*

Penduga pengganti ini sering digunakan apabila data pengamatan berukuran kecil. Pengamatan dalam L dibagi secara random menjadi V bagian (*fold*) yang saling lepas dengan berukuran kurang lebih sama besar untuk setiap kelas. Salah satu

bagian (*fold*) akan berperan sebagai data *testing*, sedangkan *fold* yang lain akan berperan menjadi data *learning* dalam prosedur pembentukan pohon. Kemudian seluruh prosedur pembentukan pohon akan dilakukan pembentukan model sebanyak V kali, dengan *fold* yang berbeda dari data setiap kali melakukan pembentukan pohon kemudian dihitung rata-rata akurasi dari 10 bagian model yang sudah dibentuk (Lewis, 2000).

Metode *cross validation* bisa dilakukan untuk menghindari adanya kasus *overlapping* pada data testing. Nilai V yang sering digunakan dan dijadikan standar adalah 10 sehingga menjadi *10-fold cross validation* dikarenakan nilai 10 merupakan nilai terbaik untuk mendapatkan estimasi *error* dari proses klasifikasi yang sudah dilakukan. Selain itu dengan membagi data menjadi 10 bagian akan menghasilkan proporsi data yang seimbang atau *balance*. Sehingga nilai akurasi yang dihasilkan dengan proporsi data yang *balance* akan lebih akurat (Witten, 2011). Oleh karena itu metode *10-fold cross validation* dijadikan sebagai standar yang digunakan untuk proses klasifikasi dalam membagi data learning dan testing. Jika menggunakan metode *10-fold cross validation* akan dihasilkan 10 nilai akurasi dari data *learning* dan data *testing* kemudian 10 nilai akurasi tersebut dirata-rata. Berikut ini merupakan ilustrasi prosedur pembagian data dengan *10-fold cross validation* untuk setiap bagian (*fold*) berdasarkan data *learning* dan data *testing* yang ditunjukkan pada Gambar 2.5 sebagai berikut.



Gambar 2.5 Ilustrasi Prosedur *Cross Validation 10-Fold*

(Sumber : Witten, 2011)

Misalkan $d^{(v)}(x)$ adalah hasil pengklasifikasian, maka penduga sampel uji untuk $R(T_i^{(v)})$ adalah sebagai berikut:

$$R(T_i^{(v)}) = \frac{1}{N_v} \sum_{(x_n, j_n) \in L_v} X(d^{(v)}(x_n) \neq j_n) \quad (2.12)$$

dimana $N_v \cong N/V$ adalah jumlah pengamatan dalam L_v . Selanjutnya, dilakukan prosedur yang sama dengan menggunakan semua pengamatan dalam L untuk membentuk deret pohon T_i . Penduga validasi silang lipat V untuk $T_i^{(v)}$ adalah

$$R^{cv}(T_i) = \frac{1}{V} \sum_{v=1}^V R^{cv}(T_i^{(v)}) \quad (2.13)$$

Sehingga, pohon klasifikasi yang optimum dipilih T^* dengan $R^{cv}(T^*) = \min R^{cv}(T_i)$.

2.4.4 Ilustrasi CART dengan Pemilihan Pemilah Indeks Gini

Algoritma CART dengan pemilihan pemilah yang pertama yakni menggunakan indeks gini dalam pembentukan pohon klasifikasi dapat diilustrasikan dengan menggunakan 2 variabel prediktor yakni umur penderita (tahun) dan faktor genetik. Kemudian satu variabel respon yang terdiri dari 3 kelas yakni kelas 1 adalah penderita retardasi mental tipe ringan, kelas 2 adalah penderita retardasi mental tipe sedang, dan kelas 3 adalah penderita retardasi mental tipe berat. Variabel yang digunakan dalam ilustrasi CART dijelaskan dalam Tabel 2.4 sebagai berikut.

Tabel 2.4 Variabel Penelitian Ilustrasi CART

Variabel	Nama Variabel	Kategori	Skala Pengukuran
Y	Tipe Retardasi Mental	1: ringan 2: sedang 3: berat	Ordinal
X ₁	Umur Penderita (Tahun)	-	Rasio
X ₂	Faktor Genetik	1: ada keturunan 2: tidak ada	Nominal

Pada ilustrasi CART menggunakan data bangkitan sebanyak 10 sampel pengamatan. Data sampel yang digunakan untuk ilustrasi pembentukan pohon klasifikasi CART ditampilkan dalam Tabel 2.5 sebagai berikut.

Tabel 2.5 Data Sampel Ilustrasi CART

Data Sampel ke-	X_1	X_2	Y
1	5	1	3
2	5	2	2
3	6	2	2
4	10	1	1
5	12	1	3
6	6	1	3
7	10	2	2
8	10	2	1
9	12	2	1
10	6	2	1

Data sampel pengamatan tersebut akan membentuk simpul induk (*parent nodes*) yang kemudian akan memilah menjadi simpul anak (*child nodes*). Algoritma CART yang pertama adalah pembentukan pohon klasifikasi, dengan pemilihan pemilah menggunakan indeks gini. Sebelum melakukan proses pemilihan pemilah dalam analisis CART, terlebih dahulu menghitung nilai keheterogenan simpul induk berdasarkan fungsi keheterogenan indeks gini yang telah didefinisikan pada persamaan (2.3). Pada saat memilah simpul induk menjadi dua simpul anak, dilakukan penentuan pemilah dari semua kemungkinan variabel dan nilai variabel (*threshold*). Sebelumnya mencari nilai *goodness of split* menggunakan persamaan (2.5) dari kemungkinan *threshold* pada setiap variabel yang dihasilkan kemudian dilakukan perbandingan dan memilih nilai *threshold* pada variabel yang memiliki *goodness of split* yang paling tinggi berdasarkan persamaan (2.6). *Threshold* variabel terpilih tersebut selanjutnya dijadikan sebagai pemilah

terbaik untuk simpul awal. Berikut ini merupakan contoh perhitungan manual dari proses pemilihan pemilah terbaik.

a. Perhitungan Indeks Gini untuk Variabel Y

Tabel 2.6 Proporsi Pengamatan pada Respon Y

Variabel respon (Y)	$N_i(t)$	$p(i t)$
Kelas 1	4	$4/10 = 0,4$
Kelas 2	3	$3/10 = 0,3$
Kelas 3	3	$3/10 = 0,3$
Total	10	1

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j$$

$$i(t) = [p(1|t)p(2|t)] + [p(1|t)p(3|t)] + [p(2|t)p(3|t)]$$

$$i(t) = (0,4)(0,3) + (0,4)(0,3) + (0,3)(0,3) = 0,33$$

b. Penentuan *Threshold* dan Perhitungan *Goodness of Split* Variabel X_1 (Umur Penderita)

Variabel X_1 memiliki skala pengukuran kontinyu, sehingga *threshold* pada variabel X_1 berupa nilai tengah (median) dari dua nilai amatan sampel berurutan yang berbeda dari variabel X_1 . Diperoleh kemungkinan nilai median sebagai berikut.



Nilai median = 5,5 8 11

Threshold untuk variabel X_1 dengan nilai median 5,5 maka untuk data $X_1 \leq 5,5$ akan dipilah ke simpul kiri sedangkan data $X_1 > 5,5$ akan dipilah ke simpul kanan. Perhitungannya pada Tabel 2.7 sebagai berikut.

Tabel 2.7 Perhitungan *Goodness of Split* untuk *Threshold* $X_1=5,5$

Simpul kiri		Simpul kanan	
$N_L = 2$	$p_L = 0,2$	$N_R = 8$	$p_R = 0,8$
$N(1 t_L) = 0$	$p(1 t_L) = 0$	$N(1 t_R) = 4$	$p(1 t_R) = 4/8$
$N(2 t_L) = 1$	$p(2 t_L) = 1/2$	$N(2 t_R) = 2$	$p(2 t_R) = 2/8$
$N(3 t_L) = 1$	$p(3 t_L) = 1/2$	$N(3 t_R) = 2$	$p(3 t_R) = 2/8$

Ket: L=left (kiri); R=right (kanan)

Simpul kiri : $i(t_L)=[(0)(1/2)]+[(0)(1/2)]+[(1/2)(1/2)]=1/4$

Simpul kanan : $i(t_R)=[(4/8)(2/8)]+[(4/8)(2/8)]+[(2/8)(2/8)]=5/16$

Maka, perhitungan *goodness of split* untuk *threshold* $X_1=5,5$ adalah sebagai berikut.

$$\phi(s,t)=0,33-0,2(1/4)-0,8(5/16)=0,03$$

Pada *threshold* dengan nilai median 8 pada data $X_1 \leq 8$ akan dipilah ke simpul kiri sedangkan data $X_1 > 8$ akan dipilah ke simpul kanan. Perhitungannya akan ditampilkan pada Tabel 2.8 sebagai berikut.

Tabel 2.8 Perhitungan *Goodness of Split* untuk *Threshold* $X_1=8$

Simpul kiri		Simpul kanan	
$N_L = 5$	$p_L = 0,5$	$N_R = 5$	$p_R = 0,5$
$N(1 t_L)=1$	$p(1 t_L)=1/5$	$N(1 t_R)=3$	$p(1 t_R)=3/5$
$N(2 t_L)=2$	$p(2 t_L)=2/5$	$N(2 t_R)=1$	$p(2 t_R)=1/5$
$N(3 t_L)=2$	$p(3 t_L)=2/5$	$N(3 t_R)=1$	$p(3 t_R)=1/5$

Simpul kiri : $i(t_L)=[(1/5)(2/5)]+[(1/5)(2/5)]+[(2/5)(2/5)]=8/25$

Simpul kanan : $i(t_R)=[(3/5)(1/5)]+[(3/5)(1/5)]+[(1/5)(1/5)]=7/25$

Maka, perhitungan *goodness of split* untuk *threshold* $X_1=8$ adalah sebagai berikut.

$$\phi(s,t)=0,33-0,5(8/25)-0,5(7/25)=0,03$$

Nilai *threshold* dengan nilai median 11 pada variabel $X_1 \leq 11$ akan dipilah ke simpul kiri sedangkan untuk data $X_1 > 11$ akan dipilah ke simpul kanan. Perhitungannya akan ditampilkan pada Tabel 2.9 berikut ini.

Tabel 2.9 Perhitungan *Goodness of Split* untuk *Threshold* $X_1=11$

Simpul kiri		Simpul kanan	
$N_L = 8$	$p_L = 0,8$	$N_R = 2$	$p_R = 0,2$
$N(1 t_L)=3$	$p(1 t_L)=3/8$	$N(1 t_R)=1$	$p(1 t_R)=1/2$
$N(2 t_L)=3$	$p(2 t_L)=3/8$	$N(2 t_R)=0$	$p(2 t_R)=0$
$N(3 t_L)=2$	$p(3 t_L)=2/8$	$N(3 t_R)=1$	$p(3 t_R)=1/2$

Simpul kiri : $i(t_L)=[(3/8)(3/8)]+[(3/8)(2/8)]+[(3/8)(2/8)]=21/64$

Simpul kanan : $i(t_R)=[(1/2)(0)]+[(1/2)(1/2)]+[(0)(1/2)]=1/4$

Maka, perhitungan *goodness of split* untuk *threshold* $X_1=11$ adalah sebagai berikut.

$$\phi(s,t)=0,33-0,8(21/64)-0,2(1/4)=0,0175$$

c. Penentuan *Threshold* dan Perhitungan *Goodness of Split* Variabel X_2 (Faktor Genetik)

Variabel X_2 merupakan variabel yang bersifat kategorik dengan jumlah kategori sebanyak dua, sehingga banyaknya kemungkinan *threshold* adalah sebanyak satu kemungkinan. Hasil tersebut diperoleh dari perhitungan sebagai berikut.

$$2^{L-1}-1=2^{2-1}-1=1 \text{ kemungkinan pemilah.}$$

Oleh karena itu, data pada variabel X_2 yang memiliki kategori 1 dipilah sebagai simpul kiri, sedangkan data yang memiliki kategori 2 dipilah sebagai simpul kanan. Perhitungan nilai *goodness of split* untuk *threshold* variabel X_2 ditunjukkan pada Tabel 2.10 sebagai berikut.

Tabel 2.10 Perhitungan *Goodness of Split* untuk *Threshold* $X_2=1:2$

Simpul kiri		Simpul kanan	
$N_L = 4$	$p_L = 0,4$	$N_R = 6$	$p_R = 0,6$
$N(1 t_L)=1$	$p(1 t_L)=1/4$	$N(1 t_R)=3$	$p(1 t_R)=3/6$
$N(2 t_L)=0$	$p(2 t_L)=0$	$N(2 t_R)=3$	$p(2 t_R)=3/6$
$N(3 t_L)=3$	$p(3 t_L)=3/4$	$N(3 t_R)=0$	$p(3 t_R)=0$

Simpul kiri : $i(t_L)=[(1/4)(0)]+[(1/4)(3/4)]+[(0)(3/4)]=3/16$

Simpul kanan : $i(t_R)=[(3/6)(3/6)]+[(3/6)(0)]+[(3/6)(0)]=9/36$

Maka, perhitungan *goodness of split* untuk *threshold* X_2 1:2 adalah sebagai berikut.

$$\phi(s,t)=0,33-0,4(3/16)-0,6(9/36)=0,105$$

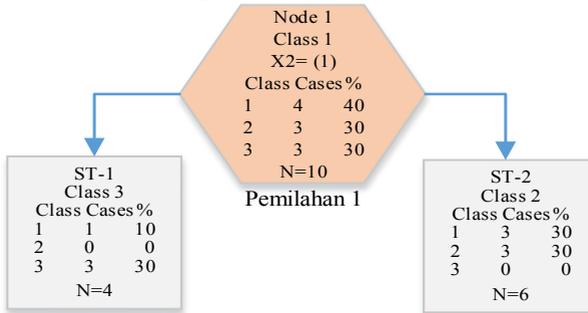
Setelah dilakukan perhitungan nilai *goodness of split* selanjutnya dilakukan perbandingan nilai *goodness of split* dari variabel X_1 dan X_2 untuk menentukan pemilah terbaik

Tabel 2.11 Perbandingan *Goodness of Split* dari Pemilah 1

Variabel	<i>Threshold</i>	<i>Goodness of Split</i> ($\phi(s,t)$)
	5,5	0,03
X_1	8	0,03
	11	0,0175
X_2	1:2	0,105*

**goodness of split* terbesar

Berdasarkan Tabel 2.11, dapat diketahui pemilah terbaik dipilih dari *threshold* dengan *goodness of split* terbesar. Nilai *goodness of split* terbesar terdapat pada variabel pemilah X_2 (faktor genetik) dengan *threshold* 1:2. Selanjutnya, dilakukan pemilahan pada data sampel berdasarkan pemilah terbaik atau pemilah utama tersebut. Sehingga data sampel pada variabel X_2 yang memiliki kategori 1 masuk ke dalam simpul kiri sedangkan data yang memiliki kategori 2 masuk dalam simpul kanan. Berikut ini merupakan hasil pembentukan pohon berdasarkan pemilahan utama yang ditunjukkan pada Gambar 2.6 sebagai berikut.



Keterangan: S1: simpul 1 (simpul induk), S2: simpul 2, S3: simpul 3

Gambar 2.6 Pohon Klasifikasi Ilustrasi Hasil Pemilahan I

Berdasarkan Gambar 2.6, dapat diketahui simpul kanan yang dihasilkan terdiri dari 5 data dan belum homogen, untuk itu dilakukan pemilahan kembali secara *recursive* hingga didapatkan pohon klasifikasi maksimal dengan simpul yang homogen. Data sampel yang berasal dari simpul kanan digunakan untuk pemilahan kedua yang ditunjukkan pada Tabel 2.12 sebagai berikut.

Tabel 2.12 Data Sampel Ilustrasi CART Pemilahan 2

Data Sampel ke-	X_1	X_2	Y
2	5	2	2
3	6	2	2
7	10	2	2
8	10	2	1
9	12	2	1
10	6	2	1

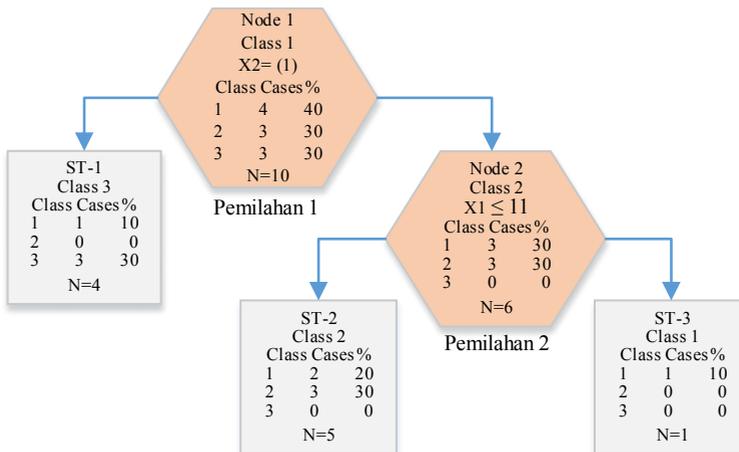
Hasil perbandingan *goodness of split* setiap kemungkinan *threshold* dari sampel pada simpul kanan yang digunakan sebagai pemilah kedua ditunjukkan dalam Tabel 2.13 sebagai berikut.

Tabel 2.13 Perbandingan *Goodness of Split* dari Pemilah 2

Variabel	Threshold	<i>Goodness of Split</i> ($\Phi(s,t)$)
X ₁	5,5	0,05
	8	0,03
	11	0,0508*
X ₂	1:2	0

**goodness of split* terbesar

Berdasarkan Tabel 2.13 dapat diketahui nilai *goodness of split* terbesar yakni 0,0508 diperoleh pada variabel X₁ dengan *threshold* sebesar 11, sehingga pemilah terbaik yang dihasilkan adalah variabel umur penderita dan berperan sebagai pemilahan II. Oleh karena itu, data sampel berjumlah 6 data akan dipilah berdasarkan pemilahan 2 menjadi dua anak simpul, dimana data variabel X₁ yang bernilai kurang dari 11 akan masuk ke dalam simpul kiri dan sampel dengan nilai lebih dari 11 akan masuk ke simpul kanan. Sehingga, hasil pemilahan untuk simpul kanan pada kedalaman 2 dapat digambarkan pada pohon klasifikasi berikut ini.



Gambar 2.7 Pohon Klasifikasi Hasil Pemilahan II

Berdasarkan Breiman, Friedman, Olshen, dan Stone (1993) pembentukan pohon akan berhenti ketika jumlah anggota pada simpul terminal berjumlah kurang dari sama dengan 5 pengamatan atau dengan jumlah *nodes* minimum yang telah ditentukan. Maka pohon yang terbentuk pada Gambar 2.7 dapat dikatakan sebagai pohon maksimal dan simpul terminal yang terbentuk adalah sebanyak 3 ditunjukkan oleh kotak yang berwarna kuning.

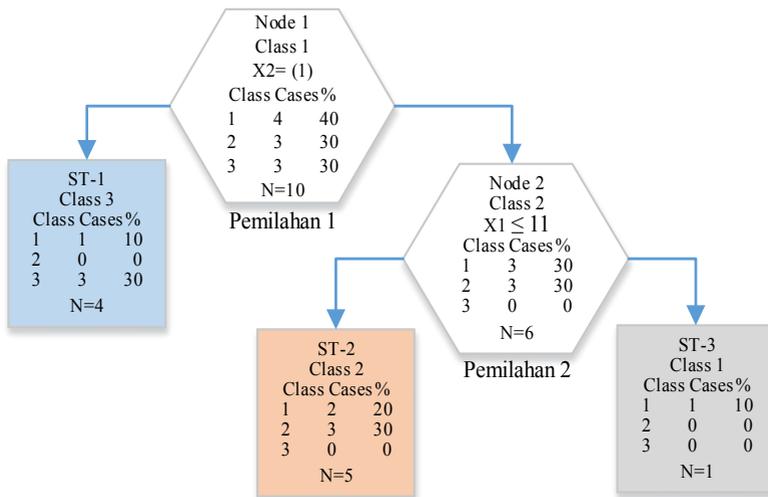
Selanjutnya dilakukan pelabelan kelas pada tiap *terminal nodes* untuk mengetahui karakteristik masing-masing kelas pada variabel respon. Sehingga dapat memudahkan dalam hal pengklasifikasian dengan cara melihat karakteristik yang telah dihasilkan. Pemberian label kelas yaitu label kelas 1 untuk penderita retardasi mental tipe ringan, label kelas 2 untuk penderita retardasi mental tipe sedang, dan label kelas 3 untuk penderita retardasi mental tipe berat yang ditentukan berdasarkan persamaan (2.7) dengan mencari nilai peluang kelas $j=1$ yang tertinggi. Penandaan label pada simpul terminal dapat dilihat pada Tabel 2.14 berikut.

Tabel 2.14 Pemberian Label pada Simpul Terminal 1

Simpul	Kelas	$N(j t)$	$p(j t)$	Label kelas
S2	1	1	25%	RM Berat
	2	0	0	
	3	3	75%	
Total		4	100%	
S4	1	2	40%	RM Sedang
	2	3	60%	
	3	0	0	
Total		5		
S5	1	1	100%	RM Ringan
	2	0	0	
	3	0	0	
Total		1	100%	

Berdasarkan Tabel 2.14, dapat diketahui label kelas untuk masing-masing simpul terminal berdasarkan proporsi tertinggi pada masing-masing kelas. Pada simpul terminal 2 dapat diketahui dari 4 data sampel pengamatan terdapat 1 data termasuk kelas 1

dan 3 data sampel pengamatan yang termasuk kelas 3, karena simpul terminal 2 banyak didominasi oleh data kelas 3 sehingga simpul terminal 2 dapat diberi label kelas 3 yang ditunjukkan oleh kotak berwarna biru. Sedangkan simpul terminal 4 diberi label kelas 2 yang ditunjukkan oleh kotak berwarna merah muda, dan simpul terminal 5 diberi label kelas 3 yang ditunjukkan oleh kotak berwarna abu-abu pada Gambar 2.8. Berdasarkan 10 data sampel pengamatan dapat diketahui sebanyak 7 data yang tepat diklasifikasikan, sedangkan sebanyak 3 data mengalami kesalahan pengklasifikasian. Proses tersebut dapat dilanjutkan dengan melakukan pemangkasan pohon (*pruning*) apabila pohon klasifikasi yang terbentuk terlalu kompleks dengan menggunakan persamaan (2.8) untuk mendapatkan pohon klasifikasi yang optimal.



Gambar 2.8 Pohon Klasifikasi dengan Label Kelas

2.4.5 Ilustrasi CART dengan Pemilihan Pemilahan Indeks Twoing

Pembentukan pohon klasifikasi pada ilustrasi CART yang kedua adalah dengan menggunakan metode pemilahan pemilahan

menggunakan indeks twoing. Data sampel pengamatan yang digunakan terdapat pada Tabel 2.4 dan Tabel 2.5. Berikut ini merupakan perhitungan indeks twoing berdasarkan variabel prediktor umur penderita (X_1) dan faktor genetik (X_2).

a. Nilai *Impurity* untuk Variabel Umur Penderita (X_1)

Variabel X_1 (umur penderita) merupakan variabel dengan skala kontinyu. Nilai *threshold* dapat dicari dengan menggunakan nilai median dari satu sampel pengamatan yang digunakan. Setelah diurutkan didapatkan 4 data sampel yang berbeda sehingga akan diperoleh 3 nilai median berdasarkan perhitungan dengan menggunakan indeks gini sebelumnya.

Threshold untuk variabel X_1 dengan nilai median yang pertama yakni sebesar 5,5, sehingga untuk data $X_1 \leq 5,5$ akan dipilah ke simpul kiri sedangkan data $X_1 > 5,5$ akan dipilah ke simpul kanan. Perhitungan proporsi yang dihasilkan oleh simpul kiri dan simpul kanan ditampilkan pada Tabel 2.15 berikut ini.

Tabel 2.15 Proporsi Pengamatan Simpul Kiri dan Simpul Kanan dengan *Threshold* $X_1=5,5$

Simpul kiri		Simpul kanan	
$N_L = 2$	$p_L = 0,2$	$N_R = 8$	$p_R = 0,8$
$N(1 t_L)=0$	$p(1 t_L)=0$	$N(1 t_R)=4$	$p(1 t_R)=4/8$
$N(2 t_L)=1$	$p(2 t_L)=1/2$	$N(2 t_R)=2$	$p(2 t_R)=2/8$
$N(3 t_L)=1$	$p(3 t_L)=1/2$	$N(3 t_R)=2$	$p(3 t_R)=2/8$

Ket: L=left (kiri); R=right (kanan)

$$i(t) = \frac{P_L P_R}{4} \left[\sum_j |p(j|t_L) - p(j|t_R)| \right]^2$$

$$i(t) = \frac{(0,2)(0,8)}{4} [|(0 - 4/8)| + |(1/2 - 2/8)| + |(1/2 - 2/8)|]^2$$

$$i(t) = 0,04(0,5 + 0,25 + 0,25)^2$$

$$i(t) = 0,04$$

Nilai median yang kedua sebagai *threshold* adalah sebesar 8, sehingga untuk data $X_1 \leq 8$ akan dipilah ke simpul kiri sedangkan data $X_1 > 8$ akan dipilah ke simpul kanan. Perhitungannya pada Tabel 2.16 berikut ini.

Tabel 2.16 Proporsi Pengamatan Simpul Kiri dan Simpul Kanan dengan *Threshold* $X_1=8$

Simpul kiri		Simpul kanan	
$N_L = 5$	$p_L = 0,5$	$N_R = 5$	$p_R = 0,5$
$N(1 t_L)=1$	$p(1 t_L)=1/5$	$N(1 t_R)=3$	$p(1 t_R)=3/5$
$N(2 t_L)=2$	$p(2 t_L)=2/5$	$N(2 t_R)=1$	$p(2 t_R)=1/5$
$N(3 t_L)=2$	$p(3 t_L)=2/5$	$N(3 t_R)=1$	$p(3 t_R)=1/5$

$$i(t) = \frac{(0,5)(0,5)}{4} [|(1/5-3/5)| + |(2/5-1/5)| + |(2/5-1/5)|]^2$$

$$i(t) = 0,0625(0,4+0,2+0,2)^2$$

$$i(t) = 0,04$$

Threshold dengan nilai median yang ketiga yakni sebesar 11, sehingga untuk data $X_1 \leq 11$ akan dipilah ke simpul kiri sedangkan data $X_1 > 11$ akan dipilah ke simpul kanan. Perhitungan proporsi pengamatan untuk simpul kiri dan simpul kanan dapat dilihat pada Tabel 2.17 berikut ini.

Tabel 2.17 Proporsi Pengamatan Simpul Kiri dan Simpul Kanan dengan *Threshold* $X_1=11$

Simpul kiri		Simpul kanan	
$N_L = 8$	$p_L = 0,8$	$N_R = 2$	$p_R = 0,2$
$N(1 t_L)=3$	$p(1 t_L)=3/8$	$N(1 t_R)=1$	$p(1 t_R)=1/2$
$N(2 t_L)=3$	$p(2 t_L)=3/8$	$N(2 t_R)=0$	$p(2 t_R)=0$
$N(3 t_L)=2$	$p(3 t_L)=2/8$	$N(3 t_R)=1$	$p(3 t_R)=1/2$

$$i(t) = \frac{(0,8)(0,2)}{4} [|(3/8-1/2)| + |(3/8-0)| + |(2/8-1/2)|]^2$$

$$i(t) = 0,04(0,125+0,375+0,25)^2$$

$$i(t) = 0,0225$$

b. Nilai *Impurity* untuk Variabel Faktor Genetik (X_2)

Variabel X_2 adalah variabel yang memiliki jumlah kategori sebanyak 2, sehingga banyaknya kemungkinan *threshold* adalah sebanyak satu kemungkinan. Data pada variabel X_2 yang memiliki kategori 1 dipilah sebagai simpul kiri, sedangkan data yang memiliki kategori 2 dipilah sebagai simpul kanan. Perhitungan nilai *impurity* variabel X_2 ditunjukkan sebagai berikut.

Tabel 2.18 Proporsi Pengamatan Simpul Kiri dan Simpul Kanan dengan *Threshold* $X_2=1 : 2$

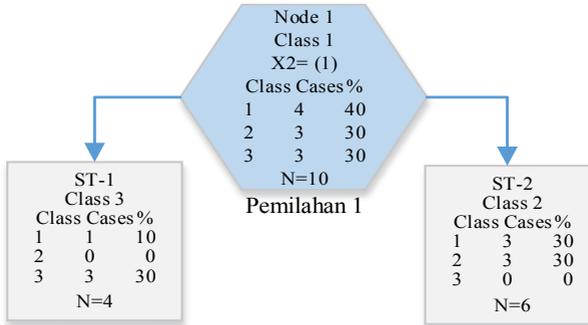
Simpul kiri		Simpul kanan	
$N_L = 4$	$p_L = 0,4$	$N_R = 6$	$p_R = 0,6$
$N(1 t_L)=1$	$p(1 t_L)=1/4$	$N(1 t_R)=3$	$p(1 t_R)=3/6$
$N(2 t_L)=0$	$p(2 t_L)=0$	$N(2 t_R)=3$	$p(2 t_R)=3/6$
$N(3 t_L)=3$	$p(3 t_L)=3/4$	$N(3 t_R)=0$	$p(3 t_R)=0$

$$i(t) = \frac{(0,4)(0,6)}{4} [|(1/4 - 3/6)| + |(0 - 3/6)| + |(3/4 - 0)|]^2$$

$$i(t) = 0,06(0,25 + 0,5 + 0,75)^2$$

$$i(t) = 0,135$$

Berdasarkan Tabel 2.18, dapat diketahui nilai *impurity* terbesar adalah sebesar 0,135 yang dihasilkan oleh *threshold* dari variabel $X_2=1:2$ yang berfungsi sebagai simpul utama. Sehingga ilustrasi pohon klasifikasi berdasarkan pemilah 1 akan digambarkan sebagai berikut.



