



TUGAS AKHIR - KS184822

**KLASIFIKASI HASIL TES DETEKSI KANKER
PAYUDARA BERDASARKAN GEJALA KLINIS
MENGUNAKAN METODE *CLASSIFICATION TREE*
DI RSUD NUR HIDAYAH BANTUL**

**DESINTYA RACHMA ANGGRAENI PUTRI
NRP 062115 4000 0113**

**Dosen Pembimbing
Dr. Santi Wulan Purnami, M.Si.**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019**



TUGAS AKHIR - KS184822

**KLASIFIKASI HASIL TES DETEKSI KANKER
PAYUDARA BERDASARKAN GEJALA KLINIS
MENGUNAKAN METODE *CLASSIFICATION TREE*
DI RSUD NUR HIDAYAH BANTUL**

**DESINTYA RACHMA ANGGRAENI PUTRI
NRP 062115 4000 0113**

**Dosen Pembimbing
Dr. Santi Wulan Purnami, M.Si.**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019**



FINAL PROJECT - KS184822

**CLASSIFICATION THE RESULTS OF BREAST
CANCER DETECTION TEST BASED ON CLINICAL
SYMPTOMS USING CLASSIFICATION TREE METHOD
IN RSUD NUR HIDAYAH BANTUL**

**DESINTYA RACHMA ANGGRAENI PUTRI
SN 062115 4000 0113**

**Supervisor
Dr. Santi Wulan Purnami, M.Si.**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019**

LEMBAR PENGESAHAN

KLASIFIKASI HASIL TES DETEKSI KANKER PAYUDARA BERDASARKAN GEJALA KLINIS MENGGUNAKAN METODE *CLASSIFICATION TREE* DI RSUD NUR HIDAYAH BANTUL

TUGAS AKHIR

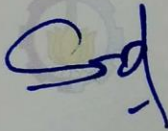
Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Statistika
pada

Program Studi Sarjana Departemen Statistika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember

Oleh :


Desintya Rachma Anggraeni Putri
NRP. 062115 4000 0113

Disetujui oleh Pembimbing :
Dr. Santi Wulan Purnami, M.Si.
NIP. 19720923 199803 2 001

()



Mengetahui,
Ketua Departemen Statistika



Dr. Suhartono
NIP. 19710929 199512 1 001

SURABAYA, JULI 2019

(Halaman ini sengaja dikosongkan)

**KLASIFIKASI HASIL TES DETEKSI KANKER PAYUDARA
BERDASARKAN GEJALA KLINIS MENGGUNAKAN
METODE *CLASSIFICATION TREE* DI RSUD NUR
HIDAYAH BANTUL**

Nama Mahasiswa : Desintya Rachma Anggraeni Putri
NRP : 062115 4000 0113
Departemen : Statistika
Dosen Pembimbing : Dr. Santi Wulan Purnami, M.Si.

Abstrak

*Kanker payudara adalah tumor ganas yang terbentuk dari sel-sel payudara yang tumbuh dan berkembang biak tanpa terkendali sehingga dapat menyebar di antara jaringan atau organ di sekitar payudara atau bahkan dapat menyebar ke bagian tubuh lainnya. Kanker payudara merupakan kanker yang paling sering terjadi diantara wanita. Tingginya kasus kanker payudara menjadikan perhatian khusus bagi kaum wanita untuk selalu waspada terhadap terjadinya kanker payudara. Tujuan dari penelitian ini ingin dilakukan klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis menggunakan metode *Classification Tree* dan *Bootstrap Aggregating (Bagging)*. Selanjutnya dilakukan evaluasi ketepatan klasifikasi dengan meninjau *accuracy*, *sensitifity*, *specificity* dan *AUC*. Hasil dari penelitian ini menunjukkan bahwa model pohon klasifikasi optimal yang terbentuk adalah memiliki 4 simpul terminal dengan kedalaman 2. Sedangkan model terbaik yang terbentuk dengan menggunakan metode *Bootstrap Aggregating* adalah dengan 20 replikasi *bootstrap*. Hasil dari perbandingan ketepatan klasifikasi pohon yang diperoleh dengan menggunakan metode *Bootstrap Aggregating* tidak lebih baik dari metode *Classification Tree*, karena pada penelitian ini metode *Classification Tree* menghasilkan ketepatan klasifikasi lebih tinggi dibandingkan dengan metode *Bootstrap Aggregating*. Walaupun hasil dari *Bootstrap Aggregating* tidak selalu meningkatkan hasil akurasi dari metode *Classification Tree*, akan tetapi dapat menghasilkan model yang konsisten.*

Kata Kunci: Bootstrap Aggregating, Classification Tree, Gejala Klinis, Kanker Payudara, K-Fold Cross Validation

(Halaman ini sengaja dikosongkan)

CLASSIFICATION THE RESULTS OF BREAST CANCER DETECTION TEST BASED ON CLINICAL SYMPTOMS USING CLASSIFICATION TREE METHOD IN RSUD NUR HIDAYAH BANTUL

Name : Desintya Rachma Anggraeni Putri
Student Number : 062115 4000 0113
Department : Statistics
Supervisor : Dr. Santi Wulan Purnami, M.Si.