Keterangan: S1: simpul 1 (simpul induk), S2: simpul 2, S3: simpul 3

Gambar 2.9 Pohon Klasifikasi Ilustrasi 2 Hasil Pemilahan I

Berdasarkan pohon klasifikasi ilustrasi 2 berdasarkan pemilah 1 seperti yang digambarkan pada Gambar 2.9, dapat diketahui simpul kanan masih bersifat heterogen untuk itu perlu dilakukan pemilahan kembali pada simpul kanan dengan menggunakan data sampel pengamatan ke 2,3,7,8,9, dan 10.

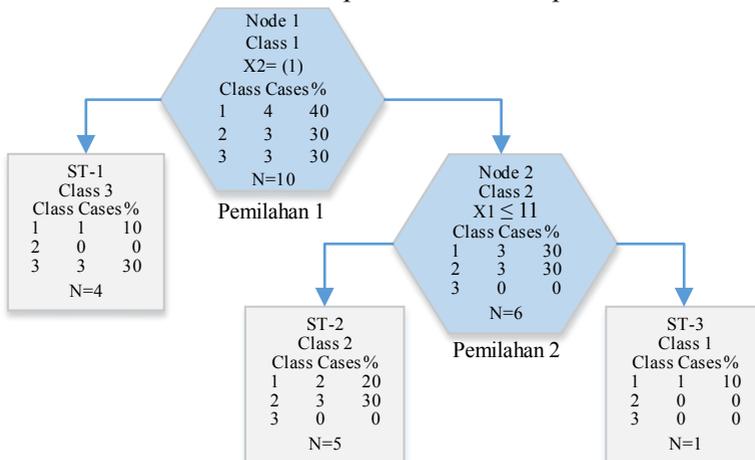
Kemudian nilai *impurity* yang dihasilkan ditunjukkan pada Tabel 2.19 berikut ini.

Tabel 2.19 Nilai Impurity Variabel X_1 dan X_2 Pada Simpul Kanan

Variabel	Threshold	Impurity ($i(t)$)
X_1	5,5	0,0499*
	8	0,0278
	11	0,0499*
X_2	1:2	0

* *impurity* maksimum

Nilai *impurity* maksimum terdapat pada *threshold* variabel X_1 (umur penderita) dengan median 5,5 dan 11. Sehingga pemilah terbaik yang dihasilkan adalah variabel umur penderita dan berperan sebagai pemilah 2. Selanjutnya, 6 data sampel akan dipilah berdasarkan proses pemilahan 2 menjadi dua anak simpul baru. Berdasarkan kedua pilihan pemilah dipilih pemilah terbaik yakni variabel umur penderita dengan median 11 sehingga apabila data variabel X_1 bernilai kurang dari 11 akan dipilah menjadi simpul kiri dan data sampel dengan nilai lebih dari 11 akan dipilah menjadi simpul kanan. Hasil pemilahan untuk simpul kanan pada kedalaman 2 dalam sebuah ilustrasi pohon klasifikasi pada Gambar 2.10.



Gambar 2.10 Pohon Klasifikasi Pemilahan 2 Pada Simpul Kanan

Pemberhentian pohon berdasarkan Breiman, Friedman, Olshen, dan Stone (1993) terjadi ketika jumlah anggota pada simpul terminal berjumlah ≤ 5 pengamatan atau sampai dengan batas minimum yang telah ditentukan. Sehingga jumlah simpul terminal yang dihasilkan adalah sebanyak 3 simpul yakni simpul yang diberi tanda S1, S4, dan S5.

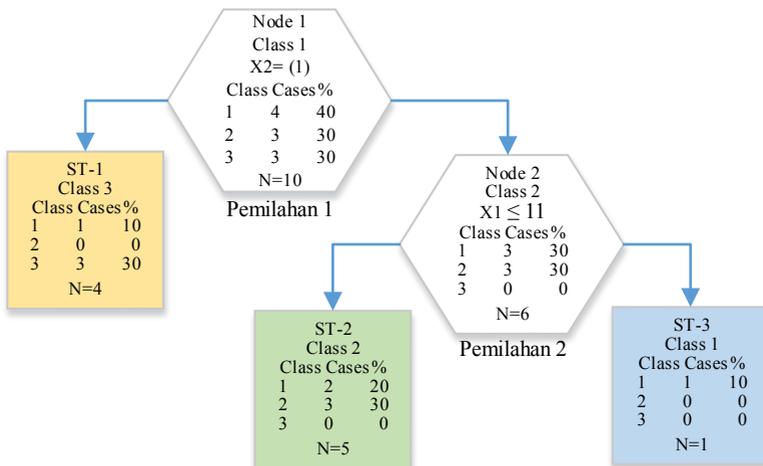
Setelah pemberian label kelas dilakukan, tahapan selanjutnya adalah pemangkasan pohon (*pruning*). Jika pohon klasifikasi yang terbentuk terlalu kompleks maka dapat dilakukan *pruning* dengan menggunakan persamaan (2.9) untuk mendapatkan pohon klasifikasi yang optimal. *Pruning* dilakukan untuk menghindari adanya kasus *overfitting* dan *underfitting*.

Tahap selanjutnya dalam pembentukan pohon klasifikasi yakni pelabelan kelas pada tiap *terminal nodes* untuk mengetahui karakteristik yang dihasilkan oleh masing-masing tipe pada variabel respon. Pemberian label kelas dilakukan dengan mencari nilai proporsi tertinggi dari masing-masing kelas pada variabel respon. Label kelas 1 adalah penderita retardasi mental tipe ringan, label kelas 2 adalah penderita retardasi mental tipe sedang, dan label kelas 3 adalah penderita retardasi mental tipe berat. Penandaan label pada simpul terminal dapat dilihat pada Tabel 2.20 berikut.

Tabel 2.20 Pemberian Label pada Simpul Terminal 2

Simpul	Kelas	$N(j t)$	$p(j t)$	Label kelas
S2	1	1	25%	RM Berat
	2	0	0	
	3	3	75%	
Total		4	100%	
S4	1	2	40%	RM Sedang
	2	3	60%	
	3	0	0	
Total		5		
S5	1	1	100%	RM Ringan
	2	0	0	
	3	0	0	
Total		1	100%	

Berdasarkan Tabel 2.20, dapat diketahui proporsi masing-masing simpul terminal berdasarkan tiap kelas pada variabel respon yang dihasilkan dengan menggunakan indeks gini dan juga dengan menggunakan indeks twoing menghasilkan ketepatan klasifikasi yang sama yakni sebanyak 7 data sampel pengamatan tepat diklasifikasikan, sedangkan sebanyak 3 data sampel pengamatan mengalami kesalahan pengklasifikasian. Simpul terminal 2 dapat diberi label kelas 3 yang ditunjukkan oleh kotak berwarna kuning. Sedangkan simpul terminal 4 diberi label kelas 2 yang ditunjukkan oleh kotak berwarna hijau, dan simpul terminal 5 diberi label kelas 3 yang ditunjukkan oleh kotak berwarna biru pada Gambar 2.11 berikut.



Gambar 2.11 Pohon Klasifikasi dengan Label Kelas Ilustrasi 2

Pohon klasifikasi yang terbentuk dapat digunakan untuk mengklasifikasikan pasien RM baru. Pengklasifikasian pasien RM dilakukan dengan cara menelusuri pohon klasifikasi yang terbentuk dari atas yakni dimulai dari simpul induk kemudian ke bawah sampai dengan simpul terminal dan diperoleh kelas yang menunjukkan tipe dari retardasi mental yang diderita oleh pasien. Berikut ini merupakan ilustrasi dalam mengklasifikasikan pasien

RM baru menggunakan variabel umur penderita (X_1) dan faktor genetik (X_2) berdasarkan Tabel 2.4 yang dijelaskan pada Tabel 2.21 sebagai berikut.

Tabel 2.21 Pasien RM Baru dengan Karakteristik Variabel Umur Penderita (X_1) dan Faktor Genetik (X_2)

Pasien	X_1	X_2
P1	5	1
P2	4	2
P3	5	1
P4	12	1
P5	14	2
P6	10	1
P7	4	2
P8	8	2
P9	10	2
P10	12	2

Berikut ini adalah aturan (*rules*) yang digunakan untuk mengklasifikasi pasien RM baru.

1. Untuk pasien 1 (P1) pengklasifikasi dapat dilakukan dengan melihat karakteristik yang dimiliki oleh pasien tersebut. Apakah pasien P1 memiliki faktor genetik kategori 1 (ada keturunan)? Jika iya, maka pasien tersebut merupakan penderita RM berat.
2. Untuk pasien 2 (P2), apakah pasien tersebut memiliki faktor genetik kategori 1 (ada keturunan)? Jika tidak, maka dilihat umur pasien tersebut. Apakah pasien tersebut memiliki umur ≤ 11 tahun? Jika iya, maka pasien tersebut merupakan pasien penderita RM sedang.
3. Untuk pasien 5 (P5) dapat dilihat pada faktor genetik yang dimiliki. apakah pasien tersebut memiliki faktor genetik kategori 1 (ada keturunan)? Jika tidak, maka dilihat umur pasien tersebut. Apakah pasien tersebut memiliki umur ≤ 11

tahun? Jika tidak, maka pasien tersebut merupakan pasien dengan tipe RM ringan.

Sehingga hasil pengklasifikasian pasien RM baru berdasarkan karakteristik variabel umur penderita (X_1) dan faktor genetik (X_2) yang dimiliki adalah sebagai berikut.

Tabel 2.22 Hasil Klasifikasi Pasien RM Baru

Pasien	X_1	X_2	Tipe RM (Y)
P1	5	1	RM Berat
P2	4	2	RM Sedang
P3	5	1	RM Berat
P4	12	1	RM Berat
P5	14	2	RM Ringan
P6	10	1	RM Berat
P7	4	2	RM Sedang
P8	8	2	RM Sedang
P9	10	2	RM Sedang
P10	12	2	RM Ringan

Berdasarkan 10 pasien RM baru dapat diketahui tipe retardasi mental yang diderita oleh pasien. Sebanyak 4 pasien menderita RM berat, sebanyak 4 pasien menderita RM sedang, dan sebanyak 2 pasien menderita RM ringan.

2.4.6 Ukuran Ketepatan Klasifikasi

Menurut Lewis (2000), masalah klasifikasi mencakup empat hal utama. Pertama adalah variabel respon yang bersifat kategorik. Kedua adalah adanya variabel prediktor yang berpengaruh terhadap variabel respon. Ketiga adalah data *learning*, yang digunakan untuk menentukan model klasifikasi. Kemudian yang keempat adalah adanya data *testing*, yang digunakan untuk memprediksi ketepatan klasifikasi berdasarkan model yang dihasilkan oleh data *learning*. Beberapa indikator yang dapat digunakan untuk mengukur ketepatan klasifikasi diantaranya adalah melalui perhitungan *apparent error rate* (APER), *total accuracy rate* (1-APER), dan

akurasi untuk masing-masing kelas variabel respon. Untuk menentukan ketepatan klasifikasi dibentuk tabel *crossstab* sebagai berikut.

Tabel 2.23 *Crosstab* Ketepatan Klasifikasi

Kelas Observasi Y	Kelas Prediksi Y			Total
	1	2	3	
1	n_{11}	n_{12}	n_{13}	$N_{1.}$
2	n_{21}	n_{22}	n_{23}	$N_{2.}$
3	N_{31}	N_{32}	N_{33}	$N_{3.}$
Total	$N_{.1}$	$N_{.2}$	$N_{.3}$	N

Tabel *Crosstab* untuk menghitung ketepatan klasifikasi ditunjukkan dalam Tabel 2.23 dengan

n_{11} = Jumlah observasi dari variabel Y kategori 1 yang tepat diprediksikan sebagai variabel Y kategori 1

n_{21} = Jumlah observasi dari variabel Y kategori 2 yang salah diprediksikan sebagai variabel Y kategori 1

n_{31} = Jumlah observasi dari variabel Y kategori 3 yang salah diprediksikan sebagai variabel Y kategori 1

n_{12} = Jumlah observasi dari variabel Y kategori 1 yang salah diprediksikan sebagai variabel Y kategori 2

n_{22} = Jumlah observasi dari variabel Y kategori 2 yang tepat diprediksikan sebagai variabel Y kategori 2

n_{32} = Jumlah observasi dari variabel Y kategori 3 yang tepat diprediksikan sebagai variabel Y kategori 2

n_{13} = Jumlah observasi dari variabel Y kategori 1 yang tepat diprediksikan sebagai variabel Y kategori 3

n_{23} = Jumlah observasi dari variabel Y kategori 2 yang tepat diprediksikan sebagai variabel Y kategori 3

n_{33} = Jumlah observasi dari variabel Y kategori 3 yang tepat diprediksikan sebagai variabel Y kategori 3

$N_{1.}$ = Jumlah observasi dari variabel Y kategori 1

$N_{2.}$ = Jumlah observasi dari variabel Y kategori 2

$N_{3.}$ = Jumlah observasi dari variabel Y kategori 3

$N_{.1}$ = Jumlah prediksi dari variabel Y kategori 1

$N_{.2}$ = Jumlah prediksi dari variabel Y kategori 2

$N_{.3}$ = Jumlah prediksi dari variabel Y kategori 3

N = Jumlah total observasi/prediksi.

Ukuran-ukuran yang digunakan untuk menghitung ketepatan klasifikasi adalah sebagai berikut.

1. *Apparent error rate* (APER) merupakan proporsi observasi yang diprediksi secara tidak benar (ukuran kesalahan klasifikasi total) dengan formula sebagai berikut.

$$\text{Apparent Error Rate (APER)} = \frac{\text{Jumlah prediksi salah}}{\text{Jumlah total prediksi}} \quad (2.14)$$

2. *Total accuracy rate* atau (1-APER) merupakan proporsi observasi yang diprediksi secara benar (ukuran ketepatan klasifikasi total) dengan formula sebagai berikut.

$$\text{Total accuracy rate} = 1 - \text{APER} = \frac{n_{11} + n_{22} + n_{33}}{N} \quad (2.15)$$

2.5 Retardasi Mental

Retardasi mental merupakan suatu keadaan penyimpangan tumbuh kembang seorang anak sedangkan peristiwa tumbuh kembang itu sendiri merupakan proses hakiki atau utama yang terjadi pada anak. Retardasi mental dapat disebabkan oleh adanya gangguan pada fase prenatal, perinatal maupun postnatal. Mengingat beratnya beban keluarga maupun masyarakat yang harus ditanggung dalam penatalaksanaan retardasi mental, maka pencegahan yang efektif merupakan pilihan terbaik (Sularyo & Kadim, 2000). Sedangkan menurut *American Association of Mental Retardation* (2002) retardasi mental adalah kelainan pada fungsi intelektual dan terjadi disabilitas atau ketidakmampuan dalam hal menyesuaikan diri (perilaku adaptif) yang ditandai dengan fungsi intelektual di bawah rata-rata dan mengakibatkan pertumbuhan dan perkembangan anak menjadi tidak sempurna. Periode perkembangan menurut definisi ini adalah mulai dari lahir sampai umur 18 tahun.

2.5.1 Klasifikasi Retardasi Mental

Berdasarkan *The ICD-10 Classification of Mental and Behavioural Disorders*, WHO (1998) telah mengklasifikasikan penderita retardasi mental menjadi 4 golongan sebagai berikut.

1. *Mild Retardation* (Retardasi Mental Ringan)

Retardasi mental ringan dikategorikan sebagai retardasi mental dapat dididik (*educable*) yang memiliki tingkat *IQ* 50-69. Anak mengalami gangguan berbahasa tetapi masih mampu menguasainya untuk keperluan bicara sehari-hari dan untuk wawancara klinik. Umumnya, mereka juga mampu mengurus diri sendiri secara independen (makan, minum, berpakaian, mengontrol saluran cerna dan kandung kemih), meskipun tingkat perkembangannya sedikit lebih lambat dari ukuran normal (Stromme, 2000).

2. *Moderate Retardation* (Retardasi Mental Sedang)

Retardasi mental sedang dikategorikan sebagai retardasi mental dapat dilatih (*trainable*) yang memiliki tingkat *IQ* 35-49. Pada kelompok ini anak mengalami keterlambatan perkembangan pemahaman dan penggunaan bahasa, serta pencapaian akhirnya terbatas. Pencapaian kemampuan mengurus diri sendiri dan ketrampilan motorik juga mengalami keterlambatan, dan beberapa diantaranya membutuhkan pengawasan dari orang tua. Kemajuan di sekolah terbatas, sebagian masih bisa belajar membaca, menulis dan berhitung (Tomb, 2004).

3. *Severe Retardation* (Retardasi Mental Berat)

Kelompok retardasi mental berat ini hampir sama dengan retardasi mental sedang yang memiliki tingkat *IQ* 20-34, dalam hal gambaran klinis, penyebab organik, dan keadaan-keadaan yang terkait. Perbedaan utama adalah pada retardasi mental berat ini biasanya mengalami kerusakan motorik yang bermakna atau adanya defisit neurologis (Stromme, 2000).

4. *Profound Retardation* (Retardasi Mental Sangat Berat)

Retardasi mental sangat berat yang memiliki tingkat *IQ* <20, secara praktis anak sangat terbatas kemampuannya dalam mengerti dan menuruti permintaan atau instruksi. Umumnya anak sangat terbatas dalam hal mobilitas, dan hanya mampu melakukan komunikasi nonverbal (Tomb, 2004).

Penderita retardasi mental dengan berat ringannya gejala yang timbul mampu diklasifikasikan berdasarkan fungsi intelektual dan kemampuan untuk menyesuaikan diri. Fungsi intelektual seseorang dapat diukur dengan menggunakan tes *IQ*

(*intelligent quotient*). Contoh dari fungsi intelektual adalah kemampuan untuk berpikir, membuat perencanaan, memahami sebab akibat, dan memecahkan masalah. Sedangkan, kemampuan menyesuaikan diri adalah kemampuan yang dipelajari seseorang agar dapat menjalankan fungsi sehari-hari (Shevell, Ashwal, dan Donley, 2003). Kemampuan ini terdiri dari kemampuan konseptual, contohnya kemampuan membaca, menulis, dan berbahasa; kemampuan sosial, contohnya kepercayaan diri, sikap bertanggung jawab, interpersonal, dan mematuhi aturan; kemampuan praktis, contohnya kegiatan sehari-hari seperti makan, minum, membersihkan rumah, bekerja, dan keselamatan diri; dan kemampuan menyesuaikan diri juga dapat dinilai dengan menggunakan serangkaian tes yang sudah distandarisasi.

2.5.2 Faktor Etiologi (Faktor Penyebab)

Terjadinya retardasi mental tidak dapat dipisahkan dari tumbuh kembang seorang anak. Seperti diketahui faktor penentu tumbuh kembang seorang anak pada garis besarnya adalah faktor genetik yang menentukan sifat bawaan anak tersebut (faktor internal) dan faktor lingkungan (faktor eksternal). Lingkungan pada anak dalam konteks tumbuh kembang adalah suasana yang terbentuk dimana anak tersebut berada (Sularyo & Kadim, 2000).

Etiologi retardasi mental tipe klinis atau biologikal dapat dibagi dalam 3 fase yakni fase prenatal, perinatal dan postnatal. Penyebab prenatal diantaranya adalah kelainan kromosom, kelainan genetik atau herediter, gangguan metabolik, sindrom dismorfik, infeksi intrauterin, dan intoksikasi. Sedangkan, penyebab perinatal diantaranya adalah prematuritas, asfiksia, kernikterus, hipoglikemia, meningitis, hidrosefalus, perdarahan, dan intraventrikular. Penyebab postnatal diakibatkan oleh infeksi (meningitis dan ensefalitis), trauma, kejang lama, dan intoksikasi seperti keracunan (Sularyo dan Kadim, 2000).

2.5.3 Gejala Psikiatrik (Tingkah Laku)

Tingkah laku yang sering ditunjukkan oleh penderita retardasi mental dapat dijadikan sebagai penentu dalam pengklasifikasian penderita RM. Seperti diketahui tingkah laku yang

ditunjukkan oleh masing-masing tipe penderita RM berbeda-beda. Banyak dari penderita RM yang menunjukkan tingkah laku normal seperti anak-anak pada umumnya hanya saja mereka yang menderita RM mengalami keterbatasan dalam melakukan suatu kegiatan. Sedangkan tidak sedikit juga anak-anak penderita RM menunjukkan tingkah laku yang hiperaktif, bahkan diantaranya ada yang tidak dapat dikontrol sehingga butuh pengawasan yang ketat. Gejala psikiatrik yang sering ditunjukkan lainnya adalah psikosa atau gejala seperti orang ketakutan jika bertemu atau berada di lingkungan yang baru (Prasadio, 1976). Sekitar 30% sampai 50% penyebab dari retardasi mental tidak diketahui penyebabnya (Moeschler, 2013).

2.6 Penelitian Sebelumnya

Penelitian sebelumnya mengenai retardasi mental dilakukan oleh Napolion (2010) menghasilkan kesimpulan bahwa penilaian tingkat kecerdasan untuk mendiagnosa pasien dengan retardasi mental harus berdasarkan semua informasi yang tersedia, termasuk temuan klinis, perilaku adaptif, dan hasil tes psikometrik. Prasadio (1976) dalam penelitiannya menghasilkan faktor infeksi sebagai penyebab retardasi mental cukup menonjol yakni sebanyak 37 persen dari 141 penderita yang telah diperiksa.

Sedangkan penelitian mengenai metode *Classification and Regression Trees* (CART) yang dilakukan oleh Rahmawati (2015) mengenai klasifikasi dengan metode CART yang dimodifikasi dengan metode *Random Forest* menghasilkan akurasi korban kecelakaan menggunakan yang menggunakan kombinasi data *training* dan *testing* sebesar 95%:5% dengan berbagai kombinasi pohon dimana yang memberikan ketepatan klasifikasi pada data *testing* terbesar adalah pada kombinasi pohon sebanyak 50. Ketepatan klasifikasi dengan kombinasi pohon sebanyak 50 untuk data *training* sebesar 78,69% dan data *testing* sebesar 58,33%.

Penelitian yang dilakukan oleh Seftiana (2014) mengenai klasifikasi rumah tangga sangat miskin berdasarkan paket bantuan rumah tangga yang diharapkan menghasilkan klasifikasi RTSM menurut bantuan rumah tangga yang diharapkan dengan metode

CART dilakukan dengan menggunakan proporsi data *learning* dan data *testing* sebesar 80% dan 20% karena menghasilkan *total accuracy rate* (1-APER) yang tinggi dengan jumlah *terminal node* yang tidak terlalu besar. Nilai 1-APER yang diperoleh yaitu sebesar 65,5% untuk data *learning* dan 62,8% untuk data *testing* dengan jumlah *terminal node* sebesar 23.

Penelitian juga dilakukan oleh Inayah (2014) mengenai klasifikasi rumah tangga di Kabupaten Jombang menggunakan metode *Classification and Regresion Trees* (CART) menghasilkan pemilah terbaik yang digunakan untuk pengklasifikasian adalah menggunakan indeks twoing dikarenakan jumlah simpul terminal yang dihasilkan lebih sedikit dibandingkan ketika menggunakan metode pemilah indeks gini yakni sebanyak 53 simpul. Sedangkan, akurasi yang dihasilkan kedua indeks sama yakni sebesar 39,7%.

Sedangkan penelitian dengan menggunakan metode SMOTE pernah dilakukan oleh Trapsilasiwi (2014) mengenai klasifikasi kanker payudara dan kanker serviks menghasilkan kesimpulan bahwa ketepatan klasifikasi kanker payudara sebelum menggunakan metode SMOTE adalah sebesar 89,2% dan setelah menggunakan metode SMOTE akurasi meningkat yakni sebesar 96,8%. Sedangkan pada klasifikasi kanker serviks sebelum menggunakan SMOTE menghasilkan ketepatan akurasi sebesar 40,4% dan sesudah menggunakan SMOTE meningkat menjadi 59,%.

“Halaman ini sengaja dikosongkan”

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

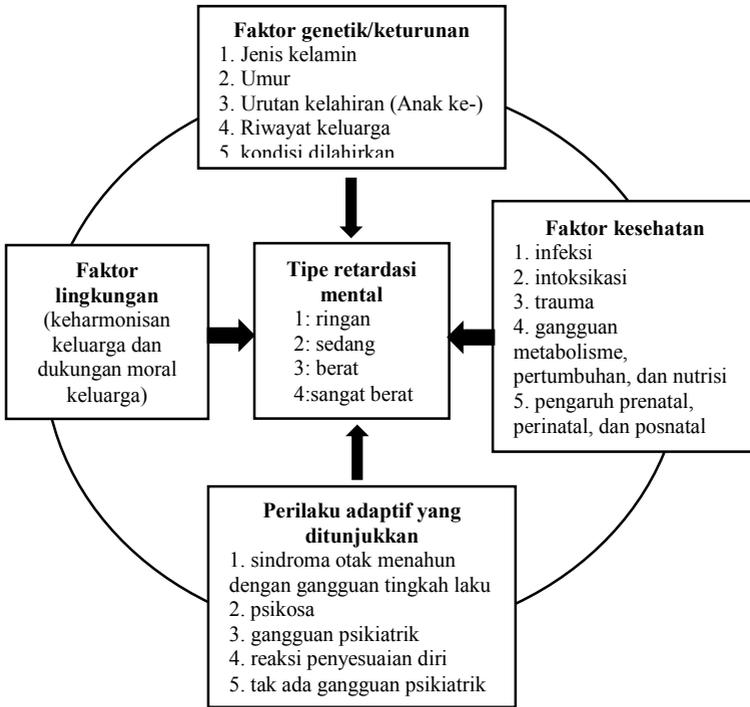
Data yang digunakan dalam penelitian tugas akhir ini merupakan data sekunder yang diambil pada poli tumbuh kembang di Rumah Sakit Jiwa Menur Provinsi Jawa Timur. Pengambilan data dilakukan berdasarkan pencatatan data rekam medis pasien penderita retardasi mental yang sedang menjalani proses rawat jalan pada tahun 2013-2014 dengan kriteria pasien berusia dibawah 18 tahun. Jumlah data yang tersedia berdasarkan pencatatan data rekam medis adalah sebanyak 101 pasien yang menderita retardasi mental di Rumah Sakit Jiwa Menur Provinsi Jawa Timur. Struktur data yang digunakan dalam penelitian ini ditampilkan dalam Tabel 3.1 sebagai berikut.

Tabel 3.1 Struktur Data

Data ke-	Y	X ₁	X ₂	...	X ₇
1	y ₁	x ₁₁	x ₁₂	...	x ₁₇
2	y ₂	x ₂₁	x ₂₂	...	x ₂₇
⋮	⋮	⋮	⋮	⋮	⋮
N	y _n	x _{n1}	x _{n2}	...	x _{n7}

3.2 Kerangka Konsep Penelitian

Kerangka konsep penelitian menunjukkan hubungan antara variabel prediktor dengan variabel respon. Jika dilihat berdasarkan penelitian sebelumnya dapat diketahui klasifikasi penderita retardasi mental dilakukan berdasarkan tingkat kecerdasan pasien dan juga perilaku adaptif yang ditunjukkan oleh pasien. Selain itu faktor lingkungan dan faktor genetik juga menentukan tipe penderita retardasi mental. Semakin kecil nilai *IQ* yang dimiliki oleh anak, maka resiko anak tersebut menderita retardasi mental dengan tipe berat semakin besar. Selain itu, kondisi kelahiran *premature* juga mengakibatkan resiko anak menderita retardasi mental semakin besar. Kerangka konsep penelitian berdasarkan teori H.L. Blum (1956) adalah sebagai berikut.



Gambar 3.1 Kerangka Konsep Penelitian Teori H.L. Blum

(Sumber : Notoadmodjo, 2010)

3.3 Variabel Penelitian

Variabel penelitian yang digunakan terdiri dari variabel respon (Y_n) dan variabel prediktor (X_n). Variabel prediktor yang digunakan merupakan faktor-faktor yang mempengaruhi pengklasifikasian penderita retardasi mental penentuannya berdasarkan penelitian-penelitian sebelumnya dan juga ketersediaan data rekam medis pasien di Rumah Sakit Jiwa Menur Provinsi Jawa Timur. Penentuan variabel respon (Y_n) yang terdiri dari tipe ringan, sedang, dan berat dilakukan berdasarkan kondisi mental yang dialami oleh penderita retardasi mental salah satunya dengan menggunakan faktor etiologik maupun gejala psikiatrik dan juga variabel lainnya yang berpengaruh terhadap pengklasifikasian.

Variabel respon dan juga variabel prediktor dapat dilihat pada Tabel 3.2 sebagai berikut.

Tabel 3.2 Variabel Penelitian

Variabel	Nama Variabel	Kategori	Skala Pengukuran
Y	Tipe Retardasi mental	1: Penderita retardasi mental tipe <i>mild</i> 2: Penderita retardasi mental tipe <i>moderate</i> 3: Penderita retardasi mental tipe <i>severe</i>	Ordinal
X ₁	Urutan Kelahiran Anak	-	Ordinal
X ₂	Kondisi saat dilahirkan	1: <i>premature</i> 2: tidak <i>premature</i>	Nominal
X ₃	Umur penderita (tahun)	-	Rasio
X ₄	Jenis Kelamin	1: laki-laki 2: perempuan	Nominal
X ₅	Faktor Genetik (Keturunan)	1: ada 2: tidak Ada	Nominal
X ₆	Faktor Etiologik	1: infeksi 2: intoksikasi 3: trauma 4: gangguan metabolisme, pertumbuhan, dan nutrisi 5: pengaruh prenatal	Nominal
X ₇	Gejala Psikiatrik	1: sindroma otak menahun dengan gangguan tingkah laku 2: psikosa 3: gangguan psikiatrik 4: reaksi penyesuaian diri 5: tak ada gangguan psikiatrik	Nominal

Klasifikasi variabel respon terdiri dari 3 tipe yang disesuaikan dengan data yang tersedia di Rumah Sakit Jiwa Menur Surabaya, sedangkan penggolongan *American Association on Mental Retardation (2002)* yang dapat dijelaskan sebagai berikut.