Abstract

Breast cancer is a malignant tumor that starts when cells in the breast begin to grow and out of control, so it can spread between tissues, organs around the breast or to other parts of body. Breast cancer occurs almost entirely in woman. The high cases of breast cancer make special attention for women to always be watchful of breast cancer's occurrence. The purpose of this research is classified the result of breast cancer detection tests based on clinical symptoms using Classification Tree and Bootstrap Aggregating (Bagging) method. The evaluation are based on accuracy, sensitivity, specificity and AUC. This research shows that the optimal classification tree model has 4 terminal nodes with the depth is 2. While the best model formed using The Bootstrap Aggregating method is with 20 bootstrap's replications. The Bootstrap Aggregating method is not always better than the Classification Tree method because in this research the accuracy of classification tree is higher than the Bootstrap Aggregating method. Even though the result of Bootstrap Aggregating is not always increasing the accuracy of the Classification Tree method, it can produce a consistent model.

Keywords: *Bootstrap Aggregating, Breast Cancer, Classification Tree, Clinical Symptoms, K-Fold Cross Validation*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan hidayah dan karunia-Nya, sehingga dalam penyusunan laporan Tugas Akhir yang berjudul “**Klasifikasi Hasil Tes Deteksi Kanker Payudara Berdasarkan Gejala Klinis Menggunakan Metode *Classification Tree* Di RSUD Nur Hidayah Bantul**” dapat diselesaikan dengan baik dan tepat pada waktunya untuk memenuhi persyaratan akademis dari Departemen Statistika Institut Teknologi Sepuluh Nopember.

Dengan terlaksananya Tugas Akhir dan pembuatan laporan ini maka tak lupa penulis mengucapkan terima kasih kepada :

1. Bapak, ibu dan mas Daru atas segala do’a, nasehat, kasih sayang, dan dukungan yang diberikan kepada penulis demi kesuksesan dan kelancaran dalam mengerjakan laporan Tugas Akhir ini.
2. Ibu Dr. Santi Wulan Purnami, M.Si. selaku Ketua Program Studi Sarjana Departemen Statistika FMKSD ITS dan dosen pembimbing yang senantiasa memberikan perhatian, ilmu, bimbingan maupun pengarahan baik selama menyelesaikan Tugas Akhir ini.
3. Bapak Dr. Bambang Widjanarko Otok., S.Si., M.Si. dan Jerry Dwi Trojoyo P., S.Si., M.Si., Ph.D. selaku dosen penguji yang senantiasa memberikan kritik dan saran demi kesempurnaan Tugas Akhir ini.
4. Seluruh dosen Departemen Statistika ITS atas segala ilmu yang telah diberikan selama menjadi mahasiswa aktif di Departemen Statistika ITS.
5. Sahabat-sahabat penulis : Shindy, Dian, Shisel, Yogi, Irsa, Rima, Yesika, Diki, Ajeng yang selama ini telah membantu, mendukung, dan mendengarkan keluh kesah penulis selama masa perkuliahan berlangsung.
6. Teman-teman seperjuangan TA, khususnya Dinda, Chiska, Liana yang selama ini telah berjuang bersama dan saling memberikan semangat.
7. Mas Amri, mas Musa, Juwita, Rhavida terkait diskusi mengenai metode CART dan *Bootstrap Aggregating*.

8. Serta teman-teman angkatan 2015 Vivacious dan semua pihak yang telah membantu dan memberikan dukungan, baik secara langsung maupun tidak langsung yang tidak dapat disebutkan 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 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
PAGE COVER	iii
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR TABEL	xvii
DAFTAR GAMBAR	xix
DAFTAR LAMPIRAN	xxi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	6
BAB II TINJAUAN PUSTAKA	7
2.1 Statistika Deskriptif.....	7
2.2 <i>Classification and Regression Tree (CART)</i>	7
2.3.1 Pembentukan Pohon Klasifikasi Maksimal.....	10
2.3.2 Pemangkasan Pohon Klasifikasi Maksimal (<i>Pruning</i>).....	13
2.3.3 Penentuan Pohon Klasifikasi Optimal.....	15
2.3 <i>K-Fold Cross Validation</i>	16
2.4 <i>Bootstrap</i>	17
2.4.1 <i>Bootstrap Aggregating</i>	18
2.5 Evaluasi Ketepatan Klasifikasi.....	19
2.6 Kanker Payudara	21
2.6.1 Gejala Klinis.....	23
2.6.2 Faktor Predisposisi	23
BAB III METODOLOGI PENELITIAN	25
3.1 Kerangka Konsep	25

3.2	Desain Penelitian	26
3.3	Sumber Data	27
3.4	Variabel Penelitian	27
3.5	Struktur Data	30
3.6	Langkah Analisis	31
3.7	Diagram Alir.....	32
BAB V	ANALISIS DAN PEMBAHASAN	35
4.1	Karakteristik Data.....	35
4.1.1	Hasil Tes Deteksi Terhadap Kanker Payudara.....	35
4.1.2	Usia Terhadap Status Pemeriksaan Kanker Payudara	36
4.1.3	Status Pernikahan Terhadap Status Pemeriksaan Kan-ker Payudara	37
4.1.4	Gejala Klinis Terhadap Status Pemeriksaan Kanker Payudara	38
4.1.5	Kualitas Tidur Terhadap Status Pemeriksaan Kanker Payudara	40
4.1.6	Riwayat Pengobatan Terhadap Status Pemerik- saan Kanker Payudara	41
4.2	Klasifikasi Hasil Tes Deteksi Kanker Payudara dengan Menggunakan Metode <i>Classification Tree</i>	42
4.2.1	Pembentukan Pohon Klasifikasi Maksimal.....	44
4.2.2	Pemangkasan Pohon Klasifikasi Maksimal	47
4.2.3	Penentuan Pohon Klasifikasi Optimal.....	48
4.2.4	Evaluasi Ketepatan Klasifikasi Menggunakan Metode <i>Classification Tree</i>	52
4.3	Klasifikasi Hasil Tes Deteksi Kanker Payudara dengan Menggunakan Metode <i>Bootstrap Aggregating</i>	54
4.3.1	Pemodelan <i>Bootstrap Aggregating</i> Dengan 10, 15, 20, 25, 30, dan 35 Replikasi <i>Bootstrap</i>	54
4.3.2	Evaluasi Ketepatan Klasifikasi Menggunakan Metode <i>Bootstrap Aggregating</i> Dengan 10, 15, 20, 25, 30, dan 35 Replikasi <i>Bootstrap</i>	57
4.4	Perbandingan Ketepatan Klasifikasi Menggunakan Metode <i>Classification Tree</i> dan <i>Bootstrap Aggregating</i> ..	58