1. *Mild* adalah penderita retardasi mental dengan kategori ringan yang masih dapat dididik (*educable*).
2. *Moderate* adalah penderita retardasi mental sedang yang dikategorikan sebagai RM yang dapat dilatih (*trainable*).
3. *Severe* adalah penderita retardasi mental berat dikategorikan sebagai retardasi mental yang hampir sama dengan retardasi mental sedang dalam hal gambaran klinis, penyebab organik, dan keadaan-keadaan yang terkait.

Penentuan variabel prediktor dilakukan berdasarkan penelitian yang telah dilakukan sebelumnya mengenai penderita retardasi mental sebagai berikut.

1. Urutan kelahiran anak
Sebanyak 45 kasus retardasi mental dari 141 pasien yang menderita merupakan anak pertama dari ibu yang memiliki usia 16-25 tahun (Prasadio, 1976).
2. Kondisi saat dilahirkan
Kondisi penderita pada saat dilahirkan, dimana kelahiran *premature* mengakibatkan resiko seorang anak mengalami retardasi mental lebih besar (Prasadio, 1976).
3. Umur penderita
Retardasi mental mengacu pada proses pertumbuhan dan perkembangan seorang anak pada saat usia seorang anak di-dibawah 18 tahun (Sularyo dan Kadim, 2000).
4. Jenis kelamin
Adanya perbedaan perilaku seksual remaja retardasi mental perempuan dengan remaja retardasi mental laki-laki (Retnaningtyas dan Setyaningsih, 2009).
5. Faktor genetik (keturunan)
Terjadinya retardasi mental dipengaruhi oleh berbagai faktor, termasuk genetik dan lingkungan (Moeschler, 2013).
6. Faktor etiologik
Dalam diagnosa etiologik yakni pengaruh utama yang menyebabkan seorang anak menderita RM banyak disebabkan oleh infeksi dan gangguan metabolisme (Prasadio, 1976).

7. Gejala psikiatrik
Dalam diagnosa psikiatrik yakni tingkah laku yang ditunjukkan oleh pasien RM banyak disebabkan oleh sindroma otak menahun dengan gangguan tingkah laku (Prasadio, 1976).

3.4 Langkah Analisis Penelitian

Langkah analisis yang dilakukan dalam penelitian klasifikasi penderita retardasi mental di Rumah Sakit Jiwa Menur Provinsi Jawa Timur menggunakan *synthetic minority oversampling technique* (SMOTE) - *classification and regression trees* (CART) adalah sebagai berikut.

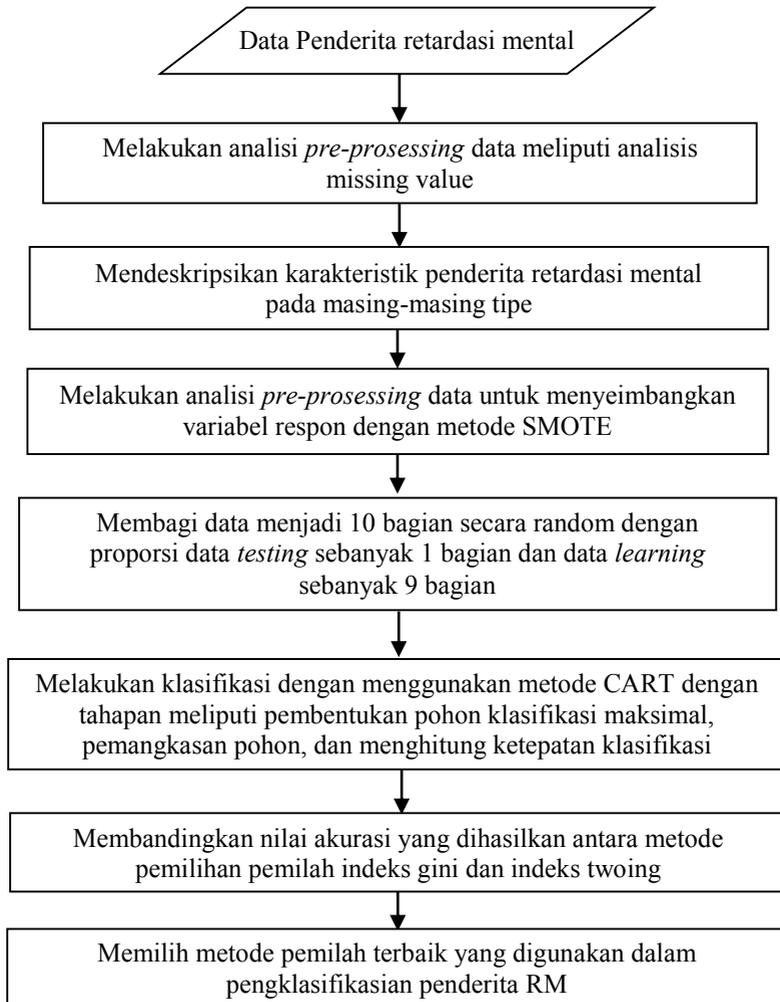
1. Melakukan analisis *pre-processing* data meliputi analisis *missing value*. Setelah dilakukan pemeriksaan adanya kasus *missing value* dilakukan penanganan untuk mengatasi adanya kasus *missing value* dengan menggunakan metode inputasi. Namun apabila data yang *missing* jumlahnya kurang dari 30% maka dilakukan eliminasi, artinya data yang mengandung *missing* tidak diikutkan dalam analisis.
2. Mendeskripsikan karakteristik penderita retardasi mental di Rumah Sakit Jiwa Menur Surabaya berdasarkan tiga tipe pada variabel respon. Jika variabel prediktor berskala rasio dihitung nilai rata-rata, standar deviasi, nilai minimum, dan nilai maksimum. Sedangkan untuk variabel prediktor dengan skala kategorik menggunakan *pie-chart* dan *bar-chart*.
3. Melakukan analisis *pre-processing* data selanjutnya yaitu menyeimbangkan jumlah data pada variabel respon dengan menggunakan metode *synthetic minority oversampling technique* (SMOTE). Pendekatan ini bekerja dengan membuat “*synthetic*” data, yaitu data replikasi dari data minor dengan menggunakan *k-nearest neighbors* berdasarkan jarak *Euclidean* antara kedua data terdekat dengan menggunakan *software* WEKA.
4. Membagi data secara random sebanyak 10 bagian yakni dengan proporsi 9 bagian akan dijadikan data *learning* dan

satu bagian dijadikan sebagai data *testing* dikarenakan dalam menentukan pohon optimal digunakan metode *10-fold cross validation estimate*.

5. Melakukan analisis klasifikasi dengan menggunakan metode CART menggunakan *software CART for windows version 4* melalui tahapan sebagai berikut.
 - a. Membentuk pohon klasifikasi maksimal dengan menggunakan data *learning* melalui tahapan sebagai berikut.
 1. Melakukan pemilihan pemilah (*classifier*) berdasarkan variabel prediktor menurut aturan pemilahan indeks gini dan indeks twoing yang kemudian hasil pemilahan dievaluasi dengan menggunakan kriteria *goodness of split*.
 2. Menentukan jumlah *terminal nodes* dengan cara menghentikan pembentukan pohon hingga dicapai batasan minimum pengamatan dan diperoleh *terminal nodes* yang homogen.
 3. Melakukan penandaan label kelas pada *terminal nodes* berdasarkan aturan jumlah terbanyak dari tiap kelas yang ada pada variabel respon.
 - b. Memangkas pohon (*pruning*) berdasarkan kriteria *cost complexity minimum*.
 - c. Melakukan seleksi pohon klasifikasi yang optimal dengan *10-fold cross validation*.
 - d. Menghitung ketepatan klasifikasi pohon hasil bentukan dengan menggunakan data *learning* dan melakukan validasi dengan menggunakan data *testing* pada pohon klasifikasi yang terbentuk. Ukuran ketepatan klasifikasi yang digunakan adalah *apparent error rate* (APER), *total accuracy rate* (1-APER), dan akurasi pada masing-masing tipe pada variabel respon.
6. Membandingkan hasil ketepatan klasifikasi berdasarkan perhitungan dengan menggunakan indeks gini dan indeks twoing dengan melihat nilai akurasi tertinggi dan jumlah *terminal nodes* yang paling minimum.

3.5 Diagram Alir Penelitian

Tahapan dalam penelitian berdasarkan langkah analisis penelitian akan digambarkan melalui diagram alir penelitian berdasarkan pada Gambar 3.2 sebagai berikut.



Gambar 3.2 Diagram Alir Penelitian

“Halaman ini sengaja dikosongkan”

BAB IV

ANALISIS DAN PEMBAHASAN

4.1 Karakteristik Deskriptif Penderita Retardasi Mental di Rumah Sakit Jiwa Menur Provinsi Jawa Timur

Jumlah penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 berjumlah 101 pasien. Berdasarkan jumlah tersebut sebanyak 59 pasien termasuk penderita retardasi mental dengan tipe ringan, sebanyak 30 pasien termasuk penderita retardasi mental dengan tipe sedang, dan sebanyak 12 pasien termasuk tipe pasien tipe berat dengan koding 1 untuk penderita RM tipe ringan, 2 untuk penderita RM tipe sedang, dan 3 untuk penderita RM tipe berat yang ketiganya dijadikan sebagai variabel respon dalam penelitian ini. Selain itu juga diperoleh informasi mengenai indikator-indikator yang mempengaruhi klasifikasi pasien penderita RM di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur, indikator tersebut diantaranya adalah urutan kelahiran anak, kondisi pasien pada saat dilahirkan (*premature* atau tidak *premature*), umur penderita (tahun), jenis kelamin, faktor genetik (keturunan), faktor etiologik, dan gejala psikiatrik. Indikator-indikator tersebut berguna untuk menentukan pasien masuk dalam klasifikasi retardasi mental tipe ringan, sedang, atau berat.

Namun, dari indikator atau variabel yang didapatkan untuk mengklasifikasikan penderita retardasi mental (RM), tidak semua memberikan informasi yang lengkap dan jelas. Terdapat data yang hilang (*missing value*) dari variabel atau indikator yang digunakan. Untuk itu, perlu dilakukan *pre-processing* data sebelum melangkah pada tahap analisis selanjutnya. Setelah ditelusuri, diperoleh adanya *missing value* yang terdapat dalam satu variabel yaitu variabel urutan kelahiran anak (X_1) yakni sebanyak 4 kasus dari 101 data yang diperoleh atau sebanyak 3,96%. Jumlah *missing value* yang terdapat dalam variabel urutan kelahiran anak berjumlah <30%, sehingga diputuskan keempat data (kasus) yang mengandung *missing value* pada variabel urutan kelahiran anak dihilangkan

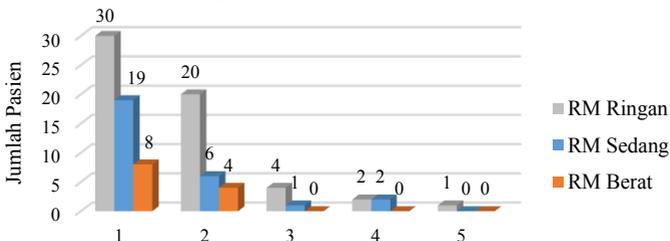
dalam analisis guna memberikan informasi yang lebih akurat. Keempat data *missing value* tersebut terdapat pada kelas RM ringan sebanyak 2 data dan RM sedang sebanyak 2 data. Sehingga jumlah data yang tersisa yakni sebanyak 97 data dengan perbandingan 57 data adalah pasien RM tipe ringan, 28 data adalah pasien RM tipe sedang, dan 12 data adalah pasien RM tipe berat. Perbandingan banyaknya penderita retardasi mental berdasarkan tiga klasifikasi yang terdapat di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 akan digambarkan pada Gambar.4.1 sebagai berikut.



Gambar 4.1 Jumlah Penderita RM di RSJ Menur Berdasarkan Masing-Masing Tipe

Jumlah pasien penderita RM untuk masing-masing tipe tidak seimbang, tipe RM ringan merupakan tipe yang paling banyak diderita oleh pasien di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014. Sedangkan tipe RM yang paling sedikit ditemukan adalah pasien dengan tipe RM berat.

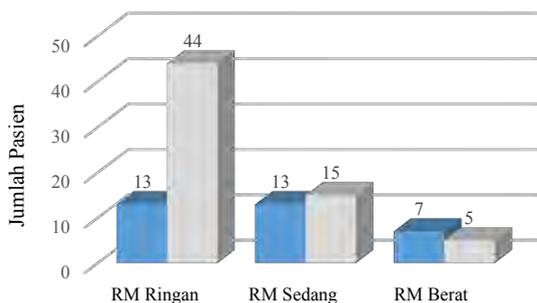
Selain itu, didapatkan informasi mengenai indikator urutan kelahiran anak berdasarkan masing-masing tipe penderita RM di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 pada gambar 4.2 berikut ini.



Gambar 4.2 Perbandingan Urutan Kelahiran Anak Berdasarkan Masing-Masing Tipe

Berdasarkan Gambar 4.2, didapatkan informasi perbandingan urutan kelahiran anak berdasarkan tipe RM yang diderita oleh pasien RSJ Menur Surabaya Provinsi Jawa Timur yakni pada tipe RM ringan sebanyak 30 pasien merupakan anak pertama, sebanyak 20 pasien merupakan anak kedua, sebanyak 4 pasien merupakan anak ketiga, sebanyak 2 pasien merupakan anak keempat, dan sebanyak 1 pasien merupakan anak kelima. Sedangkan, jika dilihat pada tipe RM sedang dapat diketahui sebanyak 19 pasien merupakan anak pertama, sebanyak 6 pasien merupakan anak kedua, sebanyak 1 pasien merupakan anak ketiga, dan sebanyak 2 pasien merupakan anak keempat. Sebanyak 12 pasien RM tipe berat merupakan anak pertama yakni sebanyak 8 pasien dan 4 pasien adalah anak kedua. Berdasarkan perbandingan 3 tipe RM dapat diketahui yang merupakan anak pertama adalah sebanyak 57 pasien dari 97 pasien yang diteliti atau sebanyak 58,76% (Lampiran B). Sehingga, dapat diperoleh kesimpulan pasien yang mengalami retardasi mental didominasi oleh anak pertama.

Kemudian untuk indikator yang kedua adalah kondisi pasien pada saat dilahirkan. yakni pasien yang diagnosis menderita retardasi mental, dilahirkan dengan kondisi *premature* atau normal. Perbandingan kondisi pasien yang menderita retardasi mental (RM) pada saat dilahirkan dapat dilihat berdasarkan Gambar 4.3 sebagai berikut.



Gambar 4.3 Perbandingan Kondisi Dilahirkan Berdasarkan Masing-Masing Tipe

Perbandingan kondisi pasien pada saat dilahirkan (*premature* atau tidak *premature*) yang digambarkan pada Gambar 4.3 dapat diperoleh informasi untuk pasien tipe RM ringan yang berjumlah 57 pasien, 13 diantaranya lahir dengan kondisi *premature* dan sebanyak 44 pasien lahir dengan kondisi normal. Untuk pasien tipe RM sedang sebanyak 28 pasien, 13 diantaranya lahir dengan kondisi *premature* dan sebanyak 15 pasien lahir dengan kondisi normal. Sedangkan untuk pasien tipe RM berat dapat diketahui dari 12 pasien yang diteliti 7 pasien diantaranya dilahirkan dengan kondisi *premature* dan 5 pasien lainnya lahir dengan keadaan normal (Lampiran B). Berdasarkan informasi tersebut dapat disimpulkan bahwa pasien RM ringan didominasi oleh anak yang lahir secara normal, sedangkan pasien RM berat didominasi oleh anak yang dilahirkan secara *premature*, sehingga dapat diindikasikan kelahiran *premature* dapat meningkatkan resiko anak tersebut diklasifikasikan menderita retardasi mental dengan tipe berat semakin tinggi.

Untuk indikator atau variabel dengan skala kontinyu yakni umur penderita berdasarkan batasan masalah yang terdapat pada subbab 1.5 menunjukkan pasien RM yang diteliti merupakan pasien yang berumur dibawah 18 tahun. Berikut ini deskriptif karakteristiknya berdasarkan masing-masing tipe RM yang ditunjukkan pada Tabel 4.1.

Tabel 4.1 Deskriptif Karakteristik Umur Penderita Berdasarkan Masing-Masing Tipe

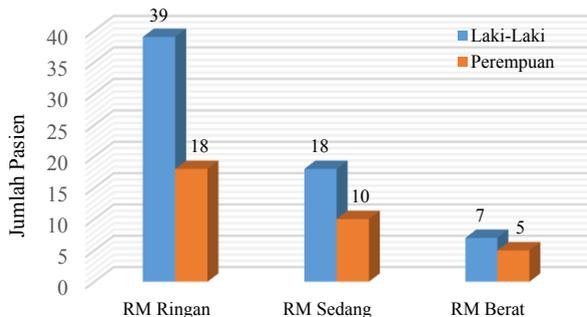
Tipe RM	Rata-Rata	Deviasi Standar	Minimum	Maksimum
Ringan	10,789	3,886	2	17
Sedang	12,143	2,978	6	17
Berat	13,08	4,72	3	17

Berdasarkan Tabel 4.1 dapat diketahui beberapa informasi mengenai indikator umur penderita untuk masing-masing tipe RM. Untuk pasien tipe RM ringan rata-rata umur pasien adalah $10,783 \approx 11$ tahun, dengan nilai deviasi standar yang cukup besar yakni sebesar 3,886 menunjukkan data umur penderita untuk tipe

RM ringan sangat bervariasi. Dapat diketahui pula pasien penderita RM tipe ringan yang paling kecil berumur 2 tahun dan yang paling besar berumur 17 tahun. Sedangkan, untuk pasien RM tipe sedang menunjukkan rata-rata pasien berumur $12,143 \approx 12$ tahun, dengan deviasi standar sebesar 2,978 menunjukkan umur penderita pada data pasien RM tipe sedang sangat bervariasi.

Berdasarkan Tabel 4.1 dapat diketahui, pasien RM tipe sedang yang paling kecil berumur 6 tahun dan yang paling besar berumur 17 tahun. Untuk pasien RM tipe berat rata-rata umur pasien yang berada di RSJ Menur adalah $13,08 \approx 13$ tahun, dengan deviasi standar sebesar 4,72 yang menunjukkan variasi umur penderita pasien RM tipe berat sangat beragam. Selain itu, dapat diketahui pasien RM tipe berat yang paling kecil berusia 3 tahun sedangkan yang paling besar berusia 17 tahun. Sehingga didapatkan informasi penderita retardasi mental di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur merupakan pasien yang masih digolongkan pada usia anak-anak.

Untuk indikator selanjutnya yaitu jenis kelamin pasien berdasarkan masing-masing tipe RM yang ditunjukkan pada Gambar 4.4 sebagai berikut.

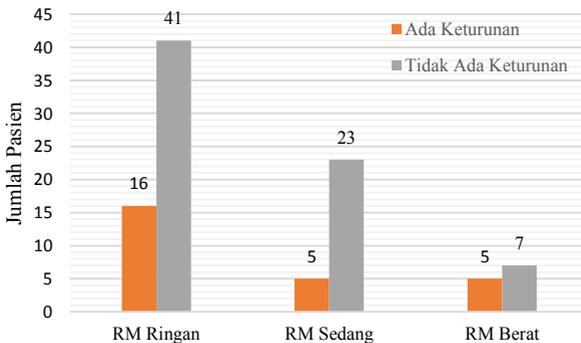


Gambar 4.4 Perbandingan Jenis Kelamin Berdasarkan Masing-Masing Tipe

Perbandingan jenis kelamin berdasarkan Gambar 4.4 menunjukkan untuk RM tipe ringan dapat diketahui dari 57 pasien, sebanyak 39 pasien merupakan pasien berjenis kelamin laki-laki

dan sebanyak 18 pasien berjenis kelamin perempuan. Untuk pasien tipe RM sedang dari 28 pasien, 18 diantaranya merupakan pasien dengan jenis kelamin laki-laki dan 10 pasien sisanya merupakan pasien dengan jenis kelamin perempuan. Sedangkan untuk pasien tipe RM berat dengan jumlah 12 pasien, sebanyak 7 pasien merupakan pasien laki-laki dan sebanyak 5 pasien merupakan pasien perempuan. Berdasarkan 97 pasien penderita RM di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 sebanyak 64 pasien atau 65,98% merupakan pasien laki-laki (Lampiran B). Sehingga dapat disimpulkan penderita RM banyak didominasi oleh pasien berjenis kelamin laki-laki.

Sedangkan, untuk indikator faktor genetik yakni indikasi bahwa pasien menderita RM berdasarkan keturunan dari anggota keluarga yang sebelumnya juga menderita RM atau untuk penderita retardasi mental yang tidak ditemukan adanya anggota keluarga yang menderita RM. Perbandingan faktor genetik berdasarkan masing-masing tipe RM dapat diketahui pada Gambar 4.5 sebagai berikut.

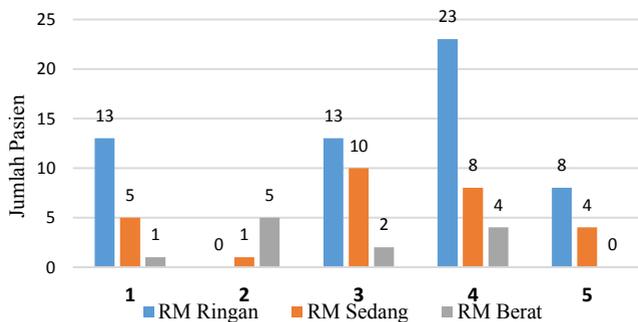


Gambar 4.5 Perbandingan Faktor Genetik Berdasarkan Masing-Masing Tipe

Berdasarkan Gambar 4.5 dapat diketahui, untuk pasien RM tipe ringan dengan jumlah 57 pasien, 16 diantaranya diketahui merupakan pasien yang didapatkan adanya anggota keluarga yang juga menderita RM dan 41 diantaranya merupakan pasien yang tidak didapatkan adanya anggota keluarga yang menderita RM.

Untuk pasien RM tipe sedang dapat diketahui dari 28 pasien, sebanyak 5 pasien merupakan pasien yang diketahui memiliki anggota keluarga yang juga menderita RM dan 23 pasien lainnya tidak memiliki anggota keluarga yang juga menderita RM. Sedangkan untuk pasien RM dengan tipe berat dengan jumlah 12 pasien, sebanyak 5 pasien memiliki anggota keluarga yang juga menderita RM dan sebanyak 7 pasien tidak memiliki anggota keluarga yang menderita RM.

Untuk indikator faktor etiologik yang merupakan faktor penyebab utama pasien menderita retardasi mental yang dikoding 1 untuk pasien yang disebabkan oleh infeksi, koding 2 untuk pasien yang disebabkan oleh intoksikasi, koding 3 untuk pasien yang disebabkan oleh trauma, koding 4 untuk pasien yang disebabkan oleh gangguan metabolisme, pertumbuhan, dan nutrisi, selanjutnya koding 5 untuk pasien yang disebabkan oleh pengaruh prenatal. Perbandingan faktor etiologik digambarkan pada Gambar 4.6 berikut ini.

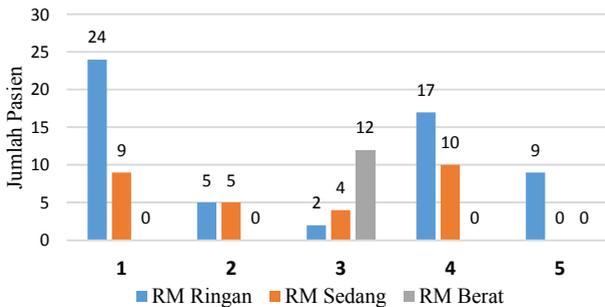


Gambar 4.6 Perbandingan Faktor Etiologik Berdasarkan Masing-Masing Tipe

Untuk pasien RM tipe ringan dapat diketahui penyebab utama pasien menderita RM tipe ringan disebabkan oleh gangguan metabolisme, pertumbuhan dan nutrisi dengan kasus terbanyak yakni sebanyak 23 pasien dari 57 pasien yang dikategorikan sebagai pasien RM tipe ringan. Untuk pasien RM tipe sedang dapat diketahui trauma seperti trauma kejang, demam (panas), dan jatuh merupakan faktor utama penyebab terjadinya RM tipe sedang

dengan jumlah kasus terbanyak yakni 10 pasien dari 28 pasien. Sedangkan faktor etiologik atau penyebab utama terjadinya RM pada tipe berat adalah disebabkan oleh intoksikasi atau keracunan zat-zat kimia selama dalam kandungan atau pada saat setelah dilahirkan, dengan jumlah kasus terbanyak yakni sebanyak 5 pasien dari 12 pasien. Dapat diketahui, faktor etiologik yang dimiliki oleh masing-masing pasien retardasi mental berdasarkan tipe ringan, sedang, dan berat sangat berbeda-beda (Lampiran B). Sehingga dapat disimpulkan faktor etiologik sangat berpengaruh dalam pengklasifikasian pasien retardasi mental atau disebut sebagai faktor penting untuk mengklasifikasikan pasien RM.

Untuk indikator yang terakhir yakni gejala psikiatrik atau gejala tingkah laku yang ditimbulkan oleh pasien RM yang dikoding 1 untuk pasien RM dengan gejala sindroma otak menahun dengan gangguan tingkah laku, koding 2 untuk psikosa, koding 3 untuk gangguan psikiatrik, koding 4 untuk reaksi penyesuaian diri, dan koding 5 untuk pasien yang tidak mengalami gangguan psikiatrik. Perbandingan gejala psikiatrik yang ditunjukkan pasien berdasarkan setiap tipe penderita RM digambarkan pada Gambar 4.7 berikut ini.



Gambar 4.7 Perbandingan Gejala Psikiatrik Berdasarkan Masing-Masing Tipe

Berdasarkan Gambar 4.7 dapat diketahui untuk pasien RM tipe ringan gejala psikiatrik yang paling banyak ditimbulkan adalah sindroma otak menahun dengan gangguan tingkah laku dengan jumlah kasus terbanyak yakni sebanyak 24 pasien dari 57 pasien.

Untuk pasien RM dengan tipe sedang gejala psikiatrik atau gejala tingkah laku yang paling banyak ditunjukkan adalah reaksi penyesuaian diri yakni sebanyak 10 pasien dari 28 pasien. Sedangkan, untuk pasien RM dengan tipe sangat berat gejala psikiatrik yang sering ditimbulkan adalah gangguan psikiatrik yakni sebanyak 12 kasus atau dapat dikatakan semua pasien dengan tipe RM berat dikategorikan memiliki tingkah laku yang sama yakni adanya gangguan psikiatrik. Jika dilihat pada Gambar 4.7, dapat diketahui penderita RM ringan, sedang, dan berat memiliki gejala psikiatrik yang berbeda (Lampiran B). Sehingga indikator gejala psikiatrik dapat dikatakan sebagai indikator yang sangat penting untuk mengklasifikasikan penderita retardasi mental di Rumah Sakit Jiwa Menur Provinsi Jawa Timur.

4.2 Analisis *Pre-processing* dengan Metode SMOTE untuk *Imbalanced Data*

Jumlah data pasien retardasi mental di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 berjumlah 97 pasien. Persentase jumlah data pada masing-masing variabel respon yang dikategorikan menjadi 3 tipe yakni tipe satu adalah penderita RM dengan tipe ringan, tipe dua adalah pasien RM dengan tipe sedang, dan tipe tiga adalah pasien RM dengan tipe berat yang digambarkan pada Gambar 4.8 berikut ini.



Gambar 4.8 Persentase Banyaknya Variabel Respon Berdasarkan Tipe RM

Jika dilihat pada Gambar 4.8 dapat diketahui jumlah persentase jumlah anggota untuk masing-masing variabel respon tidak seimbang yakni sebanyak 59% pasien merupakan penderita

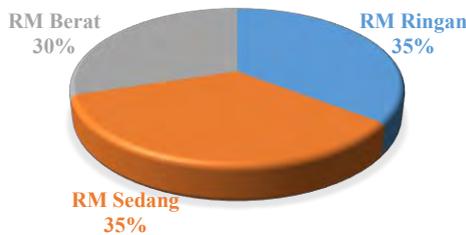
RM tipe ringan, sebanyak 29% merupakan pasien RM tipe sedang, dan sebanyak 12% pasien merupakan penderita RM tipe berat. Berdasarkan jumlah persentase yang dihasilkan dapat diketahui kelas mayor yakni kelas dengan jumlah anggota terbanyak terdapat pada kelompok RM tipe 1 atau RM tipe ringan. Untuk itu perlu dicari *synthetic* data dengan menggunakan persamaan (2.2) berdasarkan tetangga terdekat dengan menggunakan jarak euclidean pada persamaan (2.1) untuk kelas minor yakni kelompok RM tipe sedang dan tipe berat.

Untuk kelompok minor RM tipe sedang dengan jumlah anggota sebanyak 28 data maka perlu dilakukan replikasi sebanyak satu kali untuk menyeimbangkan jumlah anggota dengan kelas mayor yakni kelompok RM tipe ringan, karena replikasi yang dilakukan hanya sekali sehingga jumlah tetangga terdekat sama dengan nol untuk setiap anggota RM tipe sedang. Sehingga titik koordinat data hasil replikasi yang dihasilkan (x_{syn}) akan sama dengan titik koordinat data yang direplikasi (x_i). Sedangkan, untuk kelompok minor RM tipe berat dengan jumlah anggota sebanyak 12 pasien, replikasi yang perlu dilakukan adalah sebanyak 4 kali untuk mendapatkan jumlah anggota yang seimbang dengan anggota pada kelas mayor. Dengan melakukan replikasi sebanyak 4 kali maka diperoleh jumlah tetangga terdekat untuk masing-masing data adalah sebanyak 3 data. Titik koordinat data hasil replikasi yang dihasilkan (x_{syn}) akan berbeda dengan titik koordinat data yang direplikasi (x_i). Distribusi data setelah dilakukan replikasi dengan metode SMOTE ditunjukkan pada Tabel 4.2 sebagai berikut.

Tabel 4.2 Distribusi Data Masing-Masing Tipe RM Setelah Dilakukan SMOTE

Sebelum Direplikasi		Setelah Direplikasi		Jumlah Replikasi
Mayor	Minor	Mayor	Minor	
(Tipe 1)	(Tipe 2)	(Tipe 1)	(Tipe 2)	1 kali
57 (59%)	28 (29%)	57 (35%)	56 (35%)	4 kali
	(Tipe 3)		(Tipe 3)	
	12 (12%)		48 (30%)	

Sehingga, jumlah data yang semula berjumlah 97 data akan bertambah sampai dapat dikatakan jumlah anggota pada masing-masing kelas seimbang. Selain jumlah data pada variabel respon yang akan bertambah, jumlah data pada masing-masing variabel prediktor juga akan bertambah mengikuti jumlah data pada variabel respon. Dengan menyeimbangkan jumlah anggota pada masing-masing tipe pada variabel respon diharapkan tidak terjadi kasus *underfitting* atau *overfitting* dan menghasilkan tingkat akurasi yang baik. Sehingga jumlah anggota untuk masing-masing tipe pada variabel respon setelah dilakukan replikasi digambarkan pada Gambar 4.9 sebagai berikut.



Gambar 4.9 Persentase Banyaknya Variabel Respon Setelah Menggunakan SMOTE

Berdasarkan Gambar 4.9, dapat diketahui persentase banyaknya anggota pada masing-masing tipe pada variabel respon sudah seimbang. Semula dari jumlah data penelitian yang digunakan sebanyak 97 data kini bertambah menjadi 161 data. Data tersebut berasal dari *synthetic* data yang dihasilkan berdasarkan replikasi menggunakan SMOTE untuk menyeimbangkan jumlah data pada kelas mayor. Sehingga jumlah data pada RM tipe ringan (tipe 1) yang menjadi kelas mayor tetap berjumlah 57 data atau sebesar 35%, untuk jumlah data RM sedang (tipe 2) yang semula berjumlah 28 data setelah dilakukan replikasi sebanyak satu kali kini jumlahnya bertambah menjadi 56 data dengan persentase 35%, dan untuk RM tipe berat (tipe 3) yang semula berjumlah 12 data setelah direplikasi sebanyak 4 kali jumlahnya bertambah menjadi

48 data dengan persentase 30%. Setelah didapatkan jumlah data dengan proporsi kelas yang seimbang, dilanjutkan dengan tahapan analisis selanjutnya yaitu analisis *classification and regression trees* (CART).

4.3 Analisis CART untuk Klasifikasi Penderita RM di RSJ Menur Provinsi Jawa Timur

Berikut ini adalah hasil analisis dengan menggunakan metode CART untuk klasifikasi penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 dengan menggunakan data hasil *pre-processing* SMOTE yang berjumlah 161 data. Analisis CART bertujuan untuk mengetahui ketepatan klasifikasi yang telah dihasilkan berdasarkan 3 tipe retardasi mental yakni retardasi mental ringan (tipe 1), retardasi mental sedang (tipe 2), dan retardasi mental berat (tipe 3) berdasarkan pohon klasifikasi maksimal dan pohon klasifikasi optimal dengan menggunakan fungsi keheterogenan indeks gini dan indeks twoing. Penentuan pohon klasifikasi optimal dengan menggunakan metode *cross validation V-fold estimate* karena jumlah data sampel pengamatan yang digunakan ukurannya kecil. Jumlah *fold* yang digunakan adalah sebanyak 10 *fold*, sehingga dari jumlah data sebanyak 161 data akan dibagi menjadi 10 bagian yang masing-masing bagian berjumlah kurang lebih sebanyak 16 data. Namun, sebelum pembentukan pohon klasifikasi berikut ini akan dicari terlebih dahulu beberapa kemungkinan pemilahan berdasarkan masing-masing variabel prediktor.

4.3.1 Kemungkinan Pemilah untuk Setiap Variabel Prediktor

Dalam pembentukan pohon klasifikasi dibutuhkan variabel terpenting yang berperan sebagai pemilah terbaik yang akan menghasilkan nilai *threshold* untuk menentukan *goodness of split*. Variabel pemilah dan *threshold* dipilih dari beberapa kemungkinan pemilah dari masing-masing variabel prediktor berdasarkan jenis skala data pada variabel prediktor tersebut. Berikut ini perhitungan dalam menentukan kemungkinan pemilah berdasarkan variabel prediktor yang ditunjukkan pada Tabel 4.3 sebagai berikut.

Tabel 4.3 Banyaknya Kemungkinan Pemilah Variabel Prediktor

Variabel	Nama Variabel	Skala Data	Jumlah Kategori/nilai amatan sampel	Kemungkinan Pemilah
X ₁	Urutan kelahiran anak	Ordinal	5	$5-1=4$ pemilah
X ₂	Kondisi saat dilahirkan	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X ₃	Umur penderita (tahun)	Rasio	5	$5-1=4$ pemilah
X ₄	Jenis kelamin	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X ₅	Faktor genetik (keturunan)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X ₆	Faktor etiologik	Nominal	5	$2^{5-1} - 1 = 15$ pemilah
X ₇	Gejala psikiatrik	Nominal	5	$2^{5-1} - 1 = 15$ pemilah

Berdasarkan Tabel 4.3, dapat diketahui banyaknya kemungkinan pemilah pada masing-masing variabel prediktor. Untuk variabel X₁ (urutan kelahiran anak) kemungkinan pemilah yang dihasilkan adalah sebanyak 4 pemilah, untuk variabel X₂ (kondisi saat dilahirkan) menghasilkan kemungkinan pemilah hanya sebanyak 1 pemilah, untuk variabel umur penderita (X₃) jumlah kemungkinan pemilah yang dihasilkan adalah 4 pemilah. Jika dilihat jumlah kemungkinan pemilah untuk variabel X₄ (jenis kelamin) adalah sebanyak 1 pemilah, untuk variabel faktor genetik/keturunan (X₅) menghasilkan kemungkinan pemilah sebanyak 1 pemilah. Sedangkan, untuk variabel X₆ (faktor etiologik) kemungkinan pemilah yang dihasilkan adalah sebanyak 15 pemilah dan untuk variabel X₇ (gejala psikiatrik) jumlah kemungkinan pemilah yang dihasilkan adalah sebanyak 15 pemilah.

4.3.2 Analisis CART Penderita RM di RSJ Menur dengan Pemilahan Pemilah Indeks Gini

Analisis CART diawali dengan pembentukan pohon klasifikasi maksimal yang menghasilkan jumlah simpul terminal terbanyak dengan kedalaman yang tinggi, setelah itu dilakukan pemangkasan pohon dengan menggunakan pendekatan *cost complexity minimum* sampai didapatkan pohon klasifikasi optimal dengan metode *10-fold cross validation estimate*. Berikut ini merupakan tahapan dalam analisis CART untuk klasifikasi penderita retardasi mental (RM) di Rumah Sakit jiwa Menur Surabaya Provinsi Jawa Timur.

a. Pembentukan Pohon Klasifikasi Maksimal

Pembentukan pohon klasifikasi maksimal diawali dengan pemilihan *classifier* yakni variabel prediktor terpenting yang digunakan sebagai pemilah utama. Berdasarkan variabel prediktor yang digunakan, dilakukan pemilihan pemilah berdasarkan skor kontribusi terbesar yang dihasilkan oleh masing-masing variabel prediktor. Besarnya kontribusi masing-masing variabel prediktor ditampilkan dalam Tabel 4.4 berikut.

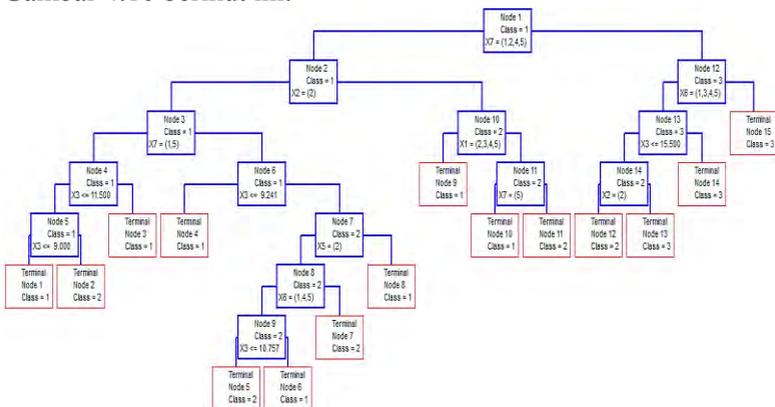
Tabel 4.4 Variabel Terpenting dari Pohon Klasifikasi Maksimal dengan Pemilah Indeks Gini

Variabel	NamaVariabel	Skor
X ₇	Gejala Psikiatrik	100,00
X ₆	Faktor Etiologik	77,77
X ₃	Umur Penderita	31,39
X ₂	Kondisi Dilahirkan	23,35
X ₁	Urutan Dilahirkan	15,94
X ₅	Faktor Genetik	7,07
X ₄	Jenis Kelamin	0,24

Berdasarkan Tabel 4.4, dapat diketahui variabel terpenting yang digunakan sebagai pemilah utama atau dapat dikatakan sebagai simpul induk (*parent node*) pada pohon klasifikasi maksimal adalah X₇ (gejala psikiatrik) dibuktikan dari skor kontribusi tertinggi yang dihasilkan yakni sebesar 100, selain itu beberapa variabel prediktor juga dapat dikatakan berpengaruh terhadap klasifikasi dalam pemilahan pemilah yaitu variabel X₆

(faktor etiologik) dengan skor kontribusi sebesar 77,77, dan variabel lainnya hanya memberikan kontribusi dibawah skor 50 untuk selengkapnya dapat dilihat pada Tabel 4.4. Gejala psikiatrik mampu menurunkan keheterogenan terbesar sehingga simpul yang dihasilkan akan lebih homogen.

Selanjutnya proses pemilahan terjadi sampai terbentuk pohon klasifikasi maksimal, dimana sifat dari pohon klasifikasi maksimal adalah menghasilkan simpul terminal terbanyak dengan kedalaman tertinggi. Jumlah simpul terminal yang dihasilkan dari pohon klasifikasi maksimal menggunakan pemilah indeks gini adalah sebanyak 15 simpul dengan tingkat kedalaman yang dihasilkan sebesar 8. Topology pohon maksimal dengan menggunakan pemilahan pemilah indeks gini ditunjukkan pada Gambar 4.10 berikut ini.

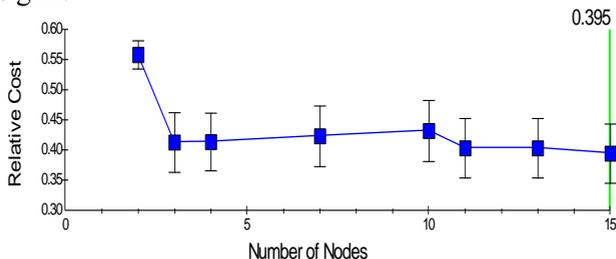


Gambar 4.10 Topology Pohon Klasifikasi Maksimal dengan Pemilahan Pemilah Indeks Gini

b. Pemangkasan (*Prunning*) Pohon Klasifikasi Maksimal

Setelah dihasilkan pohon klasifikasi maksimal, langkah selanjutnya adalah melihat apakah pohon klasifikasi maksimal tersebut perlu dilakukan pemangkasan (*prunning*) atau tidak. *Prunning* bertujuan untuk menghindari adanya kasus *overfitting* (nilai akurasi yang dihasilkan melebihi kenyataan yang ada) atau *underfitting* (nilai akurasi yang dihasilkan sangat kurang

dibandingkan dengan nilai *actual*). Untuk melihat perlu tidaknya dilakukan pemangkasan pohon, dapat dilihat pada nilai *relative cost* yang dihasilkan. Pemangkasan pohon dilakukan dengan menggunakan pendekatan *cost complexity minimum* dan dalam menentukan pohon klasifikasi optimal digunakan metode *10-fold cross validation estimate*, yakni membagi data menjadi 10 bagian dimana 9 bagian akan digunakan untuk membentuk pohon klasifikasi dan 1 bagian lainnya digunakan untuk pendugaan kesalahan klasifikasi atau untuk mengestimasi *misclassification error rate*. Setiap hasil pemangkasan memiliki nilai *relative cost* tertentu, kemudian dipilih hasil pemangkasan dengan nilai *relative cost* yang paling minimum. Berikut ini akan ditunjukkan *plot relative cost* pada klasifikasi penderita retardasi mental di Rumah Sakit Jiwa Menur Provinsi Jawa Timur dengan pemilihan pemilah indeks gini.



Gambar 4.11 Plot *Relative Cost* dalam Klasifikasi Penderita RM dengan pemilah Indeks Gini

Jika dilihat pada *plot relative cost* pohon klasifikasi dengan simpul terminal sebanyak 15 simpul sudah dapat dikatakan optimal ditunjukkan dengan nilai *relative cost* yang dihasilkan sebesar 0,395, ditandai dengan garis berwarna hijau. Sedangkan, untuk biaya kesalahan (*cross-validation relative cost*) menghasilkan nilai yang paling minimum yakni sebesar $0,395 \pm 0,049$ atau bekisar antara 0,444 sampai 0,346 dengan nilai kompleksitas sebesar 0,000. Berikut ini akan ditunjukkan urutan dalam pembentukan pohon klasifikasi sampai didapatkan pohon klasifikasi optimal yang ditampilkan pada Tabel 4.5 berikut.

Tabel 4.5 Urutan Pembentukan Pohon Klasifikasi (*Tree Sequence*) 1

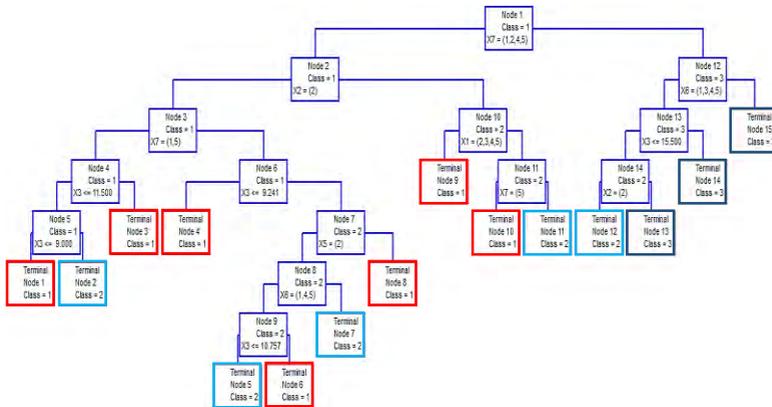
<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Cross-Validated Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity</i>
1**	15	0,395 ± 0,049	0,150	0,000
2	13	0,404 ± 0,049	0,159	0,003
3	11	0,404 ± 0,049	0,168	0,003
4	10	0,433 ± 0,051	0,186	0,012
5	7	0,424 ± 0,050	0,240	0,012
6	4	0,415 ± 0,048	0,302	0,014
7	3	0,414 ± 0,049	0,355	0,035
8	2	0,559 ± 0,023	0,518	0,108
9	1	1,000 ± 0,000	1,000	0,322

** *pohon klasifikasi optimal*

Berdasarkan Tabel 4.5, jika dilihat pada biaya kesalahan (*cross-validation relative cost*) yang dihasilkan, pohon klasifikasi 1 memiliki biaya kesalahan yang paling minimum, sehingga pohon klasifikasi dengan simpul terminal sebanyak 15 simpul sudah layak atau dapat dikatakan optimal tanpa harus dilakukan pemangkasan terlebih dahulu.

c. Pemilihan Pohon Klasifikasi Optimal

Pohon klasifikasi maksimal yang dihasilkan juga dapat dikatakan sebagai pohon klasifikasi optimal karena tidak perlu dilakukan *prunning*. Langkah selanjutnya dalam analisis CART adalah menentukan pohon klasifikasi optimal yang digunakan untuk mengetahui akurasi yang dihasilkan dari klasifikasi penderita retardasi mental di RSJ Menur berdasarkan aturan pemilahan indeks gini. Variabel-variabel apa saja yang berperan penting dalam pemilahan dapat dilihat pada besarnya nilai skor atau kontribusi yang diberikan oleh masing-masing variabel. Berdasarkan Tabel 4.4, dapat diketahui variabel terpenting dengan skor tertinggi yang berperan sebagai simpul utama adalah variabel gejala psikiatrik. Proses pemilahan dengan pemilah indeks gini yang dilakukan oleh variabel terpenting yakni gejala psikiatrik dalam pembentukan pohon optimal secara visual akan ditampilkan oleh Gambar 4.12 berikut ini.



Gambar 4.12 Topology Pohon Klasifikasi Optimal dengan Pemilahan Pemilah Indeks Gini

Berdasarkan total data sebanyak 161 data yang dilabelkan menjadi kelas 1 (RM ringan) akan memilah menjadi 2 simpul anak yakni simpul kiri dan simpul kanan berdasarkan gejala psikiatrik yang ditunjukkan. Jika pasien RM tersebut menunjukkan tingkah laku sindroma otak menahun dengan gangguan tingkah laku (kategori 1), psikosa (kategori 2), reaksi penyesuaian diri (kategori 4), dan tidak ada gangguan psikiatrik (kategori 5) akan dipilah menjadi simpul kiri dengan label kelas 1 (RM ringan) sebanyak 103 pasien, sedangkan untuk penderita retardasi mental yang menunjukkan tingkah laku seperti adanya gangguan psikiatrik (kategori 3) akan dipilah menjadi simpul kanan dengan label kelas 3 (RM berat) sebanyak 58 pasien. Pemilahan tersebut terjadi secara terus-menerus sampai tidak dapat dipilah lagi dan menghasilkan simpul terminal yakni simpul akhir dari hasil pemilahan.

Jumlah simpul terminal yang dihasilkan oleh pohon klasifikasi optimal sama dengan pohon klasifikasi maksimal yakni sebanyak 15 simpul dengan kedalaman (*depth*) sebesar 8. Kemudian setiap simpul yang dihasilkan akan diberi label untuk mengetahui karakteristik dari masing-masing tipe (kelas) pada variabel respon. Simpul terminal yang berwarna merah merupakan klasifikasi untuk penderita retardasi mental tipe ringan (tipe 1),

sedangkan untuk simpul terminal yang berwarna biru muda merupakan klasifikasi untuk penderita retardasi mental tipe sedang (tipe 2), dan untuk simpul terminal yang berwarna biru tua merupakan klasifikasi untuk penderita retardasi mental tipe berat (tipe 3).

Masing-masing simpul terminal yang dihasilkan oleh pohon klasifikasi optimal memiliki karakteristik tertentu berdasarkan kelas pada variabel respon yang telah diprediksi. Pelabelan kelas dilakukan untuk mengetahui karakteristik yang dihasilkan. Berikut ini merupakan pemberian label kelas untuk setiap simpul terminal. Berdasarkan Tabel 4.6, dapat diketahui jumlah simpul terminal untuk kelas 1 (penderita retardasi mental ringan) adalah sebanyak 7 simpul, sedangkan jumlah simpul terminal untuk kelas 2 (penderita retardasi mental sedang) berjumlah 5 simpul, dan untuk kelas 3 (penderita retardasi mental berat) jumlah simpul terminal yang dihasilkan adalah sebanyak 3 simpul.

Tabel 4.6 Pelabelan Kelas Simpul Terminal dengan Pemilah Indeks Gini

Kelas	Simpul Terminal	Persentase (%)
RM Ringan (7 simpul terminal)	1	100
	3	100
	4	77,8
	6	60
	8	83,3
	9	87,5
	10	100
RM Sedang (5 simpul terminal)	2	55,6
	5	83,3
	7	100
	11	90,6
	12	87,5
RM Berat (3 simpul terminal)	13	75
	14	80
	15	100

Karakteristik pada masing-masing kelas pada variabel respon dapat diketahui dengan menelusuri pohon klasifikasi

optimal yang sudah terbentuk berdasarkan aturan pemilahan indeks gini, pohon klasifikasi optimal yang terbentuk bisa dilihat pada Lampiran G. Karakteristik yang kuat didapatkan pada simpul terminal yang memiliki persentase kelas terbesar (100%) pada pelabelan kelas yakni untuk karakteristik penderita retardasi mental ringan dapat dilihat pada simpul terminal 1, sedangkan untuk melihat karakteristik penderita RM sedang dapat dilihat pada simpul terminal 7, dan untuk melihat karakteristik penderita RM berat dapat dilihat pada simpul terminal 15. Berikut ini karakteristik yang dihasilkan oleh pasien retardasi mental pada masing-masing kelas.

Tabel 4.7 Karakteristik Kelas Penderita RM Berdasarkan Simpul Terminal dengan Persentase Terbesar 1

Kelas	Simpul Terminal	Karakteristik
RM Ringan	ST-1	Penderita retardasi mental yang berumur <8 tahun, dengan kondisi dilahirkan normal (tidak <i>premature</i>), gejala psikiatrik yang ditunjukkan adalah sindroma otak menahun dengan gangguan tingkah laku dan tidak ada gangguan psikiatrik
RM Sedang	ST-7	Penderita retardasi mental yang berumur >9 tahun, dengan kondisi dilahirkan normal (tidak <i>premature</i>), dengan gejala psikiatrik yang menunjukkan tingkah laku adanya gangguan psikiatrik dan reaksi penyesuaian diri, tidak ada faktor genetik/keturunan, faktor etiologik banyak disebabkan oleh intoksikasi dan trauma
RM Berat	ST-15	Penderita retardasi mental dengan gejala psikiatrik yang menunjukkan adanya gangguan psikiatrik dan faktor etiologik banyak disebabkan oleh intoksikasi

Pohon klasifikasi optimal berguna untuk mengklasifikasikan pasien RM baru dengan cara menelusuri pohon klasifikasi dari atas ke bawah. Berikut ini ilustrasi dalam mengklasifikasikan pasien RM baru untuk mengetahui tipe RM yang diderita oleh pasien

berdasarkan pohon klasifikasi optimal pada Gambar 4.12. Misalnya, terdapat 10 pasien baru berdasarkan catatan rekam medis yang dimiliki oleh pasien diperoleh karakteristik pasien yang ditunjukkan pada Tabel 4.8 berikut ini.

Tabel 4.8 Ilustrasi Data Rekam Medis Pasien Baru

Pasien	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
P1	2	2	3	1	2	4	5
P2	4	2	13	1	2	5	4
P3	1	2	17	1	2	3	3
P4	2	2	2	1	1	5	5
P5	1	2	10	1	2	1	4
P6	1	2	9	2	1	3	2
P7	1	1	15	2	2	4	1
P8	1	1	13	2	2	3	3
P9	1	1	16	2	2	4	3

Dengan melakukan penelusuran pada pohon klasifikasi yang dihasilkan dengan menggunakan indeks gini, maka hasil pengklasifikasian pasien RM baru adalah sebagai berikut.

1. Untuk pasien pertama (P1), apakah pasien tersebut memiliki gejala psikiatrik (X₇) yang menunjukkan sindroma otak menahun dengan gangguan tingkah laku (kategori 1), psikosa (kategori 2), reaksi penyesuaian diri (kategori 4), atau tidak ada gangguan psikiatrik (kategori 5)? Jika iya, maka lihat kondisi pasien tersebut saat dilahirkan (X₂). Apakah pasien tersebut dilahirkan dengan kondisi normal (tidak *premature*)? Jika iya, maka lihat kembali gejala psikiatrik (X₇) yang ditunjukkan. Apakah pasien tersebut memiliki gejala psikiatrik (X₇) yang menunjukkan sindroma otak menahun dengan gangguan tingkah laku (kategori 1) atau tidak ada gangguan psikiatrik (kategori 5)? Jika iya, maka lihat umur pasien (X₃) tersebut. Apakah pasien memiliki umur ≤ 9 tahun? Jika iya maka pasien tersebut diklasifikasikan sebagai penderita RM ringan.
2. Untuk pasien kedua (P2), apakah pasien tersebut memiliki gejala psikiatrik (X₇) yang menunjukkan sindroma otak menahun dengan gangguan tingkah laku (kategori 1),

psikosa (kategori 2), reaksi penyesuaian diri (kategori 4), atau tidak ada gangguan psikiatrik (kategori 5)? Jika iya, maka lihat kondisi pasien tersebut saat dilahirkan (X_2). Apakah pasien tersebut dilahirkan dengan kondisi normal (tidak *premature*)? Jika iya, maka lihat kembali gejala psikiatrik (X_7) yang ditunjukkan. Apakah pasien tersebut memiliki gejala psikiatrik (X_7) yang menunjukkan sindroma otak menahun dengan gangguan tingkah laku (kategori 1) atau tidak ada gangguan psikiatrik (kategori 5)? Jika tidak, maka lihat urutan pasien dilahirkan (X_1) tersebut. Apakah pasien memiliki umur ≤ 9 tahun? Jika tidak, maka lihat faktor genetik pasien (X_5). Apakah pasien tersebut tidak memiliki faktor genetik (kategori 2)? Jika iya, maka lihat faktor etiologik pasien (X_6). Apakah pasien tersebut memiliki faktor etiologik yang disebabkan oleh infeksi (kategori 1), gangguan metabolisme, pertumbuhan, dan nutrisi (kategori 4), atau pengaruh prenatal (kategori 5)? Jika iya, maka lihat kembali umur pasien (X_3). Apakah pasien memiliki umur pasien $\leq 10,7$ tahun? Jika tidak, maka pasien tersebut diklasifikasikan sebagai penderita RM ringan.

3. Untuk pasien ketiga (P3), apakah pasien tersebut memiliki gejala psikiatrik (X_7) yang menunjukkan sindroma otak menahun dengan gangguan tingkah laku (kategori 1), psikosa (kategori 2), reaksi penyesuaian diri (kategori 4), atau tidak ada gangguan psikiatrik (kategori 5)? Jika iya, maka lihat faktor etiologik pasien (X_6). Apakah pasien memiliki faktor etiologik yang disebabkan oleh infeksi (kategori 1), trauma (kategori 3), gangguan metabolisme, pertumbuhan, dan nutrisi (kategori 4), atau pengaruh prenatal (kategori 5)? Jika iya, maka lihat umur pasien (X_3). Apakah pasien memiliki umur $\leq 15,5$ tahun? Jika tidak maka pasien tersebut diklasifikasikan sebagai pasien penderita RM berat.

Begitu seterusnya untuk pasien keempat sampai dengan pasien terakhir. Berikut ini merupakan hasil klasifikasi pasien baru

yang menunjukkan tipe RM yang diderita oleh pasien yang nantinya akan menjadi variabel respon (kelas) yang dihasilkan pada Tabel 4.9.

Tabel 4.9 Hasil Klasifikasi Pasien RM Baru

Pasien	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	Tipe RM (Y)
P1	2	2	3	1	2	4	5	RM Ringan
P2	4	2	13	1	2	5	4	RM Ringan
P3	1	2	17	1	2	3	3	RM Berat
P4	2	2	2	1	1	5	5	RM Ringan
P5	1	2	10	1	2	1	4	RM Ringan
P6	1	2	9	2	1	3	2	RM Ringan
P7	1	1	15	2	2	4	1	RM Sedang
P8	1	1	13	2	2	3	3	RM Berat
P9	1	1	16	2	2	4	3	RM Berat

d. Ketepatan Klasifikasi Pohon CART dengan Pemilah Indeks Gini

Hasil klasifikasi penderita retardasi mental di (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur berdasarkan masing-masing kelas menggunakan metode *10-fold cross-validation* diukur berdasarkan total ketepatan klasifikasi, akurasi pada masing-masing kelas, dan total kesalahan klasifikasi untuk data *learning* dan data *testing*. Berikut adalah hasil ketepatan klasifikasi penderita RM untuk data *learning* dan *testing* sebelum dilakukan *pre-processing* dengan menggunakan SMOTE.

Tabel 4.10 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data *Learning* Sebelum SMOTE

Fold	Akurasi Setiap Kelas			Total Akurasi
	RM Ringan	RM Sedang	RM Berat	
1	0,953	0,724	0,733	0,839
2	0,860	0,667	0,688	0,782
3	0,976	0,714	0,647	0,828
4	0,953	0,724	0,667	0,828
5	1,000	0,719	0,667	0,839
6	0,913	0,720	0,688	0,816
7	0,977	0,733	0,786	0,862
8	0,913	0,692	0,688	0,807

Tabel 4.10 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data *Learning* Sebelum SMOTE (Lanjutan)

<i>Fold</i>	Akurasi Setiap Kelas			Total Akurasi
	RM Ringan	RM Sedang	RM Berat	
9	0,956	0,714	0,733	0,841
10	0,896	0,696	0,647	0,795
<i>Mean</i>	0,940	0,710	0,694	0,824
<i>Stdev</i>	0,041	0,019	0,042	0,023

Berdasarkan Tabel 4.10, dapat diketahui rata-rata total akurasi data *learning* dengan metode *10-fold cross-validation* sebelum dilakukan *pre-processing* dengan SMOTE adalah sebesar 82,4%. Namun, jika dilihat dari akurasi untuk masing-masing kelas akurasi untuk penderita RM ringan sangat tinggi yakni sebesar 94%, sedangkan akurasi untuk penderita RM berat sangat kecil yakni sebesar 69,4%. Hal ini menunjukkan pengklasifikasian penderita RM berat yang dihasilkan kurang akurat. Sedangkan jika dilihat dari nilai deviasi standar yang cukup kecil menunjukkan bahwa akurasi yang dihasilkan oleh *10-fold* sudah cukup stabil. Untuk mengetahui seberapa layak pohon klasifikasi yang terbentuk maka dapat diuji dengan menggunakan data *testing* sebelum dilakukan *pre-processing* dengan menggunakan metode SMOTE sebagai berikut.