BAB V KESIMPULAN DAN SARAN	61
5.1 Kesimpulan.....	61
5.2 Saran.....	62
DAFTAR PUSTAKA	63
LAMPIRAN.....	67
BIODATA PENULIS	108

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 2.1 <i>Confusion Matrix</i>	20
Tabel 2.2 Kategori Pengklasifikasian Model Berdasarkan Nilai AUC	21
Tabel 3.1 Variabel Penelitian	27
Tabel 3.2 Struktur Data Penelitian	31
Tabel 4.1 Banyaknya Kemungkinan Pemilah Variabel Prediktor	43
Tabel 4.2 Urutan Variabel Pembentuk Pohon Klasifikasi Maksimal	45
Tabel 4.3 Urutan Pembentukan Pohon Klasifikasi (<i>Tree Sequence</i>)	48
Tabel 4.4 Karakteristik Wanita Berdasarkan Simpul Terminal	52
Tabel 4.5 Evaluasi Performansi Pada Data <i>Training</i>	52
Tabel 4.6 <i>Confution Matrix</i> Data <i>Testing</i>	53
Tabel 4.7 Perbandingan Keباikan Hasil Klasifikasi Menggunakan Metode <i>Bootstrap Aggregating</i>	58
Tabel 4.8 Perbandingan Keباikan Hasil Klasifikasi Menggunakan Metode <i>Classification Tree</i> dan <i>Bootstrap Aggregating</i>	59

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Diagram Pohon Klasifikasi	10
Gambar 2.2 Ilustrasi <i>K-Fold Cross Validation</i>	17
Gambar 2.3 Ilustrasi Bagging CART	19
Gambar 2.4 Hubungan Kurva ROC Dengan AUC	21
Gambar 2.5 Anatomi Payudara Manusia	22
Gambar 3.1 Kerangka Konsep Penelitian	25
Gambar 3.2 Diagram Alir Penelitian	33
Gambar 4.1 Hasil Tes Deteksi Kanker Payudara.....	36
Gambar 4.2 Usia Terhadap Status Pemeriksaan Kanker Payudara.....	37
Gambar 4.3 Status Pernikahan Terhadap Status Kanker Payudara.....	38
Gambar 4.4 Gejala Klinis Terhadap Status Kanker Payudara.....	39
Gambar 4.5 Kualitas Tidur Terhadap Status Kanker Payudara ...	41
Gambar 4.6 Riwayat Pengobatan Terhadap Status Kanker Payudara.....	42
Gambar 4.7 Bentuk Pohon Klasifikasi Maksimal.....	46
Gambar 4.8 Plot <i>Relative Cost</i> Dan Banyaknya Simpul Terminal	47
Gambar 4.9 Topologi Pohon Klasifikasi Optimal	48
Gambar 4.10 Bentuk Pohon Klasifikasi Optimal	50
Gambar 4.11 Diagram Pohon Klasifikasi Hasil <i>Bootstrap</i>	56

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Asli Penderita dan Bukan Penderita Kanker Payudara.....	67
Lampiran 2. Data Hasil Pengkodean	69
Lampiran 3. <i>Crosstab</i> Variabel Penelitian.....	71
Lampiran 4. Informasi Pemilahan Simpul Pohon Klasifikasi Optimal	75
Lampiran 5. Informasi Proporsi Setiap Kelas Pada Setiap Simpul Pohon Klasifikasi Optimal	77
Lampiran 6. Informasi Pemilahan Simpul Pohon Klasifikasi Menggunakan Metode <i>Bootstrap Aggregating</i> Untuk 10 Replikasi	77
Lampiran 7. Ketepatan Klasifikasi Dengan Metode <i>Bootstrap Aggregating</i>	103
Lampiran 8. Salah Satu Hasil Tes Deteksi Kanker Payudara Menggunakan Ultrasonografi (USG)	105
Lampiran 9. Surat Keterangan Pengambilan Data	106
Lampiran 10. Surat Ijin Pengambilan Data Dari Pihak RSUD Nur Hidayah Bantul	107

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tubuh manusia terdiri dari ratusan miliar hingga triliunan sel. Setiap saat, sel-sel tersebut akan terus berkembang biak dan membelah diri. Ada sel yang nantinya akan mati dan rusak dikarenakan sudah tidak berfungsi kembali. Kanker hanya akan muncul apabila terjadi kesalahan pada siklus sel. Satu dari sekian banyak sel mungkin akan membelah diri secara tidak sempurna yang berakibat sel ini berubah sifat menjadi bahaya untuk tubuh (Hartwell dan Kastan, 1994). Menurut Pusat Data dan Informasi Kementerian Kesehatan RI (2016), kanker adalah suatu kondisi di mana terjadinya pertumbuhan yang tidak normal (abnormal) dari sel-sel jaringan tubuh dan bersifat ganas.