Tabel 4.11 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data *Testing* Sebelum SMOTE

Kelas Observasi	Kelas Aktual			Total	Akurasi	APER	1-APER
	Ringan	Sedang	Berat				
Ringan	39	16	2	57	0,684		
Sedang	12	12	4	28	0,429	0,361	0,639
Berat	0	1	11	12	0,917		
Total	51	29	17	97			

Berdasarkan Tabel 4.11, dapat diketahui total akurasi data *testing* sebelum dilakukan *pre-processing* dengan menggunakan metode SMOTE adalah sebesar 63,9%. Total akurasi yang dihasilkan sangat kecil, sehingga pohon klasifikasi yang sudah terbentuk dengan menggunakan data *learning* kurang layak untuk

digunakan dalam pengklasifikasian penderita RM di RSJ Menur. Total akurasi yang kecil tersebut disebabkan karena jumlah data untuk masing-masing kategori pada variabel respon tidak seimbang sehingga terjadi kasus *underfitting*. Oleh sebab itu, diperlukan metode *pre-processing* dengan menggunakan metode SMOTE untuk menyeimbangkan jumlah kategori pada variabel respon dengan harapan total akurasi yang dihasilkan nantinya akan lebih tinggi. Berikut adalah hasil ketepatan klasifikasi penderita RM untuk data *learning* dan *testing* setelah dilakukan *pre-processing* dengan menggunakan SMOTE.

Tabel 4.12 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data *Learning* Setelah SMOTE

<i>Fold</i>	Akurasi Setiap Kelas			Total Akurasi
	RM Ringan	RM Sedang	RM Berat	
1	0,904	0,935	0,935	0,924
2	0,953	0,825	0,956	0,903
3	0,915	0,833	0,977	0,903
4	0,870	0,814	1,000	0,883
5	0,932	0,807	0,955	0,889
6	0,909	0,821	0,978	0,897
7	0,846	0,875	0,956	0,889
8	0,855	0,918	0,976	0,910
9	0,896	0,836	0,976	0,897
10	0,913	0,868	0,957	0,910
<i>Mean</i>	0,899	0,853	0,966	0,901
<i>Stdev</i>	0,032	0,042	0,017	0,012

Berdasarkan Tabel 4.12. dapat diketahui rata-rata total ketepatan klasifikasi data *learning* setelah dilakukan *pre-processing* dengan menggunakan SMOTE berdasarkan pendekatan dengan metode *10-fold cross validation* lebih tinggi yakni sebesar 90,1%, artinya penderita RM tepat diklasifikasikan sesuai dengan masing-masing tipe RM yang telah diprediksi sebelumnya adalah sebesar 90,1% (Lampiran E). Selain itu, berdasarkan Tabel 4.12 dapat dilihat akurasi untuk masing-masing kelas sudah cukup

tinggi. Selain itu, kestabilan yang dihasilkan berdasarkan akurasi masing-masing *fold* sudah cukup baik dibuktikan dari nilai deviasi standar yang kecil. Dalam perhitungan akurasi untuk setiap *fold* dapat dilihat dari persentase simpul terminal yang tepat diklasifikasikan atau yang mengalami kesalahan klasifikasi (Lampiran C).

Setelah diketahui hasil klasifikasi data *learning*, selanjutnya dilakukan validasi untuk mengetahui seberapa layak pohon klasifikasi menggunakan data *learning* yang sudah terbentuk. Uji kelayakan pohon klasifikasi dilakukan dengan menggunakan data *testing* setelah dilakukan *pre-processing* dengan menggunakan SMOTE. Berikut ini merupakan hasil klasifikasi penderita RM menggunakan data *testing* yang ditunjukkan pada Tabel 4.13.

Tabel 4.13 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Gini Data *Testing* Setelah SMOTE

Kelas Observasi	Kelas Aktual			Total	Akurasi	APER	1-APER
	Ringan	Sedang	Berat				
Ringan	35	21	1	57	0,614		
Sedang	14	38	4	56	0,679	0,273	0,727
Berat	2	2	44	48	0,917		
Total	51	61	49	161			

Berdasarkan Tabel 4.13, dapat diketahui sebanyak 57 penderita RM yang termasuk kategori kelas 1 (RM ringan) yang tepat diklasifikasikan sebagai penderita RM ringan adalah sebanyak 35, sedangkan sebanyak 21 penderita RM ringan salah diklasifikasikan menjadi penderita RM sedang, dan sebanyak 1 penderita RM ringan menjadi penderita RM berat. Sehingga, akurasi untuk kelas 1 (RM ringan) adalah sebesar 61,4%. Sedangkan, dari 56 data yang diklasifikasikan menjadi kelas 2 (RM sedang) yang tepat diklasifikasikan sebagai penderita RM sedang adalah sebanyak 38 pasien, sebanyak 14 pasien mengalami kesalahan pengklasifikasian menjadi penderita RM tipe ringan, dan sebanyak 4 pasien mengalami kesalahan pengklasifikasian menjadi penderita RM berat. Sehingga, akurasi yang dihasilkan untuk kelas 2 (RM sedang) adalah sebesar 67,9%. Untuk kelas 3 (RM berat)

dengan jumlah data sebanyak 48 data, terdapat 44 pasien yang tepat diklasifikasikan menjadi penderita RM berat, sebanyak 2 pasien mengalami kesalahan pengklasifikasian menjadi penderita RM ringan, dan sebanyak 2 pasien mengalami kesalahan pengklasifikasian menjadi penderita RM berat. Sehingga, ketepatan klasifikasi untuk kelas 3 (RM berat) adalah sebesar 91,7%.

Berdasarkan Tabel 4.13, dapat diketahui total akurasi (1-APER) yang dihasilkan untuk klasifikasi penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 dengan menggunakan pemilihan pemilah indeks gini adalah sebesar 72,7%, sedangkan total kesalahan klasifikasi yang dihasilkan adalah sebesar 27,3% (Lampiran E). Persentase total ketepatan klasifikasi setelah dilakukan *pre-processing* dengan *SMOTE* yang dihasilkan lebih tinggi, sehingga pohon klasifikasi optimal yang dihasilkan sudah layak dan dapat digunakan untuk mengklasifikasikan pasien RM yang baru.

4.3.3 Analisis CART Penderita RM di RSJ Menur dengan Pemilahan Pemilah Indeks Twoing

Tahapan analisis CART dengan menggunakan metode pemilahan indeks twoing sama dengan metode pemilahan indeks gini, dimulai pada pembentukan pohon klasifikasi maksimal, pemangkasan pohon klasifikasi maksimal, kemudian dilanjutkan dengan pemilihan pohon klasifikasi optimal, dan kemudian menghitung ketepatan klasifikasi dari pohon klasifikasi optimal yang telah terbentuk. Berikut ini tahapan analisis CART menggunakan metode pemilahan indeks twoing untuk klasifikasi penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014.

a. Pembentukan Pohon Klasifikasi Maksimal

Sebelum melakukan pembentukan pohon klasifikasi, terlebih dahulu perlu dilakukan penentuan variabel terpenting, yakni variabel yang memberikan kontribusi terbesar dalam pembentukan pohon klasifikasi. Variabel terpenting yang berperan dalam pembentukan pohon klasifikasi maksimal akan ditunjukkan pada Tabel 4.14 berikut ini.

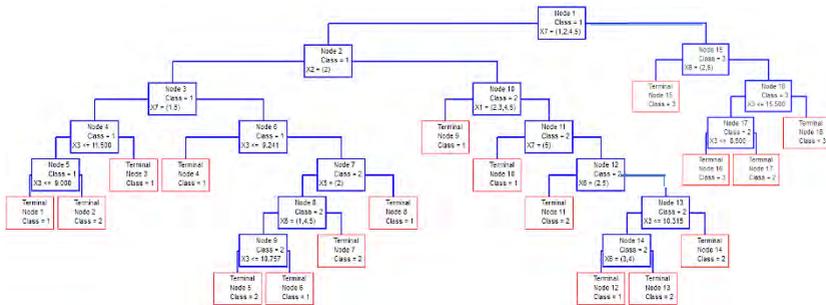
Tabel 4.14 Variabel Terpenting dari Pohon Klasifikasi Maksimal dengan Pemilahan Indeks Twoing

Variabel	Nama Variabel	Skor
X ₇	Gejala Psikiatrik	100,00
X ₆	Faktor Etiologik	81,99
X ₃	Umur Penderita	31,47
X ₂	Kondisi Dilahirkan	13,81
X ₁	Urutan Dilahirkan	12,45
X ₅	Faktor Genetik	5,56
X ₄	Jenis Kelamin	0,16

Berdasarkan Tabel 4.14 dapat diketahui variabel terpening dalam pembentukan pohon klasifikasi maksimal adalah variabel X₇ (gejala psikiatrik) ditunjukkan dengan skor kontribusi terbesar yakni sebesar 100. Sehingga dalam pembentukan pohon klasifikasi, variabel X₇ (gejala psikiatrik) akan berperan sebagai simpul utama yang akan memilah menjadi 2 simpul anak. Kemudian terdapat variabel lain yang juga berpengaruh terhadap pembentukan pohon klasifikasi yakni variabel X₆ (faktor etiologik) dengan kontribusi sebesar 81,99. Sedangkan variabel lainnya hanya memberikan skor kontribusi dibawah 50.

Dalam menentukan simpul utama yang digunakan sebagai pemilahan utama, ditentukan berdasarkan banyaknya nilai keheterogenan yang dapat direduksi oleh masing-masing variabel prediktor. Sehingga, pemilahan secara *recursive* diawali oleh pemilahan 2 simpul anak yakni simpul kiri dan simpul kanan oleh variabel X₇ (gejala psikiatrik) dengan ketentuan penderita retardasi mental dengan gejala psikiatrik yang menunjukkan tingkah laku sindroma otak menahun dengan gangguan tingkah laku (kategori 1), psikosa (kategori 2), reaksi penyesuaian diri (kategori 4), dan tidak ada gangguan psikiatrik (kategori 5) akan dipilah menjadi simpul kiri, sedangkan penderita retardasi mental dengan gejala psikiatrik yang menunjukkan tingkah laku seperti gangguan psikiatrik (kategori 3) akan dipilah menjadi simpul kanan (Lampiran K). kemudian 2 simpul anak tersebut akan memilah lagi secara *recursive* menjadi 2 simpul anak baru dan dihasilkan pohon klasifikasi maksimal.

Topology pohon klasifikasi maksimal dengan menggunakan pemilahan pemilah indeks twoing ditunjukkan pada Gambar 4.13. Jika dilihat, pohon klasifikasi maksimal dengan menggunakan pemilah indeks twoing menghasilkan 18 simpul terminal dengan kedalaman sebesar 8. Namun, pohon klasifikasi maksimal yang dihasilkan masih kompleks dalam artian simpul terminal yang dihasilkan belum homogen. Sehingga, perlu dilakukan pemangkasan pohon (*prunning*) guna memperoleh pohon klasifikasi yang optimal dengan menghasilkan simpul terminal yang homogen



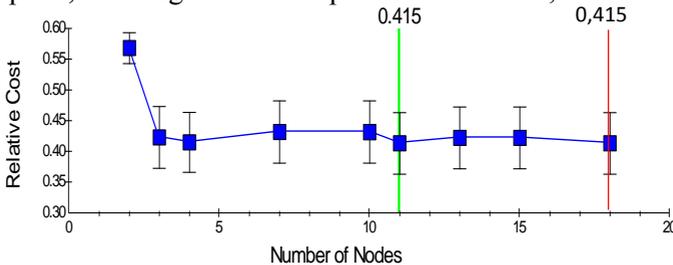
Gambar 4.13 Topology Pohon Klasifikasi Maksimal dengan Pemilahan Pemilah Indeks Twoing

b. Pemangkasan (*Prunning*) Pohon Klasifikasi Optimal

Metode pemangkasan pohon klasifikasi maksimal dengan pemilah indeks twoing sama dengan yang digunakan pada pemangkasan dengan pemilah indeks gini yakni menggunakan metode *10-fold cross validation estimate*. Untuk melihat perlu tidaknya dilakukan pemangkasan pohon, dapat dilihat pada nilai *relative cost* yang dihasilkan. Pemangkasan pohon dilakukan dengan menggunakan pendekatan *cost complexity minimum*. Sehingga pohon klasifikasi optimal yang digunakan adalah pohon klasifikasi dengan nilai *relative cost* yang minimum.

Berdasarkan Gambar 4.14, dapat diketahui jumlah pohon klasifikasi maksimal dengan jumlah simpul terminal sebanyak 18 yang diperoleh belum layak. Pohon klasifikasi maksimal ditunjukkan oleh garis merah dengan nilai *relative cost* sebesar 0,415. Kemudian, diperoleh pohon klasifikasi optimal dengan

jumlah simpul terminal yang lebih sedikit yakni sebanyak 11 simpul terminal yang ditunjukkan oleh garis berwarna hijau. Nilai *relative cost* yang dihasilkan pohon klasifikasi optimal sama dengan nilai *relative cost* pada pohon klasifikasi maksimal yakni sebesar 0,415. Sedangkan, biaya kesalahan (*cross-validation relative cost*) yang dihasilkan merupakan nilai yang paling minimum yakni sebesar $0,415 \pm 0,050$ atau bekisar antara 0,465 sampai 0,365 dengan nilai kompleksitas sebesar 0,003.



Gambar 4.14 Plot *Relative Cost* dalam Klasifikasi Penderita Retardasi Mental dengan pemilah Indeks Twoing

Urutan pembentukan pohon klasifikasi (*tree sequence*) dimulai dengan terbentuknya pohon klasifikasi maksimal kemudian dilakukan pemangkasan dan sampai didapatkan pohon klasifikasi optimal yang ditampilkan pada Tabel 4.15 berikut ini.

Tabel 4.15 Urutan Pembentukan Pohon Klasifikasi (*Tree Sequence*) 2

Pohon Klasifikasi	<i>Terminal Nodes</i>	<i>Cross-Validated Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity</i>
1	18	$0,415 \pm 0,050$	0,143	0,000
2	15	$0,423 \pm 0,050$	0,152	0,002
3	13	$0,423 \pm 0,050$	0,160	0,003
4**	11	$0,415 \pm 0,050$	0,170	0,003
5	10	$0,433 \pm 0,051$	0,187	0,012
6	7	$0,433 \pm 0,051$	0,241	0,012
7	4	$0,416 \pm 0,049$	0,302	0,014
8	3	$0,424 \pm 0,050$	0,355	0,035
9	2	$0,570 \pm 0,025$	0,518	0,108
10	1	$1,000 \pm 0,000$	1,000	0,322

** pohon klasifikasi optimal

Berdasarkan pohon klasifikasi maksimal yang ditunjukkan pada Gambar 4.13, perlu dilakukan pemangkasan sampai didapatkan pohon klasifikasi optimal dengan jumlah simpul terminal sebanyak 11 *nodes*. Setelah dilakukan pemangkasan pohon klasifikasi optimal, langkah selanjutnya adalah menentukan pohon klasifikasi optimal untuk klasifikasi penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 dengan menggunakan pemilihan pemilah indeks twoing.

c. Pemilihan Pohon Klasifikasi Optimal

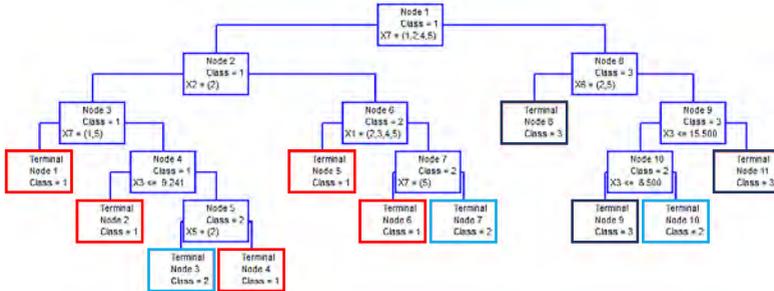
Sebelum dilakukan pemilihan pohon klasifikasi optimal dengan menggunakan metode pemilahan indeks twoing, terlebih dahulu perlu ditentukan variabel-variabel terpenting yang memberikan kontribusi atau peranan terbesar dalam pembentukan pohon klasifikasi. Berikut ini akan ditunjukkan variabel terpenting dalam pembentukan pohon klasifikasi optimal pada Tabel 4.16 sebagai berikut.

Tabel 4.16 Variabel Terpenting dalam Pembentukan Pohon Klasifikasi Optimal

Variabel	Nama Variabel	Skor
X ₇	Gejala Psikiatrik	100,00
X ₆	Faktor Etiologik	77,81
X ₃	Umur Penderita	22,66
X ₂	Kondisi Dilahirkan	13,95
X ₁	Urutan Dilahirkan	12,16
X ₅	Faktor Genetik	5,38
X ₄	Jenis Kelamin	0,01

Berdasarkan Tabel 4.16, dapat diketahui variabel terpenting yang digunakan dalam pemilahan pada simpul induk adalah variabel X₇ (gejala psikiatrik) dengan skor kontribusi tertinggi yakni sebesar 100. Sedangkan variabel lain yang penting dalam pemilahan adalah variabel X₆ (faktor etiologik) dengan skor kontribusi sebesar 77,81. Sedangkan, variabel lainnya hanya memberikan skor kontribusi dibawah 50. Proses pemilahan berdasarkan variabel terpenting dengan menggunakan metode

pemilah indeks twoing akan digambarkan secara visual dengan pohon klasifikasi optimal sebagai berikut pada Gambar 4.15 sebagai berikut.



Gambar 4.15 Topology Pohon Klasifikasi Optimal dengan Pemilahan Pemilah Indeks Twoing

Pohon klasifikasi optimal analisis CART untuk klasifikasi penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 menghasilkan jumlah simpul terminal sebanyak 11 simpul dengan tingkat kedalaman (*depth*) 6. Simpul yang dihasilkan oleh pohon klasifikasi optimal akan diberi label kelas untuk mengetahui karakteristik dari masing-masing tipe (kelas) pada variabel respon.

Berdasarkan Gambar 4.15 dapat diketahui, simpul terminal dengan warna merah merupakan label kelas untuk penderita retardasi mental dengan tipe ringan (tipe 1), sedangkan untuk simpul terminal yang berwarna biru muda merupakan klasifikasi untuk penderita retardasi mental tipe sedang (tipe 2), dan untuk simpul terminal yang berwarna biru tua merupakan klasifikasi untuk penderita retardasi mental tipe berat (tipe 3). Pemberian label pada simpul terminal berguna untuk memudahkan dalam mengetahui karakteristik atau sifat yang dihasilkan oleh masing-masing tipe penderita retardasi mental di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014. Pemberian label untuk masing-masing simpul terminal dilakukan dengan melihat persentase kelas terbesar yang terdapat pada simpul terminal tersebut.

Jumlah simpul terminal untuk kelas 1 (penderita retardasi mental ringan) adalah sebanyak 5 simpul, sedangkan jumlah simpul terminal untuk kelas 2 (penderita retardasi mental sedang) adalah 3 simpul, dan untuk kelas 3 (penderita retardasi mental berat) jumlah simpul terminal yang dihasilkan adalah sebanyak 3 simpul. Kemudian setelah dilakukan pelabelan kelas pada masing-masing simpul terminal, langkah selanjutnya adalah menelusuri pohon klasifikasi optimal berdasarkan simpul terminal pada masing-masing kelas untuk mengetahui karakteristik atau sifat yang terdapat pada masing-masing kategori atau tipe penderita retardasi mental. Pohon klasifikasi optimal dengan menggunakan pemilahan indeks twoing dapat dilihat pada Lampiran L. Pelabelan kelas untuk masing-masing simpul terminal yang dihasilkan oleh pohon klasifikasi optimal dengan metode pemilahan indeks twoing dapat dilihat pada Tabel 4.17.

Tabel 4.17 Pelabelan Kelas Simpul Terminal dengan Pemilah Indeks Twoing

Kelas	Simpul Terminal	Persentase (%)
RM Ringan (5 simpul terminal)	1	84,4
	2	77,8
	4	83,3
	5	87,5
	6	100
RM Sedang (3 simpul terminal)	3	71,4
	7	90,6
	10	88,9
RM Berat (3 simpul terminal)	8	100
	9	66,7
	11	80

Karakteristik pasien penderita retardasi mental dapat dilihat pada simpul terminal yang memiliki persentase kelas terbesar. Berdasarkan pelabelan kelas yang sudah dilakukan simpul terminal yang memiliki persentase terbesar untuk setiap kelas pada masing-masing variabel respon adalah simpul terminal 6 yang dilabelkan kelas 1 (RM ringan) dengan persentase 100%, simpul terminal 7 yang dilabelkan kelas 2 (RM sedang) dengan persentase 90,6%,

dan simpul terminal 8 yang dilabelkan kelas 3 (RM berat) dengan persentase 100%. Sehingga dengan menelusuri aturan pemilahan pada simpul terminal 6, 7, dan 8 akan didapatkan karakteristik masing-masing tipe penderita retardasi mental di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada Tabel 4.18 sebagai berikut.

Tabel 4.18 Karakteristik Kelas Penderita RM Berdasarkan Simpul Terminal dengan Persentase Terbesar 2

Kelas	Simpul Terminal	Karakteristik
RM Ringan	ST-6	Penderita retardasi mental dengan kondisi saat dilahirkan <i>premature</i> , merupakan anak pertama dan memiliki gejala psikiatrik atau tingkan laku yang ditunjukkan adalah tidak adanya gangguan psikiatrik (kategori 5)
RM Sedang	ST-7	Penderita retardasi mental dengan kondisi dilahirkan <i>premature</i> , merupakan anak pertama, dengan gejala psikiatrik atau tingkah laku yang ditimbulkan adalah sindroma otak menahun dengan gangguan tingkah laku, psikosa, dan reaksi penyesuaian diri
RM Berat	ST-8	Penderita retardasi mental dengan gejala psikiatrik yang menunjukkan tingkah laku seperti adanya gangguan psikiatrik, dengan faktor etiologik atau penderita disebabkan oleh intoksikasi (keracunan zat kimia) dan adanya pengaruh prenatal

Untuk mengklasifikasikan pasien RM baru dapat digunakan pohon klasifikasi optimal yakni dengan cara menelusuri pohon klasifikasi dari atas ke bawah. Berikut ini ilustrasi dalam mengklasifikasikan pasien RM baru untuk mengetahui tipe RM yang diderita oleh pasien berdasarkan pohon klasifikasi optimal pada Gambar 4.15. Apakah pasien tersebut meunjukkan gejala psikiatrik (X₇) seperti sindroma otak menahun dengan gangguan tingkah laku (kategori 1), psikosa (kategori 2), reaksi penyesuaian diri (kategori 4), dan tidak ada gangguan psikiatrik (kategori 5)? Jika iya, maka

lihat kondisi pasien tersebut saat dilahirkan (X_2). Apakah pasien tersebut dilahirkan dengan kondisi normal atau tidak *premature* (kategori 2)? Jika iya, maka lihat kembali gejala psikiatrik (X_7) yang ditunjukkan. Apakah pasien tersebut memiliki gejala psikiatrik (X_7) yang menunjukkan sindroma otak menahun dengan gangguan tingkah laku (kategori 1) dan tidak ada gangguan psikiatrik (kategori 5)? Jika iya, maka pasien tersebut diklasifikasikan sebagai penderita RM ringan (kelas 1). Untuk lebih jelasnya dapat dilihat pada Lampiran L.

d. Ketepatan Klasifikasi Pohon CART dengan Pemilah Indeks Twoing

Setelah terbentuk pohon klasifikasi optimal, langkah selanjutnya adalah menghitung ketepatan klasifikasi atau akurasi dari hasil klasifikasi penderita retardasi mental di (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 menggunakan *10-fold cross-validation* untuk data *learning* dan *testing*. Hasil klasifikasi berdasarkan data *learning* sebelum dilakukan *pre-processing* dengan SMOTE dapat ditunjukkan pada Tabel 4.19 berikut ini.

Tabel 4.19 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Twoing Data *Learning* Sebelum SMOTE

<i>Fold</i>	Akurasi Setiap Kelas			Total Akurasi
	RM Ringan	RM Sedang	RM Berat	
1	0,953	0,704	0,647	0,816
2	0,860	0,667	0,688	0,782
3	0,976	0,714	0,647	0,828
4	0,896	0,696	0,647	0,795
5	1,000	0,710	0,625	0,828
6	0,913	0,720	0,688	0,816
7	0,977	0,733	0,786	0,862
8	0,913	0,692	0,688	0,807
9	0,956	0,714	0,733	0,841
10	0,896	0,696	0,647	0,795
<i>Mean</i>	0,934	0,705	0,679	0,817
<i>Sidev</i>	0,043	0,017	0,046	0,023

Berdasarkan Tabel 4.19, dapat diketahui rata-rata total akurasi data *learning* dengan metode pemilah indeks twoing sebelum dilakukan *pre-processing* dengan SMOTE adalah sebesar 81,7%. Namun, jika dilihat dari akurasi untuk masing-masing kelas akurasi yang dihasilkan belum cukup baik, seperti akurasi yang dihasilkan untuk penderita RM berat hanya sebesar 67,9%. Sedangkan, jika dilihat dari nilai deviasi standar yang dihasilkan cukup kecil, hal ini menunjukkan bahwa akurasi yang dihasilkan oleh metode *10-fold cross validation* sudah cukup stabil. Untuk mengetahui seberapa layak pohon klasifikasi yang terbentuk maka dapat diuji dengan menggunakan data *testing* sebelum dilakukan *pre-processing* dengan menggunakan metode SMOTE sebagai berikut.

Tabel 4.20 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Twoing Data *Testing* Sebelum SMOTE

Kelas Observasi	Kelas Aktual			Total	Akurasi	APER	1-APER
	Ringan	Sedang	Berat				
Ringan	34	21	2	57	0,596		
Sedang	13	11	4	28	0,393	0,412	0,588
Berat	0	0	12	12	1		
Total	47	32	18	97			

Berdasarkan Tabel 4.20, dapat diketahui total akurasi data *testing* dengan metode pemilah indeks twoing sebelum dilakukan *pre-processing* dengan menggunakan metode SMOTE adalah sebesar 58,8%. Sama seperti halnya dengan indeks gini, total akurasi yang dihasilkan dengan menggunakan indeks twoing sangat kecil. Hal ini, disebabkan oleh jumlah data untuk masing-masing kategori pada variabel respon tidak seimbang sehingga terjadi kasus *underfitting*. Sehingga pohon klasifikasi yang sudah terbentuk dengan menggunakan data *learning* kurang layak untuk digunakan dalam pengklasifikasian penderita RM di RSJ Menur. sehingga, diperlukan suatu metode *pre-processing* dengan menggunakan metode SMOTE untuk mendapatkan total akurasi yang lebih tinggi. Berikut adalah hasil ketepatan klasifikasi penderita RM untuk data *learning* dan *testing* setelah dilakukan *pre-processing* dengan menggunakan SMOTE.

Tabel 4.21 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Twoing Data *Learning* Setelah SMOTE

Fold	Akurasi Setiap Kelas			Total Akurasi
	RM Ringan	RM Sedang	RM Berat	
1	0,731	0,971	0,953	0,854
2	0,953	0,828	0,955	0,903
3	0,915	0,836	0,977	0,903
4	0,891	0,814	1,000	0,890
5	0,909	0,821	0,933	0,883
6	0,929	0,810	0,978	0,897
7	0,857	0,846	0,955	0,883
8	0,865	0,885	0,976	0,903
9	0,896	0,836	0,976	0,897
10	0,913	0,868	0,957	0,910
<i>Mean</i>	0,886	0,851	0,966	0,892
<i>Stdev</i>	0,058	0,045	0,018	0,015

Berdasarkan Tabel 4.21, total ketepatan klasifikasi penderita RM di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur berdasarkan rata-rata akurasi dari 10-*fold* adalah sebesar 89,2% (Lampiran J). Sedangkan, total akurasi untuk masing-masing kelas sudah cukup tinggi yakni total akurasi untuk penderita RM kelas 1 (RM ringan) adalah sebesar 88,6%, total akurasi untuk penderita RM kelas 2 (RM sedang) adalah sebesar 85,1%, dan total ketepatan klasifikasi untuk penderita RM kelas 3 (RM berat) adalah sebesar 96,6%. Akurasi yang dihasilkan setelah dilakukan *pre-processing* dengan SMOTE lebih tinggi. Berdasarkan 10 akurasi yang dihasilkan dapat dilihat nilai deviasi standar yang dihasilkan cukup kecil sehingga dapat dikatakan akurasi yang dihasilkan untuk setiap *fold* sudah stabil.

Untuk mengetahui kelayakan pohon klasifikasi yang dibentuk dengan menggunakan data *learning*, selanjutnya akan dilakukan validasi dengan menggunakan data *testing* setelah dilakukan *pre-processing* dengan SMOTE berdasarkan pemilihan pohon optimum dengan metode 10-*fold cross validation*. Hasil ketepatan klasifikasi dengan data *testing* dengan metode pemilah indeks twoing akan ditunjukkan pada Tabel 4.22 sebagai berikut.

Tabel 4.22 Ketepatan Klasifikasi Penderita RM dengan Pemilah Indeks Twoing Data *Testing* Setelah SMOTE

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	Ringan	Sedang	Berat				
Ringan	34	21	2	57	0,596		
Sedang	15	38	3	56	0,679	0,286	0,714
Berat	3	2	43	48	0,896		
Total	52	61	48	161			

Berdasarkan Tabel 4.22, dapat diketahui akurasi untuk kelas RM ringan (tipe 1) adalah sebesar 59,6% dengan keterangan penderita RM yang termasuk kategori RM ringan (kelas 1) yang tepat diklasifikasikan sebagai penderita RM ringan adalah sebanyak 34 pasien, sedangkan sebanyak 21 penderita RM ringan salah diklasifikasikan menjadi penderita RM sedang, dan sebanyak 2 penderita RM ringan salah diklasifikasikan menjadi penderita RM berat. Akurasi untuk kelas RM sedang (tipe 2) yang dihasilkan adalah sebesar 67,9% dengan ketentuan penderita RM sedang (tipe 2) yang tepat diklasifikasikan sebagai penderita RM sedang adalah sebanyak 38 pasien, sebanyak 15 pasien mengalami kesalahan pengklasifikasian menjadi penderita RM tipe ringan, dan sebanyak 3 pasien mengalami kesalahan pengklasifikasian menjadi penderita RM berat. Sedangkan, akurasi untuk kelas RM berat (tipe 3) cukup tinggi yakni sebesar 89,6% dengan ketentuan penderita RM berat Untuk (tipe 3) sebanyak 43 pasien tepat diklasifikasikan menjadi penderita RM berat, sebanyak 3 pasien mengalami kesalahan pengklasifikasian menjadi penderita RM ringan, dan sebanyak 2 pasien mengalami kesalahan pengklasifikasian menjadi penderita RM berat.