Menurut perkiraan dari organisasi kesehatan dunia (WHO) kanker adalah pertama atau kedua penyebab utama kematian sebelum usia 70 tahun di 91 negara, dan peringkat ketiga atau keempat dalam 22 negara tambahan. Statistik kanker global tahun 2018 menunjukkan bahwa berdasarkan perkiraan GLOBOCAN akan ada sekitar 18,1 juta kasus kanker baru (17,0 juta termasuk kanker kulit nonmelanoma) dan 9,6 juta kematian akibat kanker (9,5 juta termasuk kanker kulit nonmelanoma) (Bray dkk, 2018).

Secara global, kejadian kanker payudara pada wanita telah sangat menggambarkan tingkat kejadian kanker payudara itu sendiri dibandingkan dengan laki-laki yang sedikit mengalami kanker payudara (Almutlaq dkk, 2017). Menurut data GLOBOCAN, *International Agency for Research on Cancer* (IARC), diketahui bahwa kanker payudara adalah kanker paling umum kedua di dunia setelah kanker paru-paru dan merupakan kanker yang paling sering terjadi diantara wanita dengan perkiraan 2,08 juta kasus kanker baru yang didiagnosis pada tahun 2018 (11,6% dari semua kanker). Kasus kanker payudara menempati urutan pertama dengan persentase kasus baru (setelah dikontrol oleh umur) tertinggi yaitu sebesar 43,5% dan menempati urutan kedua dengan persentase kematian

(setelah dikontrol oleh umur) yaitu sebesar 11,2%. Di Indonesia kasus kanker payudara menyumbang 16,7% (58.256) dari 348.809 kasus baru pada kanker payudara di 2018 pada semua umur baik laki-laki maupun perempuan. Dari 58.256 kasus baru tersebut pada kanker payudara sebesar 12,75% (22.692) diantaranya meninggal.

Penyebab kanker payudara sampai saat ini belum dapat diketahui secara pasti, diduga penyebab kanker payudara adalah multifaktor. Salah satu program pemerintah untuk mendeteksi adanya kanker payudara adalah SADARI (Pemeriksaan Payudara Sendiri). Sadari merupakan pengembangan kepedulian wanita terhadap kondisi payudaranya sendiri. Tindakan ini dilengkapi dengan tahapan-tahapan khusus untuk mendeteksi secara awal kanker payudara untuk dapat mengetahui perubahan-perubahan yang terjadi pada payudara. Pada umumnya perubahan pada payudara yang dapat dilihat sebagai kelainan yang terjadi dan perlu mendapat perhatian diantaranya adalah perubahan yang terjadi pada bentuk dan ukuran payudara, terdapat benjolan di sekitar payudara dan ketiak, nyeri yang tak kunjung hilang di bagian tertentu, keluarnya cairan bercak darah atau encer secara tiba-tiba, dsb (KemenKes RI, 2016). Selain gejala fisik, menurut Akman, dkk (2015) telah membuktikan bahwa seseorang yang menderita kanker payudara pada stadium lanjut memerlukan penanganan lebih intensif mulai dari terapi seperti radioterapi, kemoterapi atau kombinasi keduanya yang berdampak langsung pada kondisi fisik maupun psikologi pasien. Kondisi tersebut juga mengakibatkan kebutuhan istirahat pasien terganggu dan kualitas tidurnya dinyatakan buruk. Adapun riwayat pengobatan diduga juga memiliki pengaruh yang kuat dengan penderita kanker (Muhammad, Merriam, dan Suhami, 2012). Di Malaysia, ada beberapa pilihan pengobatan konvensional yang tersedia untuk memerangi kanker, termasuk operasi, radioterapi, kemoterapi, terapi hormon, imunoterapi dan simtomatik dan suportif terapi (Lim, 2002). Namun pengobatan tersebut memiliki kemungkinan kecil untuk sembuh, sehingga pasien penderita kanker mulai beralih ke pengobatan alternatif seperti pengobatan komplementer maupun tradisional. Pengobatan komplementer dan alternatif umumnya digu-

nakan oleh penderita kanker payudara sebagai mengatasi mekanisme untuk memerangi penyakit itu (Shaharudin dkk, 2011).

Deteksi dini pada kanker payudara mulai dari pemeriksaan mammografi dan SADARI (Pemeriksaan Payudara Sendiri) sangatlah perlu untuk mengetahui bagaimanakah karakteristik dari penderita kanker payudara itu sendiri, sehingga tidak salah dalam melakukan penanganan lebih lanjut baik dalam pemberian terapi maupun pengobatan terhadap kanker tersebut (KemenKes RI, 2016). Selain itu pentingnya deteksi dini pada kanker payudara untuk menghindari penanganan kanker di stadium lanjut sehingga lebih mudah untuk diobati.

Merujuk pada penelitian sebelumnya yang dilakukan oleh Lumachi, dkk (2002) mengenai keluhan dan resiko kanker payudara, studi kasus berbasis populasi pada 2879 wanita. Pada penelitian ini dibagi menjadi 3 kelompok dengan kriteria grup A usia < 41 tahun, grup B usia 41-55 tahun, dan grup C usia > 55 tahun. Hasil penelitian ini menunjukkan bahwa sebanyak 318 (11%) wanita memiliki kanker payudara yang dikonfirmasi secara histologis (grup A = 3,5%, grup B = 30,5% dan grup C = 66%) terhitung juga untuk 3,2% ; 16,4% dan 12% pasien dengan keluhan nyeri, benjolan pada payudara dan keluarnya cairan dari puting. Keluhan nyeri merupakan keluhan yang paling umum terjadi pada grup A, sedangkan keluhan dengan benjolan pada payudara terjadi pada grup B dan C.

Kemudian penelitian yang dilakukan oleh Bawazir, dkk (2018) menunjukkan bahwa 89% penderita kanker payudara mengalami benjolan pada payudara, 75,7% mengalami perubahan bentuk dan ukuran payudara, 75,1% mengalami perubahan pada puting, 70,7% keluar cairan atau bercak darah pada puting susu, 65,9% mengalami kemerahan pada kulit payudara, 50,5% mengalami penarikan puting susu ke dalam, 48% mengalami perubahan bentuk kulit payudara seperti kulit jeruk dan 42,6% mengalami lekukan pada kulit payudara seperti lesung pipi. Begitu juga dengan penelitian yang dilakukan oleh Seltzer (2004), menunjukkan bahwa penderita kanker payudara dengan keluhan terbanyak adalah

benjolan pada payudara terjadi pada wanita usia ≥ 70 tahun. Selain itu mereka juga mengalami nyeri dan keluar cairan dari puting susu.

Salah satu metode klasifikasi dan prediksi yang sangat kuat dan sering digunakan dalam suatu penelitian adalah metode *Classification and Regression Tree* (CART). Metode CART merupakan metode pengembangan dari *Decision Tree* berbasis regresi. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian.

Kemudian pada tahun 1996, Breiman mengembangkan salah satu metode *ensemble* yang relatif baru namun telah menjadi populer yang mana dikenal dengan nama metode Bagging (*Bootstrap Aggregating*). Ide dasar metode Bagging ini adalah menggunakan *resampling* acak dengan pengembalian pada *dataset* awal sehingga diperoleh *dataset* baru untuk membangkitkan pohon klasifikasi dengan banyak versi yang kemudian dikombinasikan untuk memperoleh prediksi akhir. Kombinasi dari banyak versi pohon klasifikasi diharapkan mampu memperbaiki akurasi klasifikasi dari pohon klasifikasi tunggal. Bagging juga bertujuan untuk mengurangi variansi prediktor dan menghindari terjadinya *overfitting*.

Penelitian yang dilakukan oleh Muttaqin (2013) mengenai metode *ensemble* pada CART untuk perbaikan klasifikasi kemiskinan di Kabupaten Jombang memberikan hasil bahwa secara umum metode *ensemble* menghasilkan akurasi klasifikasi yang lebih baik dibandingkan dengan pohon klasifikasi CART. Bagging CART mampu meningkatkan 1-APER pohon klasifikasi CART sebesar 4,52%. Dan berkaitan dengan stabilitas akurasi prediksi, metode *ensemble* terbaik adalah Bagging CART, artinya lebih cepat mencapai akurasi konvergen atau lebih sedikit memerlukan pohon klasifikasi untuk mencapai akurasi prediksi yang konvergen. Selain itu penelitian yang dilakukan oleh Istiqomah (2015) mengenai penerapan klasifikasi dan *Bootstrap Aggregating* dalam klasifikasi usia menarche menunjukkan bahwa prosedur *Bootstrap Aggregating* (Bagging) mampu meningkatkan keakuratan pendugaan po-

hon klasifikasi tunggal *Classification And Regression Tree* (CART) dari 60% menjadi 62,50%.

Berdasarkan keunggulan metode *Classification and Regression Tree* (CART) dan *Bootstrap Aggregating* (Bagging) yang digunakan pada penelitian sebelumnya, maka pada penelitian ini akan digunakan metode *Classification Tree* dan *Bootstrap Aggregating* untuk klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara. Selanjutnya akan dilakukan perbandingan evaluasi ketepatan model diantara kedua metode tersebut dengan meninjau *accuracy*, *sensitifity*, *specificity*, dan nilai AUC. Dengan adanya penelitian ini, diharapkan mendapatkan hasil klasifikasi menggunakan kedua metode tersebut dengan performansi yang lebih baik dan diketahui gejala klinis apa yang paling berpengaruh atau dominan terhadap penentuan hasil klasifikasi wanita terhadap status pemeriksaan kanker payudara.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dijelaskan sebelumnya, maka rumusan masalah dari penelitian ini adalah bagaimana deskripsi wanita dengan status pemeriksaan positif kanker payudara dan wanita dengan status pemeriksaan negatif kanker payudara berdasarkan hasil tes deteksi kanker payudara yang pernah dilakukan sebelumnya, serta mendapatkan model klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis menggunakan metode *Classification Tree* dan *Bootstrap Aggregating*, yang mana dari metode tersebut akan dilakukan perbandingan evaluasi ketepatan klasifikasi dengan meninjau akurasi, *sensitifity*, *specificity*, dan nilai AUC.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dari penelitian yang ingin dicapai adalah sebagai berikut:

1. Mendeskripsikan wanita dengan status pemeriksaan positif kanker payudara dan wanita dengan status pemeriksaan negatif kanker payudara berdasarkan hasil tes deteksi kanker payudara.