Berdasarkan akurasi pada masing-masing tipe atau kategori pada variabel respon didapatkan total akurasi (1-APER) yang dihasilkan untuk klasifikasi penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 dengan menggunakan pemilihan pemilah indeks twoing adalah sebesar 71,4% (Lampiran J). Sedangkan, total kesalahan klasifikasi (APER) yang dihasilkan adalah sebesar 28,6%. Akurasi klasifikasi penderita RM di RSJ Menur Surabaya

Provinsi Jawa Timur setelah dilakukan *pre-processing* dengan SMOTE lebih tinggi, sehingga dapat dikatakan bahwa pohon klasifikasi optimal yang dihasilkan dengan jumlah simpul terminal sebanyak 11 simpul sudah layak, sehingga pohon klasifikasi tersebut dapat digunakan untuk mengklasifikasikan penderita RM dengan menggunakan data yang baru.

4.3.4 Perbandingan Analisis CART Penderita RM di RSJ Menur dengan Pemilah Indeks Gini dan Indeks Twoing

Setelah diperoleh hasil klasifikasi penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur dengan menggunakan metode pemilahan indeks gini dan indeks twoing diperoleh total akurasi yang menunjukkan ketepatan klasifikasi. Total akurasi yang dihasilkan dan juga banyaknya simpul terminal yang dihasilkan pohon klasifikasi optimal dengan menggunakan metode pemilah indeks gini dan indeks twoing berbeda. Untuk mencari metode pemilah apa yang terbaik dapat dilihat dari nilai akurasi atau ketepatan klasifikasi yang paling tinggi dengan jumlah simpul terminal yang paling minimum. Perbandingan hasil klasifikasi penderita RM dengan menggunakan pemilah indeks gini dan indeks twoing dapat dilihat pada Tabel 4.23 sebagai berikut.

Tabel 4.23. Perbandingan Analisis CART dengan Pemilah Indeks Gini dan Indeks Twoing

Akurasi (%)	Indeks Gini	Indeks Twoing
<i>Learning</i>	90,1%	89,2%
<i>Testing</i>	72,7%	71,4%
Jumlah Simpul Terminal	15	11
Kedalaman (<i>depth</i>)	8	6

Jika dilihat berdasarkan Tabel 4.23 dapat diketahui, berdasarkan total akurasi yang dihasilkan oleh metode pemilihan pemilah indeks gini pohon klasifikasi optimal dengan menggunakan data *testing* adalah sebesar 72,7%. Sedangkan, total akurasi (1-APER) yang dihasilkan oleh metode pemilihan pemilah indeks twoing pada pohon klasifikasi optimal dengan menggunakan data

testing yakni sebesar 71,4%. Total akurasi hasil klasifikasi penderita RM di Rumah Sakit Jiwa Menur dengan menggunakan metode pemilihan pemilah indeks gini sedikit lebih tinggi dibandingkan dengan total akurasi menggunakan metode pemilihan pemilah indeks twoing.

Namun, jika dilihat berdasarkan banyaknya simpul terminal yang dihasilkan pada pohon klasifikasi optimal dengan menggunakan metode pemilah indeks gini lebih banyak yakni sebanyak 15 simpul. Sedangkan, simpul terminal yang dihasilkan oleh pohon klasifikasi optimal dengan metode pemilah indeks twoing adalah sebanyak 11 simpul. Tingkat kedalaman (*depth*) pohon klasifikasi optimal dengan indeks twoing lebih rendah yakni dengan kedalaman 6. Sehingga pohon klasifikasi optimal dengan menggunakan pemilah indeks twoing lebih *complex* dibandingkan dengan pohon klasifikasi optimal menggunakan pemilah indeks gini.

Sehingga metode pemilihan pemilah indeks gini dan indeks twoing dalam kasus pengklasifikasian penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur pada tahun 2013-2014 sama baiknya.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan untuk pengklasifikasian penderita retardasi mental (RM) di Rumah Sakit Jiwa Menur Provinsi Jawa Timur tahun 2013-2014 didapatkan kesimpulan sebagai berikut.

1. Jumlah pasien retardasi mental sebanyak 97 pasien dengan persentase penderita RM ringan sebesar 59%, penderita RM sedang sebesar 29%, dan penderita RM berat sebesar 12%. Setelah dilakukan *pre-processing* dengan menggunakan metode SMOTE diperoleh data sebanyak 161 data dengan persentase pada masing-masing kelas yang seimbang yakni jumlah pasien RM ringan sebesar 35%, jumlah pasien RM sedang sebesar 35%, dan jumlah pasien RM berat sebesar 30%.
2. Berdasarkan analisis CART, akurasi hasil pengklasifikasian penderita RM menggunakan pemilah indeks gini berdasarkan pohon klasifikasi optimal dengan variabel terpenting gejala psikiatrik adalah sebesar 72,7% dan simpul terminal yang dihasilkan sebanyak 15 simpul. Sedangkan, akurasi penderita RM menggunakan pemilah indeks twoing pada pohon klasifikasi optimal dengan variabel terpenting gejala psikiatrik adalah sebesar 71,4% dan simpul terminal yang dihasilkan sebanyak 11 simpul. Akurasi dengan menggunakan pemilah indeks gini sedikit lebih tinggi, namun pohon klasifikasi yang dihasilkan dengan pemilah indeks twoing lebih *complex*. Sehingga, kedua metode pemilah tersebut sama-sama baik digunakan untuk pengklasifikasian penderita RM di RSJ Menur.
3. Akurasi yang dihasilkan dengan menggunakan pemilah indeks gini lebih besar. Sehingga, karakteristik penderita RM yang dipilih adalah dengan menggunakan metode pemilah indeks gini. Karakteristik yang dihasilkan oleh pasien RM ringan adalah pasien RM yang berumur kurang dari 8 tahun, dengan

kondisi dilahirkan tidak *premature* (normal), gejala psikiatrik yang ditunjukkan yakni sindroma otak menahun dengan gangguan tingkah laku dan tidak adanya gangguan psikiatrik. Karakteristik pasien RM sedang adalah berumur lebih dari 9 tahun, dengan kondisi dilahirkan normal (tidak *premature*), dengan gejala psikiatrik yang menunjukkan tingkah laku adanya gangguan psikiatrik dan reaksi penyesuaian diri, tidak ada faktor genetik/keturunan, faktor etiologik banyak disebabkan oleh intoksikasi dan trauma. Sedangkan, karakteristik pasien RM berat memiliki gejala psikiatrik yang menunjukkan adanya gangguan psikiatrik dan faktor etiologik yang banyak disebabkan oleh intoksikasi tanpa memperhatikan umur pasien.

5.2 Saran

Rekomendasi atau saran yang dapat diberikan untuk penelitian selanjutnya adalah diperlukan metode alternatif dengan *ensemble* CART seperti metode *random forest*, atau dengan menggunakan metode klasifikasi lainnya seperti *support vector machine* (SVM), analisis diskriminan dan lain-lain untuk mengklasifikasikan penderita RM di RSJ Menur untuk meningkatkan akurasi. Sedangkan bagi pihak Rumah Sakit Jiwa Menur Provinsi Jawa Timur Surabaya untuk mengklasifikasikan penderita retardasi mental diperlukan beberapa variabel lainnya seperti tingkat kecerdasan (IQ) yang tidak didapatkan pada penelitian saat ini untuk memudahkan proses klasifikasi. Sehingga, hasil klasifikasi pasien akan menjadi lebih akurat.

DAFTAR LAMPIRAN

	Halaman
Lampiran A	Data Pengamatan Penderita Retardasi Mental 101
Lampiran B	Statistik Deskriptif Variabel Prediktor 102
Lampiran C	Output Pohon Klasifikasi <i>Cross-Validation 10-Fold</i> dengan Metode Pemilah Indeks Gini 104
Lampiran D	Output <i>Missclassification</i> dengan Metode Pemilihan Pemilah Indeks Gini 113
Lampiran E	Akurasi Hasil Klasifikasi dengan Pemilihan Pemilah Indeks Gini 114
Lampiran F	Informasi Simpul Terminal Pohon Klasifikasi dengan Pemilihan Pemilah Indeks Gini 116
Lampiran G	Pohon Klasifikasi Optimal dengan Menggunakan Pemilihan Pemilah Indeks Gini .. 125
Lampiran H	Output Pohon Klasifikasi <i>Cross-Validation 10-Fold</i> dengan Metode Pemilah Indeks Twoing 126
Lampiran I	Output <i>Missclassification</i> dengan Metode Pemilihan Pemilah Indeks Twoing 135
Lampiran J	Akurasi Hasil Klasifikasi dengan Pemilihan pemilah Indeks Twoing 135
Lampiran K	Informasi Simpul Terminal Pohon Klasifikasi dengan Pemilah Indeks Twoing 137
Lampiran L	Pohon Klasifikasi Optimal dengan Menggunakan Pemilah Indeks Twoing 144
Lampiran M	Surat Keterangan Penelitian di RSJ Menur Provinsi Jawa Timur 145

“Halaman ini sengaja dikosongkan”

LAMPIRAN

Lampiran A. Data Pengamatan Penderita Retardasi Mental Data Klasifikasi Penderita Retardasi Mental di Rumah Sakit Jiwa Menur Surabaya Provinsi Jawa Timur Tahun 2013-2014.

No.	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	Y
1	2	2	3	1	2	4	5	1
2	1	2	10	1	2	1	4	1
3	2	2	2	1	1	5	5	1
4	2	2	11	2	2	1	4	1
5	1	2	16	2	2	4	1	1
6	1	2	12	1	2	1	1	1
7	3	1	3	1	1	5	1	1
8	1	2	5	1	2	4	5	1
9	1	1	9	1	1	1	1	1
10	1	1	10	1	2	4	4	1
...
70	4	2	13	1	2	5	4	2
71	1	2	9	2	1	3	2	2
72	1	1	15	2	2	4	1	2
73	1	1	17	1	2	3	1	2
74	3	1	15	1	2	3	1	2
75	1	2	9	2	1	1	3	2
76*	2	2	12,58457	1	2	4	3	2
77*	1	2	10,65887	1	2	1	1	2
78*	1	1	16,98777	1	2	3	1	2
...
153	1	2	17	1	2	3	3	3
154	1	1	13	2	2	3	3	3
155	1	1	16	2	2	4	3	3
156	1	2	16	1	1	4	3	3
157	2	2	15	2	1	2	3	3
158*	1	1	10,30142	2	1	2	3	3
159*	2	1	5,077148	1	1	2	3	3
160*	1	1	16,34113	1	2	2	3	3
161*	2	1	5,048937	1	1	2	3	3

* Data SMOTE

Keterangan :

X₁ : urutan kelahiran anak

- X_2 : kondisi saat dilahirkan
 X_3 : umur penderita (tahun)
 X_4 : jenis kelamin
 X_5 : faktor genetik (keturunan)
 X_6 : faktor etiologik
 X_7 : gejala psikiatrik
 $Y(1)$: penderita retardasi mental ringan
 $Y(2)$: penderita retardasi mental sedang
 $Y(3)$: penderita retardasi mental berat

Lampiran B. Statistik Deskriptif Variabel Prediktor

Jenis_Kelamin * Tipe_RM Crosstabulation

Count	Tipe_RM			Total	
	1	2	3		
Jenis_Kelamin	1	39	18	7	64
	2	18	10	5	33
Total		57	28	12	97

Kondisi_Dilahirkan * Tipe_RM Crosstabulation

Count	Tipe_RM			Total	
	1	2	3		
Kondisi_Dilahirkan	1	13	13	7	33
	2	44	15	5	64
Total		57	28	12	97

Faktor_Etiologik * Tipe_RM Crosstabulation

Count	Tipe_RM			Total	
	1	2	3		
Faktor_Etiologik	2	13	5	1	19
	3	0	1	5	6
	4	13	10	2	25
	5	23	8	4	35
	5	8	4	0	12
Total		57	28	12	97

Gejala_Psikiatrik * Tipe_RM Crosstabulation

Count	Tipe_RM			Total	
	1	2	3		
Gejala_Psikiatrik	1	24	9	0	33
	2	5	5	0	10
	3	2	4	12	18
	4	17	10	0	27
	5	9	0	0	9
Total		57	28	12	97

Faktor Keturunan * Tipe RM Crosstabulation

Count	Tipe_RM			Total
	1	2	3	
Faktor_Keturunan	1	16	5	26
	2	41	23	71
Total		57	28	97

Urutan Dilahirkan * Tipe RM Crosstabulation

Count	Tipe_RM			Total
	1	2	3	
Urutan_Dilahirkan	1	30	19	57
	2	20	6	30
	3	4	1	5
	4	2	2	4
	5	1	0	1
Total		57	28	97

LEARNING SAMPLE VARIABLE STATISTICS

```
=====
                        CLASS
VARIABLE                1          2          3      OVERALL
-----
X3  MEAN|  10.789      12.158      13.278      12.007
     SD|   3.886       2.847       4.155       3.763
     N|   57.000      56.000      48.000     161.000
     SUM|  615.000     680.849     637.335     1933.184
X4  MEAN|   1.316       1.196       1.354       1.286
     SD|   0.469       0.401       0.483       0.453
     N|   57.000      56.000      48.000     161.000
     SUM|   75.000      67.000      65.000     207.000
X2  MEAN|   1.772       1.446       1.313       1.522
     SD|   0.423       0.502       0.468       0.501
     N|   57.000      56.000      48.000     161.000
     SUM|  101.000     81.000     63.000     245.000
X6  MEAN|   3.228       2.875       2.188       2.795
     SD|   1.363       1.294       0.607       1.230
     N|   57.000      56.000      48.000     161.000
     SUM|  184.000     161.000     105.000     450.000
X7  MEAN|   2.684       2.196       3.000       2.609
     SD|   1.627       1.299       0.000       1.271
     N|   57.000      56.000      48.000     161.000
     SUM|  153.000     123.000     144.000     420.000
X5  MEAN|   1.719       1.911       1.521       1.727
     SD|   0.453       0.288       0.505       0.447
     N|   57.000      56.000      48.000     161.000
     SUM|   98.000     107.000     73.000     278.000
X1  MEAN|   1.667       1.321       1.208       1.410
     SD|   0.893       0.690       0.410       0.729
     N|   57.000      56.000      48.000     161.000
     SUM|   95.000     74.000     58.000     227.000
KELAS MEAN|   1.000       2.000       3.000       1.944
     SD|   0.000       0.000       0.000       0.808
     N|   57.000      56.000      48.000     161.000
     SUM|   57.000     112.000     144.000     313.000
=====
```

AUTOMATIC LEVEL SETTINGS

NAME	LEVELS	MINIMUM
X4	2	1

X2	2	1
X6	5	1
X7	5	1
X5	2	1
X1	5	1

Lampiran C. Output Pohon Klasifikasi *Cross-Validation 10-Fold* dengan Metode Pemilah Indeks Gini

***** CROSS VALIDATION TREE 1 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	14	0.092	1	0.073	1	13	0.927	0.017
					2	1	0.073	
					3	0	0.000	
2	5	0.033	1	0.405	1	3	0.595	0.006
					2	2	0.405	
					3	0	0.000	
3	9	0.060	2	0.109	1	1	0.109	0.007
					2	8	0.891	
					3	0	0.000	
4	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
5	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
6	16	0.105	1	0.000	1	16	1.000	0.007
					2	0	0.000	
					3	0	0.000	
7	1	0.007	2	0.000	1	0	0.000	0.007
					2	1	1.000	
					3	0	0.000	
8	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
9	12	0.080	2	0.164	1	2	0.164	0.009
					2	10	0.836	
					3	0	0.000	
10	3	0.022	3	0.301	1	0	0.000	0.009
					2	1	0.301	
					3	2	0.699	
11	2	0.013	2	0.000	1	0	0.000	0.013
					2	2	1.000	
					3	0	0.000	
12	10	0.066	1	0.203	1	8	0.797	0.013
					2	2	0.203	
					3	0	0.000	
13	6	0.044	3	0.299	1	1	0.148	0.024
					2	1	0.151	
					3	4	0.701	
14	2	0.013	1	0.000	1	2	1.000	0.013

					2	0	0.000	
					3	0	0.000	
15	21	0.140	2	0.000	1	0	0.000	0.013
					2	21	1.000	
					3	0	0.000	
16	1	0.007	2	0.000	1	0	0.000	0.007
					2	1	1.000	
					3	0	0.000	
17	37	0.287	3	0.000	1	0	0.000	0.007
					2	0	0.000	
					3	37	1.000	

***** CROSS VALIDATION TREE 2 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	6	0.038	1	0.000	1	6	1.000	0.008
					2	0	0.000	
					3	0	0.000	
2	9	0.059	2	0.435	1	4	0.435	0.008
					2	5	0.565	
					3	0	0.000	
3	15	0.096	1	0.000	1	15	1.000	0.004
					2	0	0.000	
					3	0	0.000	
4	5	0.033	2	0.194	1	1	0.194	0.016
					2	4	0.806	
					3	0	0.000	
5	8	0.052	1	0.129	1	7	0.871	0.014
					2	1	0.129	
					3	0	0.000	
6	9	0.059	2	0.216	1	2	0.216	0.006
					2	7	0.784	
					3	0	0.000	
7	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
8	4	0.026	1	0.000	1	4	1.000	0.026
					2	0	0.000	
					3	0	0.000	
9	7	0.045	1	0.148	1	6	0.852	0.032
					2	1	0.148	
					3	0	0.000	
10	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
11	28	0.186	2	0.069	1	2	0.069	0.013
					2	26	0.931	
					3	0	0.000	
12	6	0.040	2	0.161	1	1	0.161	0.017
					2	5	0.839	
					3	0	0.000	
13	4	0.030	3	0.223	1	0	0.000	0.017

					2	1	0.223	
					3	3	0.777	
14	4	0.030	3	0.216	1	1	0.216	0.017
					2	0	0.000	
					3	3	0.784	
15	37	0.287	3	0.000	1	0	0.000	0.011
					2	0	0.000	
					3	37	1.000	

***** CROSS VALIDATION TREE 3 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	7	0.045	1	0.000	1	7	1.000	0.008
					2	0	0.000	
					3	0	0.000	
2	9	0.059	2	0.435	1	4	0.435	0.008
					2	5	0.565	
					3	0	0.000	
3	14	0.090	1	0.000	1	14	1.000	0.004
					2	0	0.000	
					3	0	0.000	
4	6	0.039	1	0.172	1	5	0.828	0.014
					2	1	0.172	
					3	0	0.000	
5	9	0.059	2	0.325	1	3	0.325	0.006
					2	6	0.675	
					3	0	0.000	
6	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
7	2	0.013	2	0.000	1	0	0.000	0.003
					2	2	1.000	
					3	0	0.000	
8	7	0.045	1	0.294	1	5	0.706	0.019
					2	2	0.294	
					3	0	0.000	
9	8	0.052	1	0.129	1	7	0.871	0.038
					2	1	0.129	
					3	0	0.000	
10	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
11	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
12	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
13	9	0.060	2	0.107	1	1	0.107	0.006
					2	8	0.893	
					3	0	0.000	

14	18	0.120	2	0.000	1	0	0.000	0.003
					2	18	1.000	
					3	0	0.000	
15	7	0.046	2	0.138	1	1	0.138	0.017
					2	6	0.862	
					3	0	0.000	
16	4	0.030	3	0.223	1	0	0.000	0.017
					2	1	0.223	
					3	3	0.777	
17	4	0.031	3	0.000	1	0	0.000	0.020
					2	0	0.000	
					3	4	1.000	
18	36	0.279	3	0.000	1	0	0.000	0.013
					2	0	0.000	
					3	36	1.000	

***** CROSS VALIDATION TREE 4 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	7	0.046	1	0.000	1	7	1.000	0.007
					2	0	0.000	
					3	0	0.000	
2	9	0.059	2	0.444	1	4	0.444	0.007
					2	5	0.556	
					3	0	0.000	
3	14	0.092	1	0.000	1	14	1.000	0.003
					2	0	0.000	
					3	0	0.000	
4	7	0.046	1	0.000	1	7	1.000	0.007
					2	0	0.000	
					3	0	0.000	
5	5	0.033	2	0.400	1	2	0.400	0.007
					2	3	0.600	
					3	0	0.000	
6	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
7	9	0.059	2	0.222	1	2	0.222	0.007
					2	7	0.778	
					3	0	0.000	
8	3	0.020	1	0.333	1	2	0.667	0.007
					2	1	0.333	
					3	0	0.000	
9	8	0.052	2	0.125	1	1	0.125	0.023
					2	7	0.875	
					3	0	0.000	
10	3	0.023	3	0.000	1	0	0.000	0.023
					2	0	0.000	
					3	3	1.000	
11	6	0.042	1	0.372	1	4	0.628	0.026
					2	0	0.000	
					3	2	0.372	

12	5	0.034	1	0.422	1	3	0.578	0.013
					2	1	0.193	
					3	1	0.229	
13	3	0.020	1	0.333	1	2	0.667	0.007
					2	1	0.333	
					3	0	0.000	
14	27	0.176	2	0.074	1	2	0.074	0.007
					2	25	0.926	
					3	0	0.000	
15	1	0.007	2	0.000	1	0	0.000	0.007
					2	1	1.000	
					3	0	0.000	
16	37	0.287	3	0.000	1	0	0.000	0.007
					2	0	0.000	
					3	37	1.000	

***** CROSS VALIDATION TREE 5 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	7	0.045	1	0.000	1	7	1.000	0.001
					2	0	0.000	
					3	0	0.000	
2	8	0.052	2	0.490	1	4	0.490	0.001
					2	4	0.510	
					3	0	0.000	
3	14	0.090	1	0.000	1	14	1.000	.512820E-03
					2	0	0.000	
					3	0	0.000	
4	9	0.058	1	0.229	1	7	0.771	0.024
					2	2	0.229	
					3	0	0.000	
5	11	0.073	2	0.176	1	2	0.176	0.032
					2	9	0.824	
					3	0	0.000	
6	5	0.032	1	0.000	1	5	1.000	0.032
					2	0	0.000	
					3	0	0.000	
7	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
8	30	0.199	2	0.097	1	3	0.097	0.006
					2	27	0.903	
					3	0	0.000	
9	8	0.052	1	0.129	1	7	0.871	0.038
					2	1	0.129	
					3	0	0.000	
10	8	0.054	2	0.261	1	1	0.118	0.016
					2	6	0.739	
					3	1	0.143	
11	3	0.022	3	0.293	1	1	0.293	0.016
					2	0	0.000	

```

                3  2  0.707
12  41  0.317  3  0.021  1  0  0.000  0.017
                2  1  0.021
                3  40  0.979
    
```

***** CROSS VALIDATION TREE 6 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	8	0.052	1	0.000	1	8	1.000	0.007
					2	0	0.000	
					3	0	0.000	
2	9	0.059	2	0.440	1	4	0.440	0.007
					2	5	0.560	
					3	0	0.000	
3	13	0.085	1	0.000	1	13	1.000	0.004
					2	0	0.000	
					3	0	0.000	
4	7	0.046	1	0.290	1	5	0.710	0.019
					2	2	0.290	
					3	0	0.000	
5	10	0.066	2	0.197	1	2	0.197	0.007
					2	8	0.803	
					3	0	0.000	
6	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
7	7	0.046	1	0.290	1	5	0.710	0.017
					2	2	0.290	
					3	0	0.000	
8	6	0.039	1	0.000	1	6	1.000	0.039
					2	0	0.000	
					3	0	0.000	
9	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
10	28	0.186	2	0.105	1	3	0.105	0.013
					2	25	0.895	
					3	0	0.000	
11	8	0.053	2	0.123	1	1	0.123	0.015
					2	7	0.877	
					3	0	0.000	
12	2	0.015	3	0.000	1	0	0.000	0.015
					2	0	0.000	
					3	2	1.000	
13	5	0.037	3	0.177	1	1	0.177	0.030
					2	0	0.000	
					3	4	0.823	
14	1	0.007	2	0.000	1	0	0.000	0.007
					2	1	1.000	
					3	0	0.000	
15	38	0.288	3	0.000	1	0	0.000	0.007

2 0 0.000
 3 38 1.000

***** CROSS VALIDATION TREE 7 *****

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	28	0.183	1	0.143	1	24	0.857	0.007
					2	4	0.143	
					3	0	0.000	
2	9	0.059	1	0.222	1	7	0.778	0.010
					2	2	0.222	
					3	0	0.000	
3	11	0.072	2	0.364	1	4	0.364	0.013
					2	7	0.636	
					3	0	0.000	
4	4	0.026	1	0.250	1	3	0.750	0.013
					2	1	0.250	
					3	0	0.000	
5	2	0.013	2	0.000	1	0	0.000	0.013
					2	2	1.000	
					3	0	0.000	
6	5	0.033	1	0.000	1	5	1.000	0.033
					2	0	0.000	
					3	0	0.000	
7	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
8	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
9	2	0.013	1	0.500	1	1	0.500	0.002
					2	1	0.500	
					3	0	0.000	
10	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
11	9	0.059	2	0.111	1	1	0.111	0.007
					2	8	0.889	
					3	0	0.000	
12	19	0.124	2	0.000	1	0	0.000	0.003
					2	19	1.000	
					3	0	0.000	
13	7	0.046	2	0.143	1	1	0.143	0.017
					2	6	0.857	
					3	0	0.000	
14	4	0.030	3	0.219	1	0	0.000	0.017
					2	1	0.219	
					3	3	0.781	
15	5	0.038	3	0.174	1	1	0.174	0.020
					2	0	0.000	
					3	4	0.826	

```

16  36      0.279  3      0.000   1   0      0.000      0.013
      2   0      0.000
      3  36      1.000

```

***** CROSS VALIDATION TREE 8 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	15	0.098	1	0.067	1	14	0.933	0.016
					2	1	0.067	
					3	0	0.000	
2	3	0.020	1	0.333	1	2	0.667	0.007
					2	1	0.333	
					3	0	0.000	
3	9	0.059	2	0.111	1	1	0.111	0.007
					2	8	0.889	
					3	0	0.000	
4	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
5	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
6	16	0.105	1	0.000	1	16	1.000	0.011
					2	0	0.000	
					3	0	0.000	
7	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
8	12	0.078	2	0.167	1	2	0.167	0.007
					2	10	0.833	
					3	0	0.000	
9	4	0.030	3	0.219	1	0	0.000	0.017
					2	1	0.219	
					3	3	0.781	
10	5	0.034	1	0.422	1	3	0.578	0.013
					2	1	0.193	
					3	1	0.229	
11	2	0.013	1	0.500	1	1	0.500	0.002
					2	1	0.500	
					3	0	0.000	
12	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
13	9	0.059	2	0.111	1	1	0.111	0.007
					2	8	0.889	
					3	0	0.000	
14	19	0.124	2	0.000	1	0	0.000	0.003
					2	19	1.000	
					3	0	0.000	
15	9	0.061	1	0.360	1	6	0.640	0.033
					2	1	0.107	
					3	2	0.253	

```

16  37      0.287  3      0.000  1  0      0.000  0.287
                        2  0      0.000
                        3  37     1.000

```

***** CROSS VALIDATION TREE 9 *****

=====
TERMINAL NODE INFORMATION
=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	27	0.176	1	0.111	1	24	0.889	0.098
					2	3	0.111	
					3	0	0.000	
2	6	0.039	1	0.000	1	6	1.000	0.016
					2	0	0.000	
					3	0	0.000	
3	10	0.065	2	0.300	1	3	0.300	0.007
					2	7	0.700	
					3	0	0.000	
4	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
5	2	0.013	2	0.000	1	0	0.000	0.003
					2	2	1.000	
					3	0	0.000	
6	9	0.059	1	0.000	1	9	1.000	0.013
					2	0	0.000	
					3	0	0.000	
7	6	0.039	2	0.333	1	2	0.333	0.013
					2	4	0.667	
					3	0	0.000	
8	28	0.183	2	0.071	1	2	0.071	0.007
					2	26	0.929	
					3	0	0.000	
9	5	0.033	1	0.400	1	3	0.600	0.007
					2	2	0.400	
					3	0	0.000	
10	9	0.061	2	0.253	1	0	0.000	0.030
					2	7	0.747	
					3	2	0.253	
11	5	0.038	3	0.174	1	1	0.174	0.030
					2	0	0.000	
					3	4	0.826	
12	37	0.287	3	0.000	1	0	0.000	0.015
					2	0	0.000	
					3	37	1.000	

***** CROSS VALIDATION TREE 10 *****

=====
TERMINAL NODE INFORMATION
=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	15	0.098	1	0.068	1	14	0.932	0.020
					2	1	0.068	
					3	0	0.000	
2	11	0.073	2	0.179	1	2	0.179	0.007
					2	9	0.821	

					3	0	0.000	
3	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
4	15	0.098	1	0.068	1	14	0.932	0.007
					2	1	0.068	
					3	0	0.000	
5	8	0.053	2	0.370	1	3	0.370	0.014
					2	5	0.630	
					3	0	0.000	
6	5	0.033	1	0.203	1	4	0.797	0.014
					2	1	0.203	
					3	0	0.000	
7	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
8	27	0.180	2	0.073	1	2	0.073	0.013
					2	25	0.927	
					3	0	0.000	
9	8	0.052	1	0.127	1	7	0.873	0.039
					2	1	0.127	
					3	0	0.000	
10	3	0.022	3	0.301	1	1	0.301	0.015
					2	0	0.000	
					3	2	0.699	
11	7	0.047	2	0.000	1	0	0.000	0.015
					2	7	1.000	
					3	0	0.000	
12	4	0.029	3	0.223	1	1	0.223	0.023
					2	0	0.000	
					3	3	0.777	
13	39	0.295	3	0.000	1	0	0.000	0.016
					2	0	0.000	
					3	39	1.000	