2. Mengklasifikasikan hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara menggunakan metode *Classification Tree* dan *Bootstrap Aggregating*.
3. Melakukan perbandingan hasil evaluasi ketepatan klasifikasi menggunakan metode *Classification Tree* dan *Bootstrap Aggregating* dengan meninjau *accuracy*, *sensitivity*, *specificity*, dan nilai AUC.

1.4 Manfaat Penelitian

Manfaat dari penelitian yang ingin diperoleh, diantaranya adalah sebagai berikut:

1. Manfaat bagi pihak RSUD Nur Hidayah Bantul
Dapat menjadikan bahan pertimbangan maupun rekomendasi dalam mendiagnosa seorang penderita dan bukan penderita kanker payudara berdasarkan hasil tes deteksi kanker payudara terhadap gejala klinis dan faktor predisposisi yang diduga mempengaruhi terjadinya kanker payudara.
2. Manfaat bagi masyarakat umum
Bagi masyarakat umum dapat memberikan informasi dan menambah pengetahuan mengenai gejala klinis apa saja yang sering dialami oleh kelompok atau individu yang memiliki status pemeriksaan positif dan negatif kanker payudara berdasarkan hasil tes deteksi yang pernah dilakukan sebelumnya yang berpengaruh secara signifikan terhadap terjadinya kanker payudara.
3. Manfaat bagi penulis
Penelitian ini diharapkan dapat mengembangkan wawasan keilmuan statistika tentang klasifikasi secara umum terutama dalam mengklasifikasikan data menggunakan metode *Classification Tree* dan *Bootstrap Aggregating*.

1.5 Batasan Masalah

Batasan masalah yang digunakan pada penelitian kali ini adalah responden yang diteliti dibatasi pada wanita yang sudah atau pernah dan belum menikah.

BAB II

TINJAUAN PUSTAKA

2.1 Statistika Deskriptif

Statistika deskriptif merupakan salah satu metode analisis statistika yang berhubungan dengan pengumpulan dan penyajian data yang disajikan dalam bentuk tabel, grafik, *boxplot*, ukuran pemusatan data maupun ukuran penyebaran data sehingga memberikan informasi yang berguna. Statistika deskriptif ini digunakan untuk melihat gambaran umum dari suatu data sebelum dilakukan analisis. Ukuran pemusatan data dapat berupa *mean* dan *median*, sedangkan ukuran penyebaran data dapat berupa *varians* dan *standard deviation* (Walpole, 1995).

a. Bar Chart

Bar Chart atau diagram batang merupakan diagram berbentuk batang-batang tegak ataupun mendatar dan sama lebar dengan batang-batang terpisah. Diagram batang dapat digunakan untuk menggambarkan perkembangan nilai-nilai suatu objek penelitian untuk menyajikan data yang bersifat kategori atau distribusi.

b. Box Plot

Box Plot merupakan suatu cara dalam statistik deskriptif yang digunakan untuk menampilkan gambar secara grafik dari data-data numerik.

c. Pie Chart

Pie Chart atau diagram lingkaran merupakan lingkaran yang berbentuk tiga dimensi dan setiap juring menunjukkan persentase dari masing-masing kelompok data. Diagram ini berbentuk lingkaran dan dibagi menjadi beberapa bagian yang sesuai dengan banyaknya kelas penyusunan.

2.2 Classification and Regression Tree (CART)

Classification and Regression Tree (CART) dikembangkan oleh Breiman, dkk (1993) merupakan metode statistika non parametrik berdasarkan kaidah pohon keputusan (*decision tree*) yang digunakan untuk menggambarkan hubungan antara variabel respon (*dependent*) dengan satu atau lebih variabel prediktor (*Indepen-*

dent), baik untuk variabel respon yang berskala kategorik maupun kontinu. Pada metode *Classification and Regression Tree* (CART) adapun model pohon yang dihasilkan bergantung pada skala variabel respon, apabila variabel respon data berskala kontinu sehingga model pohon yang dihasilkan adalah *regression trees* (pohon regresi). Sedangkan apabila variabel respon mempunyai skala kategorik sehingga pohon yang dihasilkan adalah *classification trees* (pohon klasifikasi). Tujuan utama dari pembentukan pohon klasifikasi menggunakan metode *Classification and Regression Tree* (CART) yaitu untuk mendapatkan hasil pengklasifikasian yang akurat dan menentukan prediksi struktur data (Breiman dkk, 1993).

Berikut ini merupakan kelebihan dari metode *Classification and Regression Tree* (CART).

1. Metode *Classification and Regression Tree* (CART) adalah metode analisis statistika yang bersifat non parametrik sehingga tidak memerlukan asumsi-asumsi yang mengikat seperti asumsi distribusi normal untuk variabel prediktor (*independent*).
2. Metode ini tidak hanya memberikan klasifikasi, namun juga memberikan estimasi probabilitas kesalahan pengklasifikasian.
3. Mampu mengklasifikasi interaksi antara variabel prediktor yang berpengaruh secara lokal akibat diterapkannya pengambilan keputusan secara bertahap dalam himpunan-himpunan bagian data pengukuran yang kompleks.
4. Struktur data dapat dilihat secara visual sehingga memudahkan eksplorasi dan pengambilan keputusan berdasarkan model yang diperoleh.
5. Hasil klasifikasi akhir berbentuk sederhana dan dapat digunakan untuk mengklasifikasikan data baru secara efisien.
6. Kemudahan dalam interpretasi hasil.

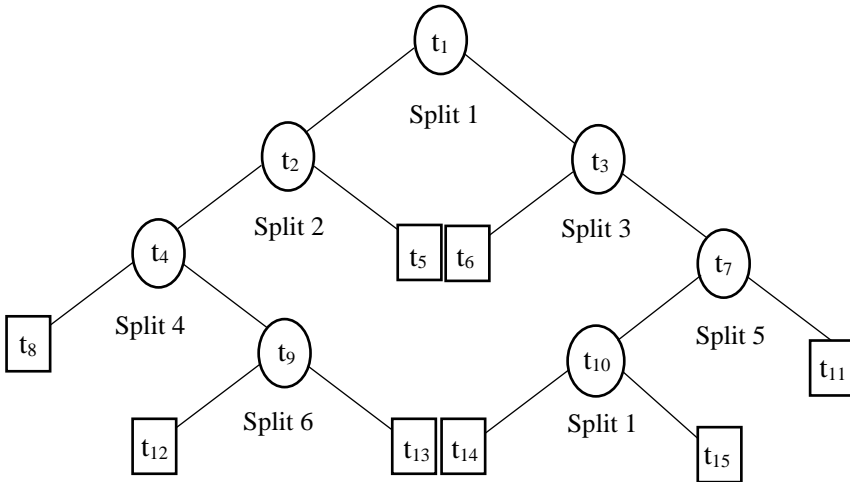
Sedangkan untuk kekurangan dari metode *Classification and Regression Tree* (CART) adalah CART mungkin tidak stabil dalam *decision trees* (pohon keputusan) karena CART sangat sensitif dengan data baru. CART sangat bergantung dengan jumlah

sampel. Jika sampel data *learning* dan *testing* berubah maka pohon keputusan yang dihasilkan juga ikut berubah.

Menurut Breiman dkk, (1993) analisis pohon klasifikasi merupakan salah satu metode klasifikasi *Binary Recursive Partitioning* (Penyekatan Rekursif Secara Biner). Istilah “*Binary*” diartikan sebagai pemilahan pada sekelompok data yang terkumpul di suatu simpul (*node*) menjadi dua kelompok yang disebut simpul anak (*child node*). Kemudian istilah “*Recursive*” diartikan sebagai pemilahan tersebut dapat terjadi berulang-ulang, lalu setiap simpul anak dapat memilah untuk membentuk simpul anak yang lain, dan seterusnya. Apabila dapat dilakukan pemilahan pada simpul anak sehingga disebut simpul dalam (*internal node*), namun jika tidak ada pemilahan setelahnya maka disebut simpul akhir (*terminal node*). Kemudian untuk istilah “*partitioning*” diartikan sebagai proses klasifikasi dapat dilakukan dengan cara memilih kumpulan data menjadi beberapa bagian atau partisi (Lewis, 2000).

Ilustrasi pohon klasifikasi ditunjukkan pada Gambar 2.1 di mana terdapat simpul awal sebagai variabel terpenting dalam menduga kelas suatu pengamatan (*parent node*) dengan simbol t_1 , kemudian simpul dalam (*internal node*) disimbolkan dengan t_2, t_3, t_4, t_7, t_9 , dan t_{10} serta simpul akhir (*terminal node*) disimbolkan dengan $t_5, t_6, t_8, t_{11}, t_{12}, t_{13}, t_{14}$, dan t_{15} di mana tidak ada lagi pemilahan setelahnya. Setiap simpul berada pada kedalaman (*depth*) tertentu di mana t_1 berada pada kedalaman 1, t_2 dan t_3 berada pada kedalaman 2 dan begitu seterusnya hingga simpul t_{12}, t_{13}, t_{14} , dan t_{15} berada pada kedalaman 5.

Terdapat tiga tahapan dalam melakukan klasifikasi dengan menggunakan metode *Classification and Regression Tree* (CART). Pertama, membentuk pohon klasifikasi dengan prosedur pembentukan menggunakan pemilahan simpul secara berulang (*recursive*). Selanjutnya dilakukan pemangkasan pohon klasifikasi (*pruning*) yang menghasilkan rangkaian pohon klasifikasi yang lebih sederhana. Tahapan terakhir, penentuan pohon klasifikasi optimal, dimana pohon klasifikasi tersebut dapat mempresentasikan informasi dari data namun tidak berlebihan (*overfitting*).



Gambar 2.1 Diagram Pohon Klasifikasi

2.3.1 Pembentukan Pohon Klasifikasi Maksimal

Proses pembentukan pohon klasifikasi dapat diawali dengan penentuan variabel dan nilai dari variabel tersebut yang layak dijadikan pemilah bagi setiap simpul. Proses pembentukan pohon klasifikasi harus memperhatikan tiga hal penting yaitu pemilihan pemilah, penentuan simpul terminal, dan penandaan label kelas (Breiman dkk, 1993). Berikut ini merupakan penjelasan dari ketiga proses dalam pembentukan pohon klasifikasi.