Lampiran D. Output *Missclassification* dengan Metode Pemilihan Pemilah Indeks Gini

```

=====
MISCLASSIFICATION BY CLASS
=====
(Cross Validation)

```

Class	Prior Prob	Wgt Count	Count	Wgt Misclass	Misclass	Cost
1	0.333	57.00	57	10.00	10	0.175
		(57.00)	57	22.00	22	0.386)
2	0.333	56.00	56	7.00	7	0.125
		(56.00)	56	18.00	18	0.321)
3	0.333	48.00	48	0.00	0	0.000
		(48.00)	48	4.00	4	0.083)

Total	1.000	161.00	161	17.00	17	
		(161.00)	161	44.00	44)	

Lampiran E. Akurasi Hasil Klasifikasi dengan Pemilihan Pemilah Indeks Gini

Akurasi Data Learning

Fold 1							
Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	47	5	0	52	0,9038462		
2	3	43	0	46	0,9347826	0,0763889	0,9236111
3	1	2	43	46	0,9347826		
Total	51	50	43	144			

Fold 2							
Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	41	2	0	43	0,9534884		
2	10	47	0	57	0,8245614	0,0965517	0,9034483
3	1	1	43	45	0,9555556		
Total	52	50	43	145			

Fold 3							
Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	43	4	0	47	0,9148936		
2	9	45	0	54	0,8333333	0,0965517	0,9034483
3	0	1	43	44	0,9772727		
Total	52	50	43	145			

Fold 4							
Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	40	3	3	46	0,8695652		
2	11	48	0	59	0,8135593	0,1172414	0,8827586
3	0	0	40	40	1		
Total	51	51	43	145			

Fold 5							
Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	41	3	0	44	0,9318182		
2	10	46	1	57	0,8070175	0,1103448	0,8896552
3	1	1	42	44	0,9545455		
Total	52	50	43	145			

Fold 6

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	40	4	0	44	0,9090909		
2	10	46	0	56	0,8214286	0,1034483	0,8965517
3	1	0	44	45	0,9777778		
Total	51	50	44	145			

Fold 7

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	44	8	0	52	0,8461538		
2	6	42	0	48	0,875	0,1103448	0,8896552
3	1	1	43	45	0,9555556		
Total	51	51	43	145			

Fold 8

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	47	5	3	55	0,8545455		
2	4	45	0	49	0,9183673	0,0896552	0,9103448
3	0	1	40	41	0,9756098		
Total	51	51	43	145			

Fold 9

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	43	5	0	48	0,8958333		
2	7	46	2	55	0,8363636	0,1034483	0,8965517
3	1	1	41	42	0,9761905		
Total	51	51	43	145			

Fold 10

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	42	4	0	46	0,9130435		
2	7	46	0	53	0,8679245	0,0896552	0,9103448
3	2	0	44	46	0,9565217		
Total	51	50	44	145			

Akurasi Data Testing

=====
CROSS VALIDATION CLASSIFICATION TABLE
=====

Actual Class	Predicted Class			Actual Total
	1	2	3	
1	35.00	21.00	1.00	57.00
2	14.00	38.00	4.00	56.00

3	2.00	2.00	44.00	48.00

PRED. TOT.	51.00	61.00	49.00	161.00
CORRECT	0.614	0.679	0.917	
SUCCESS IND.	0.260	0.331	0.619	
TOT. CORRECT	0.727			

=====
 CROSS VALIDATION CLASSIFICATION PROBABILITY TABLE
 =====

Actual Class	Predicted Class			Total
	1	2	3	
1	0.614	0.368	0.018	1.000
2	0.250	0.679	0.071	1.000
3	0.042	0.042	0.917	1.000

Lampiran F. Informasi Simpul Terminal Pohon Klasifikasi dengan Pemilihan Pemilah Indeks Gini

=====
 NODE INFORMATION
 =====

```

*****
*                               *
*           Node 1: X7           *
*           N: 161               *
*****
*****
*                               *
*           Node 2               *
*           N: 103               *
*                               *
*           Node 12              *
*           N: 58                *
*****
Node 1 was split on X7
A case goes left if X7 = (1,2,4,5)
Improvement = 0.261      Complexity Threshold = 0.322
Node   Cases  Wgt  Counts      Cost  Class
  1     161    161.00  0.667  1
  2     103    103.00  0.470  1
 12     58     58.00  0.151  3
Weighted Counts
Class   Top      Left      Right
  1     57.00    55.00    2.00
  2     56.00    48.00    8.00
  3     48.00     0.00   48.00
Within Node Probabilities
Class   Top      Left      Right
  1     0.333    0.530    0.030
  2     0.333    0.470    0.121
  3     0.333    0.000    0.849
Surrogate      Split      Assoc.  Improve.
  1 X6          s 1,3,4,5  0.710   0.257
  2 X3          s      15.058 0.297   0.072
  3 X5          s 2          0.123   0.042
Competitor      Split      Improve.
  1 X6          1,3,4,5  0.257
  2 X3          15.058  0.072
  3 X2          2          0.050
  4 X5          2          0.042
  5 X1          2,3,4,5  0.023

```

```

*****
*                               *
*           Node 2: X2           *
*           N: 103               *
*****
*****
*                               *
*           Node 3               *
*                               *
*           Node 10              *
*                               *

```

```

*                N: 61                *                *                N: 42                *
*****
Node 2 was split on X2
A case goes left if X2 = (2)
Improvement = 0.052                Complexity Threshold = 0.108
Node   Cases  Wgt Counts                Cost Class
  2     103    103.00                0.470 1
  3     61     61.00                0.299 1
  10    42     42.00                0.282 2

                Weighted Counts
Class   Top    Left    Right
  1     55.00  43.00  12.00
  2     48.00  18.00  30.00
  3      0.00   0.00   0.00

                Within Node Probabilities
Class   Top    Left    Right
  1     0.530  0.701  0.282
  2     0.470  0.299  0.718
  3     0.000  0.000  0.000

Surrogate      Split                Assoc.    Improve.
  1 X3          s 14.500                0.170    0.017
  2 X6          s 1,2,4,5                0.099    0.009
  3 X1          s 1,2,4,5                0.024    0.004

Competitor      Split                Improve.
  1 X1          2,3,5                0.041
  2 X7          5                0.026
  3 X3          7.231                0.024
  4 X5          2                0.021
  5 X6          4,5                0.017

                *****
                *                Node 3: X7                *
                *                N: 61                *
                *****
*****
*                Node 4                *                *                Node 6                *
*                N: 32                *                *                N: 29                *
*****
Node 3 was split on X7
A case goes left if X7 = (1,5)
Improvement = 0.015                Complexity Threshold = 0.012
Node   Cases  Wgt Counts                Cost Class
  3     61     61.00                0.299 1
  4     32     32.00                0.159 1
  6     29     29.00                0.453 1

                Weighted Counts
Class   Top    Left    Right
  1     43.00  27.00  16.00
  2     18.00   5.00  13.00
  3      0.00   0.00   0.00

                Within Node Probabilities
Class   Top    Left    Right
  1     0.701  0.841  0.547
  2     0.299  0.159  0.453
  3     0.000  0.000  0.000

Surrogate      Split                Assoc.    Improve.
  1 X1          s 1,3,4                0.208    0.003
  2 X4          s 2                0.140    .335557E-03
  3 X6          s 1,4                0.140    .218972E-03
  4 X3          r 9.847                0.105    0.002
  5 X5          s 2                0.035    0.004

Competitor      Split                Improve.
  1 X3          8.500                0.011
  2 X6          1,2,5                0.007
  3 X1          2,3,5                0.005
  4 X5          2                0.004
  5 X4          2                .335557E-03

                *****

```

```

*                Node 4: X3                *
*                N: 32                      *
*****
*****
*                Node 5                *
*                N: 17                  *
*****
=====
*                Node 5                *
*                N: 17                  *
*                Terminal Node 3        *
*                N: 15                  *
=====

Node 4 was split on X3
A case goes left if X3 <= 11.500
Improvement = 0.008                Complexity Threshold = 0.003

Node   Cases   Wgt   Counts           Cost   Class
  4         32     32.00      0.159 1
  5         17     17.00      0.298 1
 -3         15     15.00      0.000 1

Weighted Counts
Class   Top      Left      Right
  1     27.00    12.00    15.00
  2      5.00     5.00     0.00
  3      0.00     0.00     0.00

Within Node Probabilities
Class   Top      Left      Right
  1     0.841    0.702    1.000
  2     0.159    0.298    0.000
  3     0.000    0.000    0.000

Surrogate      Split      Assoc.   Improve.
  1 X4          s 1          0.332   0.002
  2 X6          s 1,2,5    0.265   0.007
  3 X5          s 2          0.133   0.002
  4 X1          s 3,4,5    0.129   .482115E-03

Competitor      Split      Improve.
  1 X6          1,2,5      0.007
  2 X1          1,4,5      0.004
  3 X7          5          0.002
  4 X5          2          0.002
  5 X4          2          0.002

*****
*                Node 5: X3                *
*                N: 17                      *
*****
=====
=                Terminal Node 1          =
=                N: 8                    =
=====
=                Terminal Node 2          =
=                N: 9                    =
=====

Node 5 was split on X3
A case goes left if X3 <= 9.000
Improvement = 0.016                Complexity Threshold = 0.006

Node   Cases   Wgt   Counts           Cost   Class
  5         17     17.00      0.298 1
 -1         8      8.00      0.000 1
 -2         9      9.00      0.440 2

Weighted Counts
Class   Top      Left      Right
  1     12.00     8.00     4.00
  2      5.00     0.00     5.00
  3      0.00     0.00     0.00

Within Node Probabilities
Class   Top      Left      Right
  1     0.702    1.000    0.440
  2     0.298    0.000    0.560
  3     0.000    0.000    0.000

Surrogate      Split      Assoc.   Improve.
  1 X7          s 5          0.500   0.005
  2 X1          s 2,5      0.375   0.004
  3 X6          s 2,3      0.250   0.002
  4 X5          s 1        0.250   0.002
  5 X4          s 1        0.116   .121509E-03

Competitor      Split      Improve.
  1 X1          1,4,5      0.007
  2 X7          5          0.005

```

```

3 X6          1,2,5                      0.005
4 X5          2                          0.002
5 X4          2                          .121509E-03

```

```

*****
*                Node 6: X3                *
*                N: 29                     *
*****
=====
=      Terminal Node 4      =      *      Node 7      *
=      N: 9                 =      *      N: 20      *
=====
*****

```

Node 6 was split on X3
A case goes left if X3 <= 9.241
Improvement = 0.008 Complexity Threshold = 0.018

Node	Cases	Wgt	Counts	Cost	Class
6	29		29.00	0.453	1
-4	9		9.00	0.225	1
7	20		20.00	0.446	2

Weighted Counts

Class	Top	Left	Right
1	16.00	7.00	9.00
2	13.00	2.00	11.00
3	0.00	0.00	0.00

Within Node Probabilities

Class	Top	Left	Right
1	0.547	0.775	0.446
2	0.453	0.225	0.554
3	0.000	0.000	0.000

Competitor	Split	Improve.
1 X1	2,3,5	0.006
2 X4	2	0.004
3 X6	1,3,4,5	0.004
4 X5	2	0.003
5 X7	4	.349485E-03

```

*****
*                Node 7: X5                *
*                N: 20                     *
*****
=====
*      Node 8      *      =      Terminal Node 8      =
*      N: 14       *      =      N: 6                 =
*****
=====

```

Node 7 was split on X5
A case goes left if X5 = (2)
Improvement = 0.015 Complexity Threshold = 0.023

Node	Cases	Wgt	Counts	Cost	Class
7	20		20.00	0.446	2
8	14		14.00	0.282	2
-8	6		6.00	0.169	1

Weighted Counts

Class	Top	Left	Right
1	9.00	4.00	5.00
2	11.00	10.00	1.00
3	0.00	0.00	0.00

Within Node Probabilities

Class	Top	Left	Right
1	0.446	0.282	0.831
2	0.554	0.718	0.169
3	0.000	0.000	0.000

Surrogate	Split	Assoc.	Improve.
1 X3	s 13.500	0.329	0.005
2 X1	s 1,2,3,4	0.166	0.004
3 X6	s 1,2,4,5	0.160	0.002

Competitor	Split	Improve.
1 X3	9.847	0.005
2 X1	2,3,5	0.005
3 X6	2,4,5	0.003
4 X7	4	.252282E-03
5 X4	2	.196005E-03

```

*****
*                Node 8: X6                *
*                N: 14                      *
*****
*****
*                Node 9                *      = Terminal Node 7      =
*                N: 11                *      =                N: 3      =
*****
Node 8 was split on X6
A case goes left if X6 = (1,4,5)
Improvement = 0.004          Complexity Threshold = 0.003
Node   Cases  Wgt  Counts      Cost  Class
  8     14     14.00  14.00    0.282  2
  9     11     11.00   11.00    0.360  2
 -7     3       3.00   3.00     0.000  2

          Weighted Counts
Class    Top      Left      Right
  1         4.00    4.00     0.00
  2        10.00    7.00     3.00
  3         0.00    0.00     0.00

          Within Node Probabilities
Class    Top      Left      Right
  1         0.282   0.360    0.000
  2         0.718   0.640    1.000
  3         0.000   0.000    0.000

Surrogate      Split      Assoc.  Improve.
  1 X3          s          0.667    0.002
  2 X7          s 4          0.339   .101060E-03

Competitor      Split      Improve.
  1 X1          1,3,4,5      0.003
  2 X3          9.847        0.002
  3 X4          2          .101060E-03
  4 X7          4          .101060E-03

*****
*                Node 9: X3                *
*                N: 11                      *
*****
=====
=      Terminal Node 5      =      =      Terminal Node 6      =
=      N: 6                 =      =      N: 5                 =
=====

Node 9 was split on X3
A case goes left if X3 <= 10.757
Improvement = 0.006          Complexity Threshold = 0.006
Node   Cases  Wgt  Counts      Cost  Class
  9     11     11.00  11.00    0.360  2
 -5     6       6.00   6.00    0.164  2
 -6     5       5.00   5.00    0.404  1

          Weighted Counts
Class    Top      Left      Right
  1         4.00    1.00     3.00
  2         7.00    5.00     2.00
  3         0.00    0.00     0.00

          Within Node Probabilities
Class    Top      Left      Right
  1         0.360   0.164    0.596
  2         0.640   0.836    0.404
  3         0.000   0.000    0.000

Surrogate      Split      Assoc.  Improve.
  1 X1          s 1,3,5      0.599   .141997E-03
  2 X6          s 1          0.397   .141997E-03
  3 X7          s 4          0.199    0.005

Competitor      Split      Improve.
  1 X7          4          0.005
  2 X1          3,4,5      0.002
  3 X4          2          .533430E-03
  4 X6          4,5      .141997E-03

```

```

*****
*                               *
*           Node 10: X1         *
*           N: 42              *
*****
=====
= Terminal Node 9 = *           Node 11 *
= N: 8 = *           N: 34 *
=====
*****
Node 10 was split on X1
A case goes left if X1 = (2,3,4,5)
Improvement = 0.040      Complexity Threshold = 0.035
Node   Cases  Wgt Counts  Cost Class
  10     42    42.00    0.282 2
  -9     8     8.00    0.127 1
  11    34    34.00    0.145 2

Weighted Counts
Class      Top      Left      Right
  1      12.00    7.00     5.00
  2      30.00    1.00    29.00
  3       0.00    0.00     0.00

Within Node Probabilities
Class      Top      Left      Right
  1      0.282    0.873    0.145
  2      0.718    0.127    0.855
  3      0.000    0.000    0.000

Surrogate      Split      Assoc.      Improve.
  1 X3          s        6.231      0.249      0.013
  2 X6          s 2,5    0.122      0.018
Competitor      Split      Improve.
  1 X7          4,5    0.020
  2 X5          2      0.018
  3 X6          2,5    0.018
  4 X3          6.231  0.013
  5 X4          2      0.005

*****
*                               *
*           Node 11: X7         *
*           N: 34              *
*****
=====
= Terminal Node 10 = = Terminal Node 11 =
= N: 2 = = N: 32 =
=====
Node 11 was split on X7
A case goes left if X7 = (5)
Improvement = 0.018      Complexity Threshold = 0.012
Node   Cases  Wgt Counts  Cost Class
  11    34    34.00    0.145 2
 -10     2     2.00    0.000 1
 -11    32    32.00    0.092 2

Weighted Counts
Class      Top      Left      Right
  1       5.00    2.00     3.00
  2      29.00    0.00    29.00
  3       0.00    0.00     0.00

Within Node Probabilities
Class      Top      Left      Right
  1      0.145    1.000    0.092
  2      0.855    0.000    0.908
  3      0.000    0.000    0.000

Competitor      Split      Improve.
  1 X5          2      0.007
  2 X6          2,5    0.003
  3 X3          16.994  0.003
  4 X4          2      .190408E-03

*****
*                               *
*           Node 12: X6         *
*           N: 58              *
*****

```

```

*****
*           Node 13           *
*           N: 17            *
*****
=====
=           Terminal Node 15  =
=           N: 41            =
=====

Node 12 was split on X6
A case goes left if X6 = (1,3,4,5)
Improvement = 0.040           Complexity Threshold = 0.014

Node      Cases  Wgt Counts      Cost Class
  12         58      58.00      0.151 3
  13         17      17.00      0.550 3
 -15         41      41.00      0.000 3

                Weighted Counts
Class      Top      Left      Right
  1         2.00     2.00     0.00
  2         8.00     8.00     0.00
  3        48.00     7.00    41.00

                Within Node Probabilities
Class      Top      Left      Right
  1         0.030     0.108     0.000
  2         0.121     0.441     0.000
  3         0.849     0.450     1.000

Surrogate      Split      Assoc.      Improve.
  1 X3          s 3.489      0.064      .271581E-03
Competitor      Split      Improve.
  1 X1          1,3,4,5      0.014
  2 X2          2          0.012
  3 X3          15.058      0.011
  4 X5          2          0.006
  5 X4          2          0.001

*****
*           Node 13: X3       *
*           N: 17            *
*****

*****
*           Node 14           *
*           N: 12            *
*****
=====
=           Terminal Node 14  =
=           N: 5             =
=====

Node 13 was split on X3
A case goes left if X3 <= 15.500
Improvement = 0.017           Complexity Threshold = 0.021

Node      Cases  Wgt Counts      Cost Class
  13         17      17.00      0.550 3
  14         12      12.00      0.359 2
 -14         5       5.00      0.174 3

                Weighted Counts
Class      Top      Left      Right
  1         2.00     1.00     1.00
  2         8.00     8.00     0.00
  3         7.00     3.00     4.00

                Within Node Probabilities
Class      Top      Left      Right
  1         0.108     0.079     0.174
  2         0.441     0.641     0.000
  3         0.450     0.280     0.826

Surrogate      Split      Assoc.      Improve.
  1 X1          s 2,3,4,5      0.439      0.009
  2 X6          s 3,4,5      0.030      .291620E-03
Competitor      Split      Improve.
  1 X2          2          0.009
  2 X1          2,3,4,5      0.009
  3 X6          4,5      0.002
  4 X5          2          0.002
  5 X4          2          .674822E-03

*****
*           Node 14: X2       *
*           N: 12            *
*****

=====
=           Terminal Node 12  =
=           Terminal Node 13  =
=====

```

```

===== N: 8 ===== = = N: 4 =====
Node 14 was split on X2
A case goes left if X2 = (2)
Improvement = 0.018 Complexity Threshold = 0.015
Node Cases Wgt Counts Cost Class
 14 12 12.00 0.359 2
-12 8 8.00 0.123 2
-13 4 4.00 0.222 3
Weighted Counts
Class Top Left Right
1 1.00 1.00 0.00
2 8.00 7.00 1.00
3 3.00 0.00 3.00
Within Node Probabilities
Class Top Left Right
1 0.079 0.123 0.000
2 0.641 0.877 0.222
3 0.280 0.000 0.778
Surrogate Split Assoc. Improve.
1 X3 r 6.500 0.519 0.016
2 X4 s 1 0.259 .354362E-03
3 X1 s 2,3,4,5 0.259 .354362E-03
4 X6 s 1,4,5 0.037 0.002
5 X5 s 2 0.037 0.002
Competitor Split Improve.
1 X3 8.500 0.017
2 X6 1,4,5 0.002
3 X5 2 0.002
4 X4 2 .354362E-03
5 X1 2,3,4,5 .354362E-03

```

=====
TERMINAL NODE INFORMATION
=====

[Breiman adjusted cost, lambda = 0.039]

Node	Class	Wgt Count	N	Prob	Cost	Parent Complexity
1	1	8.00	8	0.047	0.000	0.006 [0.358]
	1	8.00	8	1.000		
	2	0.00	0	0.000		
	3	0.00	0	0.000		
2	2	9.00	9	0.053	0.440	0.006 [0.773]
	1	4.00	4	0.440		
	2	5.00	5	0.560		
	3	0.00	0	0.000		
3	1	15.00	15	0.088	0.000	0.003 [0.242]
	1	15.00	15	1.000		
	2	0.00	0	0.000		
	3	0.00	0	0.000		
4	1	9.00	9	0.053	0.225	0.018 [0.559]
	1	7.00	7	0.775		
	2	2.00	2	0.225		
	3	0.00	0	0.000		
5	2	6.00	6	0.036	0.164	0.006 [0.576]
	1	1.00	1	0.164		
	2	5.00	5	0.836		
	3	0.00	0	0.000		

6	1	5.00	5	0.029	0.404	0.006 [0.853]
	1	3.00	3	0.596		
	2	2.00	2	0.404		
	3	0.00	0	0.000		
7	2	3.00	3	0.018	0.000	0.003 [0.541]
	1	0.00	0	0.000		
	2	3.00	3	1.000		
	3	0.00	0	0.000		
8	1	6.00	6	0.035	0.169	0.023 [0.583]
	1	5.00	5	0.831		
	2	1.00	1	0.169		
	3	0.00	0	0.000		
9	1	8.00	8	0.047	0.127	0.035 [0.484]
	1	7.00	7	0.873		
	2	1.00	1	0.127		
	3	0.00	0	0.000		
10	1	2.00	2	0.012	0.000	0.012 [0.607]
	1	2.00	2	1.000		
	2	0.00	0	0.000		
	3	0.00	0	0.000		
11	2	32.00	32	0.190	0.092	0.012 [0.226]
	1	3.00	3	0.092		
	2	29.00	29	0.908		
	3	0.00	0	0.000		
12	2	8.00	8	0.048	0.123	0.015 [0.478]
	1	1.00	1	0.123		
	2	7.00	7	0.877		
	3	0.00	0	0.000		
13	3	4.00	4	0.027	0.222	0.015 [0.689]
	1	0.00	0	0.000		
	2	1.00	1	0.222		
	3	3.00	3	0.778		
14	3	5.00	5	0.034	0.174	0.021 [0.597]
	1	1.00	1	0.174		
	2	0.00	0	0.000		
	3	4.00	4	0.826		
15	3	41.00	41	0.285	0.000	0.014 [0.095]
	1	0.00	0	0.000		
	2	0.00	0	0.000		
	3	41.00	41	1.000		

Lampiran H. Output Pohon Klasifikasi *Cross-Validation 10-Fold* dengan Metode Pemilah Indeks Twoing

***** CROSS VALIDATION TREE 1 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	14	0.092	1	0.073	1	13	0.927	0.017
					2	1	0.073	
					3	0	0.000	
2	5	0.033	1	0.405	1	3	0.595	0.006
					2	2	0.405	
					3	0	0.000	
3	9	0.060	2	0.109	1	1	0.109	0.007
					2	8	0.891	
					3	0	0.000	
4	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
5	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
6	16	0.105	1	0.000	1	16	1.000	0.007
					2	0	0.000	
					3	0	0.000	
7	1	0.007	2	0.000	1	0	0.000	0.007
					2	1	1.000	
					3	0	0.000	
8	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
9	15	0.102	2	0.280	1	2	0.128	0.013
					2	11	0.720	
					3	2	0.152	
10	2	0.013	2	0.000	1	0	0.000	0.013
					2	2	1.000	
					3	0	0.000	
11	10	0.066	1	0.203	1	8	0.797	0.013
					2	2	0.203	
					3	0	0.000	
12	6	0.044	3	0.299	1	1	0.148	0.024
					2	1	0.151	
					3	4	0.701	
13	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
14	21	0.140	2	0.000	1	0	0.000	0.013
					2	21	1.000	
					3	0	0.000	
15	1	0.007	2	0.000	1	0	0.000	0.007

					2	1	1.000	
					3	0	0.000	
16	37	0.287	3	0.000	1	0	0.000	0.007
					2	0	0.000	
					3	37	1.000	

***** CROSS VALIDATION TREE 2 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	6	0.038	1	0.000	1	6	1.000	0.008
					2	0	0.000	
					3	0	0.000	
2	9	0.059	2	0.435	1	4	0.435	0.008
					2	5	0.565	
					3	0	0.000	
3	15	0.096	1	0.000	1	15	1.000	0.004
					2	0	0.000	
					3	0	0.000	
4	5	0.033	2	0.194	1	1	0.194	0.016
					2	4	0.806	
					3	0	0.000	
5	8	0.052	1	0.129	1	7	0.871	0.014
					2	1	0.129	
					3	0	0.000	
6	9	0.059	2	0.216	1	2	0.216	0.006
					2	7	0.784	
					3	0	0.000	
7	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
8	4	0.026	1	0.000	1	4	1.000	0.026
					2	0	0.000	
					3	0	0.000	
9	7	0.045	1	0.148	1	6	0.852	0.032
					2	1	0.148	
					3	0	0.000	
10	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
11	28	0.186	2	0.069	1	2	0.069	0.013
					2	26	0.931	
					3	0	0.000	
12	37	0.287	3	0.000	1	0	0.000	0.011
					2	0	0.000	
					3	37	1.000	
13	3	0.022	3	0.293	1	1	0.293	0.016
					2	0	0.000	
					3	2	0.707	
14	7	0.048	2	0.162	1	0	0.000	0.016
					2	6	0.838	
					3	1	0.162	
15	4	0.030	3	0.216	1	1	0.216	0.016

2 0 0.000
 3 3 0.784

***** CROSS VALIDATION TREE 3 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	7	0.045	1	0.000	1	7	1.000	0.008
					2	0	0.000	
					3	0	0.000	
2	9	0.059	2	0.435	1	4	0.435	0.008
					2	5	0.565	
					3	0	0.000	
3	14	0.090	1	0.000	1	14	1.000	0.004
					2	0	0.000	
					3	0	0.000	
4	6	0.039	1	0.172	1	5	0.828	0.014
					2	1	0.172	
					3	0	0.000	
5	9	0.059	2	0.325	1	3	0.325	0.006
					2	6	0.675	
					3	0	0.000	
6	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
7	2	0.013	2	0.000	1	0	0.000	0.003
					2	2	1.000	
					3	0	0.000	
8	7	0.045	1	0.294	1	5	0.706	0.019
					2	2	0.294	
					3	0	0.000	
9	8	0.052	1	0.129	1	7	0.871	0.038
					2	1	0.129	
					3	0	0.000	
10	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
11	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
12	1	0.006	1	0.000	1	1	1.000	0.006
					2	0	0.000	
					3	0	0.000	
13	9	0.060	2	0.107	1	1	0.107	0.006
					2	8	0.893	
					3	0	0.000	
14	18	0.120	2	0.000	1	0	0.000	0.003
					2	18	1.000	
					3	0	0.000	
15	36	0.279	3	0.000	1	0	0.000	0.013
					2	0	0.000	
					3	36	1.000	

16	3	0.022	3	0.293	1	1	0.293	0.016
					2	0	0.000	
					3	2	0.707	
17	8	0.054	2	0.142	1	0	0.000	0.016
					2	7	0.858	
					3	1	0.142	
18	4	0.031	3	0.000	1	0	0.000	0.019
					2	0	0.000	
					3	4	1.000	
***** CROSS VALIDATION TREE 4 *****								
=====								
TERMINAL NODE INFORMATION								
=====								
Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold

1	7	0.046	1	0.000	1	7	1.000	0.007
					2	0	0.000	
					3	0	0.000	
2	9	0.059	2	0.444	1	4	0.444	0.007
					2	5	0.556	
					3	0	0.000	
3	14	0.092	1	0.000	1	14	1.000	0.003
					2	0	0.000	
					3	0	0.000	
4	7	0.046	1	0.000	1	7	1.000	0.007
					2	0	0.000	
					3	0	0.000	
5	5	0.033	2	0.400	1	2	0.400	0.007
					2	3	0.600	
					3	0	0.000	
6	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
7	9	0.059	2	0.222	1	2	0.222	0.007
					2	7	0.778	
					3	0	0.000	
8	3	0.020	1	0.333	1	2	0.667	0.007
					2	1	0.333	
					3	0	0.000	
9	8	0.052	2	0.125	1	1	0.125	0.023
					2	7	0.875	
					3	0	0.000	
10	3	0.023	3	0.000	1	0	0.000	0.023
					2	0	0.000	
					3	3	1.000	
11	6	0.042	1	0.372	1	4	0.628	0.026
					2	0	0.000	
					3	2	0.372	
12	5	0.034	1	0.422	1	3	0.578	0.013
					2	1	0.193	
					3	1	0.229	
13	3	0.020	1	0.333	1	2	0.667	0.007
					2	1	0.333	
					3	0	0.000	

14	27	0.176	2	0.074	1	2	0.074	0.007
					2	25	0.926	
					3	0	0.000	
15	1	0.007	2	0.000	1	0	0.000	0.007
					2	1	1.000	
					3	0	0.000	
16	37	0.287	3	0.000	1	0	0.000	0.007
					2	0	0.000	
					3	37	1.000	

***** CROSS VALIDATION TREE 5 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	7	0.045	1	0.000	1	7	1.000	0.001
					2	0	0.000	
					3	0	0.000	
2	8	0.052	2	0.490	1	4	0.490	0.001
					2	4	0.510	
					3	0	0.000	
3	14	0.090	1	0.000	1	14	1.000	.512820E-03
					2	0	0.000	
					3	0	0.000	
4	9	0.058	1	0.229	1	7	0.771	0.020
					2	2	0.229	
					3	0	0.000	
5	10	0.066	2	0.194	1	2	0.194	0.032
					2	8	0.806	
					3	0	0.000	
6	5	0.032	1	0.000	1	5	1.000	0.032
					2	0	0.000	
					3	0	0.000	
7	7	0.046	2	0.138	1	1	0.138	0.023
					2	6	0.862	
					3	0	0.000	
8	3	0.023	3	0.000	1	0	0.000	0.023
					2	0	0.000	
					3	3	1.000	
9	9	0.059	1	0.243	1	7	0.757	0.038
					2	1	0.112	
					3	1	0.131	
10	5	0.035	3	0.557	1	2	0.366	0.009
					2	1	0.191	
					3	2	0.443	
11	30	0.199	2	0.097	1	3	0.097	0.009
					2	27	0.903	
					3	0	0.000	
12	1	0.007	2	0.000	1	0	0.000	0.007
					2	1	1.000	
					3	0	0.000	
13	37	0.287	3	0.000	1	0	0.000	0.007
					2	0	0.000	
					3	37	1.000	

***** CROSS VALIDATION TREE 6 *****

```
=====
TERMINAL NODE INFORMATION
=====
```

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	8	0.052	1	0.000	1	8	1.000	0.007
					2	0	0.000	
					3	0	0.000	
2	9	0.059	2	0.440	1	4	0.440	0.007
					2	5	0.560	
					3	0	0.000	
3	13	0.085	1	0.000	1	13	1.000	0.004
					2	0	0.000	
					3	0	0.000	
4	7	0.046	1	0.290	1	5	0.710	0.020
					2	2	0.290	
					3	0	0.000	
5	12	0.080	2	0.246	1	3	0.246	0.026
					2	9	0.754	
					3	0	0.000	
6	6	0.039	1	0.169	1	5	0.831	0.026
					2	1	0.169	
					3	0	0.000	
7	6	0.039	1	0.000	1	6	1.000	0.039
					2	0	0.000	
					3	0	0.000	
8	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
9	28	0.186	2	0.105	1	3	0.105	0.013
					2	25	0.895	
					3	0	0.000	
10	8	0.053	2	0.123	1	1	0.123	0.015
					2	7	0.877	
					3	0	0.000	
11	2	0.015	3	0.000	1	0	0.000	0.015
					2	0	0.000	
					3	2	1.000	
12	5	0.037	3	0.177	1	1	0.177	0.030
					2	0	0.000	
					3	4	0.823	
13	1	0.007	2	0.000	1	0	0.000	0.007
					2	1	1.000	
					3	0	0.000	
14	38	0.288	3	0.000	1	0	0.000	0.007
					2	0	0.000	
					3	38	1.000	

***** CROSS VALIDATION TREE 7 *****

```
=====
TERMINAL NODE INFORMATION
=====
```

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	28	0.183	1	0.143	1	24	0.857	0.007
					2	4	0.143	

					3	0	0.000	
2	9	0.059	1	0.222	1	7	0.778	0.010
					2	2	0.222	
					3	0	0.000	
3	11	0.072	2	0.364	1	4	0.364	0.013
					2	7	0.636	
					3	0	0.000	
4	4	0.026	1	0.250	1	3	0.750	0.013
					2	1	0.250	
					3	0	0.000	
5	2	0.013	2	0.000	1	0	0.000	0.013
					2	2	1.000	
					3	0	0.000	
6	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
7	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
8	31	0.203	2	0.097	1	3	0.097	0.007
					2	28	0.903	
					3	0	0.000	
9	5	0.033	1	0.000	1	5	1.000	0.033
					2	0	0.000	
					3	0	0.000	
10	36	0.279	3	0.000	1	0	0.000	0.013
					2	0	0.000	
					3	36	1.000	
11	3	0.022	3	0.297	1	1	0.297	0.016
					2	0	0.000	
					3	2	0.703	
12	8	0.054	2	0.145	1	0	0.000	0.016
					2	7	0.855	
					3	1	0.145	
13	5	0.038	3	0.174	1	1	0.174	0.019
					2	0	0.000	
					3	4	0.826	

***** CROSS VALIDATION TREE 8 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	15	0.098	1	0.067	1	14	0.933	0.016
					2	1	0.067	
					3	0	0.000	
2	3	0.020	1	0.333	1	2	0.667	0.007
					2	1	0.333	
					3	0	0.000	
3	9	0.059	2	0.111	1	1	0.111	0.007
					2	8	0.889	
					3	0	0.000	
4	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	

					3	0	0.000	
5	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
6	16	0.105	1	0.000	1	16	1.000	0.011
					2	0	0.000	
					3	0	0.000	
7	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
8	12	0.078	2	0.167	1	2	0.167	0.007
					2	10	0.833	
					3	0	0.000	
9	4	0.030	3	0.219	1	0	0.000	0.017
					2	1	0.219	
					3	3	0.781	
10	9	0.061	1	0.360	1	6	0.640	0.033
					2	1	0.107	
					3	2	0.253	
11	5	0.034	1	0.422	1	3	0.578	0.013
					2	1	0.193	
					3	1	0.229	
12	31	0.203	2	0.097	1	3	0.097	0.013
					2	28	0.903	
					3	0	0.000	
13	37	0.287	3	0.000	1	0	0.000	0.287
					2	0	0.000	
					3	37	1.000	

***** CROSS VALIDATION TREE 9 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	27	0.176	1	0.111	1	24	0.889	0.098
					2	3	0.111	
					3	0	0.000	
2	6	0.039	1	0.000	1	6	1.000	0.016
					2	0	0.000	
					3	0	0.000	
3	10	0.065	2	0.300	1	3	0.300	0.007
					2	7	0.700	
					3	0	0.000	
4	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
5	2	0.013	2	0.000	1	0	0.000	0.003
					2	2	1.000	
					3	0	0.000	
6	9	0.059	1	0.000	1	9	1.000	0.013
					2	0	0.000	
					3	0	0.000	
7	6	0.039	2	0.333	1	2	0.333	0.013
					2	4	0.667	

					3	0	0.000	
8	28	0.183	2	0.071	1	2	0.071	0.007
					2	26	0.929	
					3	0	0.000	
9	5	0.033	1	0.400	1	3	0.600	0.007
					2	2	0.400	
					3	0	0.000	
10	37	0.287	3	0.000	1	0	0.000	0.015
					2	0	0.000	
					3	37	1.000	
11	9	0.061	2	0.253	1	0	0.000	0.030
					2	7	0.747	
					3	2	0.253	
12	5	0.038	3	0.174	1	1	0.174	0.030
					2	0	0.000	
					3	4	0.826	

***** CROSS VALIDATION TREE 10 *****

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	15	0.098	1	0.068	1	14	0.932	0.020
					2	1	0.068	
					3	0	0.000	
2	11	0.073	2	0.179	1	2	0.179	0.007
					2	9	0.821	
					3	0	0.000	
3	1	0.007	1	0.000	1	1	1.000	0.007
					2	0	0.000	
					3	0	0.000	
4	15	0.098	1	0.068	1	14	0.932	0.007
					2	1	0.068	
					3	0	0.000	
5	8	0.053	2	0.370	1	3	0.370	0.014
					2	5	0.630	
					3	0	0.000	
6	5	0.033	1	0.203	1	4	0.797	0.014
					2	1	0.203	
					3	0	0.000	
7	2	0.013	1	0.000	1	2	1.000	0.013
					2	0	0.000	
					3	0	0.000	
8	27	0.180	2	0.073	1	2	0.073	0.013
					2	25	0.927	
					3	0	0.000	
9	8	0.052	1	0.127	1	7	0.873	0.039
					2	1	0.127	
					3	0	0.000	
10	3	0.022	3	0.301	1	1	0.301	0.015
					2	0	0.000	
					3	2	0.699	
11	7	0.047	2	0.000	1	0	0.000	0.015
					2	7	1.000	

					3	0	0.000	
12	4	0.029	3	0.223	1	1	0.223	0.023
					2	0	0.000	
					3	3	0.777	
13	39	0.295	3	0.000	1	0	0.000	0.016
					2	0	0.000	
					3	39	1.000	

Lampiran I. Output *Missclassification* dengan Metode Pemilihan Pemilah Indeks Twoing

=====

MISCLASSIFICATION BY CLASS

=====

(Cross Validation)

Class	Prior Prob	Wgt	Count	Count	Wgt		Cost
					Misclass	Misclass	
1	0.333		57.00	57	9.00	9	0.158
			(57.00)	57	23.00	23	(0.404)
2	0.333		56.00	56	9.00	9	0.161
			(56.00)	56	18.00	18	(0.321)
3	0.333		48.00	48	1.00	1	0.021
			(48.00)	48	5.00	5	(0.104)
Total	1.000		161.00	161	19.00	19	
			(161.00)	161	46.00	46)	

Lampiran J. Akurasi Hasil Klasifikasi dengan Pemilihan pemilah Indeks Twoing

Akurasi Data Learning

Fold 1

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	49	16	2	67	0.7313433		
2	1	33	0	34	0.9705882	0.1458333	0.8541667
3	1	1	41	43	0.9534884		
Total	51	50	43	144			

Fold 2

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	41	2	0	43	0.9534884		
2	9	48	1	58	0.8275862	0.0965517	0.9034483
3	2	0	42	44	0.9545455		
Total	52	50	43	145			

Fold 3

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	43	4	0	47	0.9148936	0.0965517	0.9034483
2	8	46	1	55	0.8363636		

3	1	0	42	43	0.9767442		
Total	52	50	43	145			

Fold 4

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	41	3	3	46	0.8913043		
2	11	48	0	59	0.8135593	0.1103448	0.8896552
3	0	0	40	40	1		
Total	51	51	43	145			

Fold 5

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	40	3	1	44	0.9090909		
2	10	46	0	56	0.8214286	0.1172414	0.8827586
3	2	1	42	45	0.9333333		
Total	52	50	43	145			

Fold 6

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	39	3	0	42	0.9285714		
2	11	47	0	58	0.8103448	0.1034483	0.8965517
3	1	0	44	45	0.9777778		
Total	51	50	44	145			

Fold 7

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	42	7	0	49	0.8571429		
2	7	44	1	52	0.8461538	0.1172414	0.8827586
3	2	0	42	44	0.9545455		
Total	51	51	43	145			

Fold 8

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	45	4	3	52	0.8653846		
2	6	46	0	52	0.8846154	0.0965517	0.9034483
3	0	1	40	41	0.9756098		
Total	51	51	43	145			

Fold 9

Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	43	5	0	48	0.8958333	0.1034483	0.8965517

2	7	46	2	55	0.8363636		
3	1	0	41	42	0.9761905		
Total	51	51	43	145			
Fold 10							
Kelas Observasi	Kelas Aktual			Total	Akurasi Setiap Kelas	APER	1-APER
	1	2	3				
1	42	4	0	46	0.9130435		
2	7	46	0	53	0.8679245	0.0896552	0.9103448
3	2	0	44	46	0.9565217		
Total	51	50	44	145			

Akurasi Data Testing

```

=====
CROSS VALIDATION CLASSIFICATION TABLE
=====
Actual      Predicted Class
Class      1          2          3          Actual
-----
1          34.00      21.00      2.00      57.00
2          15.00      38.00      3.00      56.00
3           3.00      2.00      43.00      48.00
-----
PRED. TOT.      52.00      61.00      48.00      161.00
CORRECT         0.596      0.679      0.896
SUCCESS IND.    0.242      0.331      0.598
TOT. CORRECT    0.714

```

```

=====
CROSS VALIDATION CLASSIFICATION PROBABILITY TABLE
=====
Actual      Predicted Class
Class      1          2          3          Actual
-----
1          0.596      0.368      0.035      1.000
2          0.268      0.679      0.054      1.000
3          0.063      0.042      0.896      1.000
-----

```

Lampiran K. Informasi Simpul Terminal Pohon Klasifikasi dengan Pemilah Indeks Twoing

```

=====
NODE INFORMATION
=====

```

```

*****
*                Node 1: X7                *
*                N: 161                    *
*****
*                Node 2                    *      *                Node 8                *
*                N: 103                    *      *                N: 58                *
*****
Node 1 was split on X7
A case goes left if X7 = (1,2,4,5)
Improvement = 0.687      Complexity Threshold = 0.322

```

Node	Cases	Wgt	Counts	Cost	Class
1	161		161.00	0.667	1
2	103		103.00	0.470	1

```

      8      58      58.00      0.151 3
      Weighted Counts
Class      Top      Left      Right
1          57.00     55.00     2.00
2          56.00     48.00     8.00
3          48.00      0.00     48.00
      Within Node Probabilities
Class      Top      Left      Right
1          0.333     0.530     0.030
2          0.333     0.470     0.121
3          0.333     0.000     0.849
      Surrogate      Split      Assoc.      Improve.
1 X6      s 1,3,4,5      0.710      0.684
2 X3      s      15.058      0.297      0.193
3 X5      s 2      0.123      0.084
      Competitor      Split      Improve.
1 X6      1,3,4,5      0.684
2 X3      15.058      0.193
3 X2      2      0.122
4 X5      2      0.084
5 X1      2,3,4,5      0.060

```

```

*****
*                      Node 2: X2                      *
*                      N: 103                          *
*****
*****
*                      Node 3                      *
*                      N: 61                        *
*****
*                      Node 6                      *
*                      N: 42                        *
*****

```

```

Node 2 was split on X2
A case goes left if X2 = (2)
Improvement = 0.103      Complexity Threshold = 0.108
Node  Cases  Wgt  Counts      Cost  Class
2      103      103.00     0.470  1
3      61      61.00     0.299  1
6      42      42.00     0.282  2

```

```

      Weighted Counts
Class      Top      Left      Right
1          55.00     43.00     12.00
2          48.00     18.00     30.00
3          0.00      0.00      0.00
      Within Node Probabilities
Class      Top      Left      Right
1          0.530     0.701     0.282
2          0.470     0.299     0.718
3          0.000     0.000     0.000
      Surrogate      Split      Assoc.      Improve.
1 X3      s      14.500      0.170      0.034
2 X6      s 1,2,4,5      0.099      0.018
3 X1      s 1,2,4,5      0.024      0.009
      Competitor      Split      Improve.
1 X1      2,3,5      0.083
2 X7      5      0.051
3 X3      7.231      0.047
4 X5      2      0.042
5 X6      4,5      0.034

```

```

*****
*                      Node 3: X7                      *
*                      N: 61                          *
*****
=====
=      Terminal Node 1      =      *      Node 4      *
=      N: 32                =      *      N: 29      *
=====
*****

```

```

Node 3 was split on X7
A case goes left if X7 = (1,5)
Improvement = 0.031      Complexity Threshold = 0.012
Node  Cases  Wgt  Counts      Cost  Class
3      61      61.00     0.299  1

```

```

-1      32      32.00      0.159 1
 4      29      29.00      0.453 1
                                Weighted Counts
Class   Top      Left      Right
1       43.00    27.00    16.00
2       18.00     5.00    13.00
3        0.00     0.00     0.00
                                Within Node Probabilities
Class   Top      Left      Right
1       0.701    0.841    0.547
2       0.299    0.159    0.453
3        0.000    0.000    0.000
Surrogate Split
1 X1      s 1,3,4
2 X4      s 2
3 X6      s 1,4
4 X3      r 9.847
5 X5      s 2
Competitor Split
1 X3      8.500
2 X6      1,2,5
3 X1      2,3,5
4 X5      2
5 X4      2
                                Assoc.   Improve.
1 X1      0.208      0.006
2 X4      0.140      .671113E-03
3 X6      0.140      .437944E-03
4 X3      0.105      0.005
5 X5      0.035      0.008
Competitor Split
1 X3      8.500
2 X6      1,2,5
3 X1      2,3,5
4 X5      2
5 X4      2
                                Improve.
1 X3      0.022
2 X6      0.014
3 X1      0.011
4 X5      0.008
5 X4      .671113E-03
*****
*                Node 4: X3                *
*                N: 29                      *
*****
=====
=      Terminal Node 2      =      *                Node 5                *
=                N: 9      =      *                N: 20                *
=====
Node 4 was split on X3
A case goes left if X3 <= 9.241
Improvement = 0.016          Complexity Threshold = 0.018
Node   Cases  Wgt Counts  Cost Class
 4     29     29.00    0.453 1
-2     9      9.00     0.225 1
 5     20     20.00    0.446 2
                                Weighted Counts
Class   Top      Left      Right
1       16.00    7.00     9.00
2       13.00     2.00    11.00
3        0.00     0.00     0.00
                                Within Node Probabilities
Class   Top      Left      Right
1       0.547    0.775    0.446
2       0.453    0.225    0.554
3        0.000    0.000    0.000
Competitor Split
1 X1      2,3,5
2 X4      2
3 X6      1,3,4,5
4 X5      2
5 X7      4
                                Improve.
1 X1      0.011
2 X4      0.009
3 X6      0.007
4 X5      0.006
5 X7      .698970E-03
*****
*                Node 5: X5                *
*                N: 20                      *
*****
=====
=      Terminal Node 3      =      =      Terminal Node 4      =
=                N: 14      =      =                N: 6                =
=====
Node 5 was split on X5
A case goes left if X5 = (2)
Improvement = 0.030          Complexity Threshold = 0.023
Node   Cases  Wgt Counts  Cost Class
 5     20     20.00    0.446 2
-3     14     14.00    0.282 2
-4     6      6.00     0.169 1

```

Weighted Counts			
Class	Top	Left	Right
1	9.00	4.00	5.00
2	11.00	10.00	1.00
3	0.00	0.00	0.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.446	0.282	0.831
2	0.554	0.718	0.169
3	0.000	0.000	0.000

Surrogate		Split	Assoc.	Improve.
1	X3	s 13.500	0.329	0.010
2	X1	s 1,2,3,4	0.166	0.008
3	X6	s 1,2,4,5	0.160	0.004

Competitor		Split	Improve.
1	X3	9.847	0.011
2	X1	2,3,5	0.010
3	X6	2,4,5	0.005
4	X7	4	.504564E-03
5	X4	2	.392010E-03

```

*****
*                               *
*           Node 6: X1         *
*                               *
*           N: 42             *
*                               *
*****

```

```

=====
= Terminal Node 5 = * Node 7 *
= N: 8 = * N: 34 *
=====
*****

```

Node 6 was split on X1
A case goes left if X1 = (2,3,4,5)
Improvement = 0.081 Complexity Threshold = 0.035

Node	Cases	Wgt	Counts	Cost	Class
6	42		42.00	0.282	2
-5	8		8.00	0.127	1
7	34		34.00	0.145	2

Weighted Counts			
Class	Top	Left	Right
1	12.00	7.00	5.00
2	30.00	1.00	29.00
3	0.00	0.00	0.00

Within Node Probabilities			
Class	Top	Left	Right
1	0.282	0.873	0.145
2	0.718	0.127	0.855
3	0.000	0.000	0.000

Surrogate		Split	Assoc.	Improve.
1	X3	s 6.231	0.249	0.025
2	X6	s 2,5	0.122	0.035

Competitor		Split	Improve.
1	X7	4,5	0.040
2	X5	2	0.036
3	X6	2,5	0.035
4	X3	6.231	0.025
5	X4	2	0.011

```

*****
*                               *
*           Node 7: X7         *
*                               *
*           N: 34             *
*                               *
*****

```

```

=====
= Terminal Node 6 = = Terminal Node 7 =
= N: 2 = = N: 32 =
=====

```

Node 7 was split on X7
A case goes left if X7 = (5)
Improvement = 0.036 Complexity Threshold = 0.012

Node	Cases	Wgt	Counts	Cost	Class
7	34		34.00	0.145	2
-6	2		2.00	0.000	1
-7	32		32.00	0.092	2

Weighted Counts

```

Class      Top      Left      Right
1          5.00     2.00     3.00
2          29.00    0.00     29.00
3          0.00     0.00     0.00
          Within Node Probabilities
Class      Top      Left      Right
1          0.145    1.000    0.092
2          0.855    0.000    0.908
3          0.000    0.000    0.000
Competitor      Split
1 X5            2
2 X6            2,5
3 X3            16.994
4 X4            2
          Improve.
          0.013
          0.006
          0.006
          .380817E-03
          *****
          *                Node 8: X6                *
          *                N: 58                    *
          *****
          =====
          =      Terminal Node 8      =      *      Node 9      *
          =      N: 41                =      *      N: 17      *
          =====
          *****

Node 8 was split on X6
A case goes left if X6 = (2,5)
Improvement = 0.095      Complexity Threshold = 0.014
Node  Cases  Wgt Counts  Cost Class
  8      58      58.00    0.151 3
 -8      41      41.00    0.000 3
  9      17      17.00    0.550 3
          Weighted Counts
Class      Top      Left      Right
1          2.00     0.00     2.00
2          8.00     0.00     8.00
3          48.00    41.00    7.00
          Within Node Probabilities
Class      Top      Left      Right
1          0.030    0.000    0.108
2          0.121    0.000    0.441
3          0.849    1.000    0.450
Surrogate      Split      Assoc.  Improve.
1 X3            3.489      0.064  .645310E-03
Competitor      Split      Improve.
1 X1            2,3,4,5
2 X2            2
3 X3            15.058
4 X5            2
5 X4            2
          *****
          *                Node 9: X3                *
          *                N: 17                    *
          *****
          *****
          *                Node 10                *      =      Terminal Node 11      =
          *                N: 12                    *      =      N: 5                    *
          *****
          *****

Node 9 was split on X3
A case goes left if X3 <= 15.500
Improvement = 0.038      Complexity Threshold = 0.020
Node  Cases  Wgt Counts  Cost Class
  9      17      17.00    0.550 3
 10      12      12.00    0.359 2
 -11     5       5.00    0.174 3
          Weighted Counts
Class      Top      Left      Right
1          2.00     1.00     1.00
2          8.00     8.00     0.00
3          7.00     3.00     4.00
          Within Node Probabilities
Class      Top      Left      Right
1          0.108    0.079    0.174

```

2	0.441	0.641	0.000		
3	0.450	0.280	0.826		
	Surrogate	Split		Assoc.	Improve.
1	X1	s 2,3,4,5		0.439	0.019
2	X6	s 3,4		0.030	.688061E-03
	Competitor	Split			Improve.
1	X2	2			0.021
2	X1	2,3,4,5			0.019
3	X6	4			0.006
4	X5	2			0.006
5	X4	2			0.002

```

*****
*           Node 10: X3           *
*           N: 12                 *
*****

```

```

=====
=      Terminal Node 9      =      Terminal Node 10      =
=      N: 3                 =      N: 9                 =
=====

```

Node 10 was split on X3
 A case goes left if X3 <= 8.500
 Improvement = 0.044 Complexity Threshold = 0.014

Node	Cases	Wgt	Counts	Cost	Class
10	12		12.00	0.359	2
-9	3		3.00	0.296	3
-10	9		9.00	0.127	2

Weighted Counts

Class	Top	Left	Right
1	1.00	1.00	0.00
2	8.00	0.00	8.00
3	3.00	2.00	1.00

Within Node Probabilities

Class	Top	Left	Right
1	0.079	0.296	0.000
2	0.641	0.000	0.873
3	0.280	0.704	0.127

	Surrogate	Split		Assoc.	Improve.
1	X2	s 1		0.050	0.041
2	X5	s 1		0.050	0.004
	Competitor	Split			Improve.
1	X2	2			0.041
2	X6	1,4			0.004
3	X5	2			0.004
4	X4	2			.782835E-03
5	X1	2,3,4,5			.782835E-03

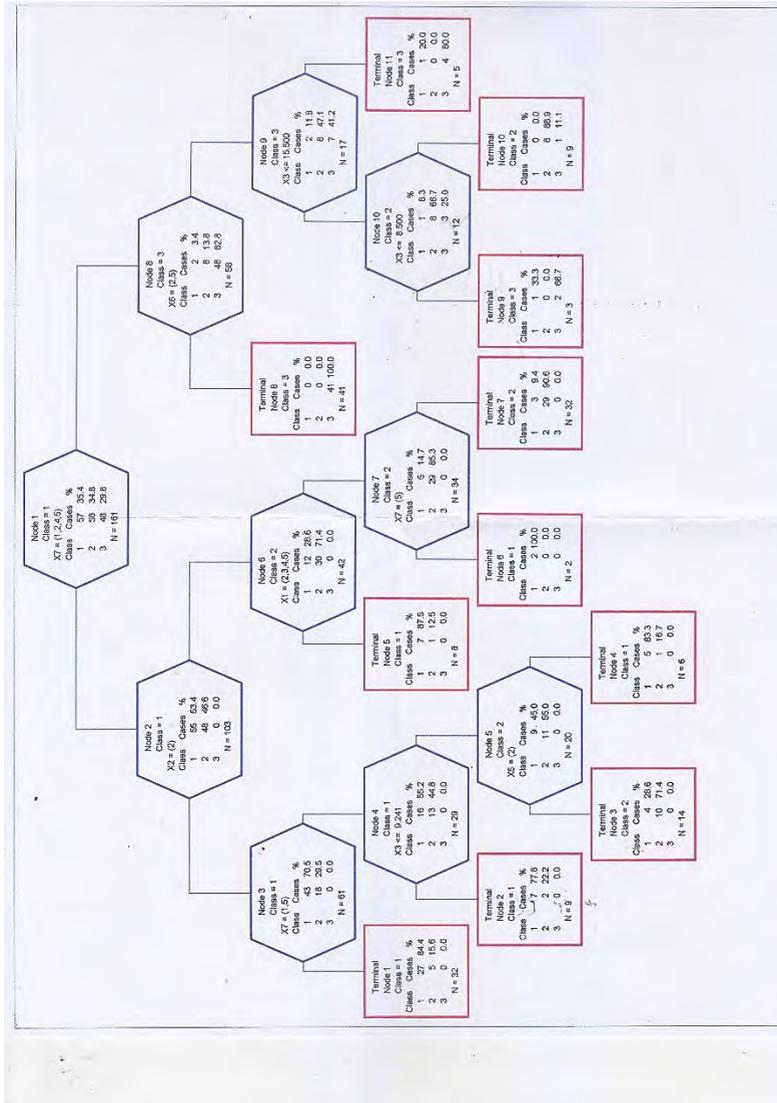
=====

[Breiman adjusted cost, lambda = 0.051]

Node	Class	Wgt	Count	N	Prob	Cost	Parent Complexity
1	1		32.00	32	0.188	0.159	0.012 [0.335]
	1		27.00	27	0.841		
	2		5.00	5	0.159		
	3		0.00	0	0.000		
2	1		9.00	9	0.053	0.225	0.018 [0.631]
	1		7.00	7	0.775		
	2		2.00	2	0.225		
	3		0.00	0	0.000		
3	2		14.00	14	0.083	0.282	0.023 [0.596]
	1		4.00	4	0.282		
	2		10.00	10	0.718		

3	0.00	0	0.000		
4 1	6.00	6	0.035	0.169	0.023 [0.658]
1	5.00	5	0.831		
2	1.00	1	0.169		
3	0.00	0	0.000		
5 1	8.00	8	0.047	0.127	0.035 [0.557]
1	7.00	7	0.873		
2	1.00	1	0.127		
3	0.00	0	0.000		
6 1	2.00	2	0.012	0.000	0.012 [0.673]
1	2.00	2	1.000		
2	0.00	0	0.000		
3	0.00	0	0.000		
7 2	32.00	32	0.190	0.092	0.012 [0.267]
1	3.00	3	0.092		
2	29.00	29	0.908		
3	0.00	0	0.000		
8 3	41.00	41	0.285	0.000	0.014 [0.125]
1	0.00	0	0.000		
2	0.00	0	0.000		
3	41.00	41	1.000		
9 3	3.00	3	0.020	0.296	0.014 [0.893]
1	1.00	1	0.296		
2	0.00	0	0.000		
3	2.00	2	0.704		
10 2	9.00	9	0.055	0.127	0.014 [0.526]
1	0.00	0	0.000		
2	8.00	8	0.873		
3	1.00	1	0.127		
11 3	5.00	5	0.034	0.174	0.020 [0.672]
1	1.00	1	0.174		
2	0.00	0	0.000		
3	4.00	4	0.826		

Lampiran L. Pohon Klasifikasi Optimal dengan Menggunakan Pemilah Indeks Twoing



**Lampiran M. Surat Keterangan Penelitian di RSJ Menur Provinsi
Jawa Timur**



PEMERINTAH PROVINSI JAWA TIMUR
RUMAH SAKIT JIWA MENUR
Jl. Raya Menur No. 120 Telp. 5021635 – 5021637 Surabaya

SURAT KETERANGAN
NOMOR : 423.4/2006 /305/2015

Yang bertanda tangan di bawah ini :

Nama : dr. Melani Handoyo
Jabatan : Kepala Instalasi DiklatLit
RSJ Menur Provinsi Jawa Timur

Menerangkan bahwa :

Nama : PUTRI MILAKHUL KHASANAH
N I M : 1311100116
Inststitusi : Fakultas Matematika Dan Ilmu Pengetahuan Alam
Jurusan Statistika FMIPA
Institut Teknologi Sepuluh Nopember Surabaya

Telah melakukan Penelitian di RS Jiwa Menur Surabaya.

Judul : Klasifikasi Penderita Retardasi Mental di Rumah
Sakit Jiwa Menur Surabaya Menggunakan
Classification and Regression Tree (CART)
Tanggal : 21 Februari s/d 02 Maret 2015

Surabaya, 03 Maret 2015
Kepala Instalasi DiklatLit

RS Jiwa Menur



“Halaman ini sengaja dikosongkan”