a. Pemilahan Pemilah

Pada tahapan pemilahan pemilah dilakukan pemilahan pada sampel data *training* berdasarkan aturan pemilahan dan kriteria *goodness of split*, dimana sampel data *training* yang digunakan masih bersifat heterogen. Pemilahan pemilah tergantung pada jenis pohon atau pada jenis variabel respon. Hasil dari proses pemilahan harus lebih homogen dibandingkan dengan simpul induknya. Tingkat heterogenan simpul tersebut dapat diukur menggunakan nilai *impurity* atau $i(t)$. Aturan pemilahan simpul induk menjadi dua simpul anak berdasarkan pada nilai yang berasal dari satu variabel prediktor. Setiap pemilahan hanya tergantung pada satu variabel

prediktor saja. Apabila x_j berskala kontinu, maka pemilahan yang diperbolehkan adalah $x_j \leq c_i$ dan $x_j > c_i$ dengan $i = 1, 2, 3, \dots, n - 1$ dan c_i adalah nilai tengah dari dua nilai amatan sampel berurutan yang berbeda dari variabel x_j . Jika suatu ruang sampel berukuran n dan terdapat sejumlah n nilai amatan sampel yang berbeda pada variabel x_j , maka terdapat $n-1$ kemungkinan pemilahan yang berbeda. Namun apabila 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. Bila kategorik berskala ordinal bertaraf L , maka akan diperoleh sebanyak $L-1$ pemilahan yang mungkin.

Tujuan dilakukan pemilahan adalah untuk mengurangi keheterogenan pada simpul utama dan mendapatkan simpul anak-simpul anak dengan tingkat homogenitas yang tinggi dan dapat mencirikan karakteristik kelas-kelas variabel respon. Untuk mengetahui kondisi tersebut, maka ada implementasi fungsi keheterogenan simpul. Hal ini dapat dilakukan dengan mendefinisikan fungsi keheterogenan simpul. Fungsi heterogenitas yang sering digunakan adalah indeks gini. Penggunaan indeks gini dalam pemilahan pemilah memiliki kelebihan, yaitu proses perhitungannya sederhana dan relatif cepat, serta mudah untuk diterapkan dalam berbagai kasus (Brieman dkk, 1993). Fungsi indeks gini dirumuskan dalam persamaan 2.1 berikut.

$$i(t) = \sum_{i,j=1} p(i|t)p(j|t), i \neq j \quad (2.1)$$

dengan,

$i(t)$: Indeks gini (fungsi heterogenitas) pada simpul t

$p(i|t)$: Proporsi kelas i pada simpul t

$p(j|t)$: Proporsi kelas j pada simpul t

Langkah berikutnya yaitu menentukan pemilah terbaik dari setiap variabel prediktor, pemilah terbaik adalah pemilah yang memaksimalkan ukuran pemilahan antara dua simpul anak tersebut. Setiap pemilahan akan dilakukan pada setiap simpul sampai diper-

oleh simpul akhir. Kemudian menentukan kriteria *goodness of split* yang merupakan suatu evaluasi pemilahan yang dilakukan oleh pemilah s pada suatu simpul t . Rumus untuk mencari nilai *goodness of split* dituliskan pada persamaan 2.2.

$$\phi(s, t) = \Delta_1(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.2)$$

dengan,

$\phi(s, t)$: Nilai *goodness of split*

$i(t)$: Fungsi keheterogenan pada simpul t

p_L : Proporsi pengamatan simpul kiri

p_R : Proporsi pengamatan simpul kanan

$i(t_L)$: Fungsi keheterogenan pada simpul kiri

$i(t_R)$: Fungsi keheterogenan pada simpul kanan

Pemilahan yang menghasilkan nilai *goodness of split* tertinggi merupakan pemilah terbaik karena dapat menghasilkan heterogenitas lebih tinggi. Pembentukan struktur pohon dilakukan dengan pencarian sebuah pemilah yang mungkin pada setiap simpul yang dimulai dari simpul utama t_1 . Suatu pemilah s^* yang mampu menghasilkan nilai penurunan keheterogenan tertinggi diantara semua pemilah s yang mungkin adalah yang terpilih sebagai pemilah simpul utama t_1 menjadi simpul t_2 dan t_3 . Konsep tersebut secara sistematis dapat dituliskan dalam persamaan 2.3 berikut ini.

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1) \quad (2.3)$$

Kemudian untuk mengembangkan pembentukan struktur pohon maka dengan cara dan konsep yang sama dilakukan pemilahan pada simpul t_2 dan t_3 masing-masing secara terpisah seterusnya sampai ditemukan simpul terminal. Kemudian setiap variabel prediktor akan menghasilkan skor untuk menunjukkan seberapa besar variabel tersebut memberikan kontribusi dalam proses pembentukan pohon. Berikut ini merupakan persamaan untuk menentukan besarnya skor setiap variabel.

$$skor = \sum_{i=1}^n \phi(s, t_i) \quad (2.4)$$

Dimana $\phi(s, t_i)$ merupakan nilai *goodness of split* pada setiap simpul. Nilai skor diperoleh dengan menjumlahkan nilai *goodness of split (improvement)* dari masing-masing variabel yang berperan sebagai *surrogate* untuk setiap simpul (CART Reference Guide, 2000).

b. Penentuan Simpul Terminal

Suatu simpul t dapat dijadikan sebagai simpul terminal apabila memenuhi salah satu kriteria sebagai berikut; hanya ada satu pengamatan ($n=1$) dalam setiap simpul anak atau adanya batasan minimum n pengamatan di setiap simpul anak yang dihasilkan, adanya batasan jumlah level atau tingkat kedalaman pohon maksimal yang ditetapkan dan semua pengamatan dalam setiap simpul anak mempunyai distribusi identik terhadap variabel prediktor sehingga tidak memungkinkan untuk dilakukan pemilahan kembali (Breiman dkk, 1993).

c. Penandaan Label Kelas

Setiap simpul terminal perlu diberi label kelas untuk mengetahui karakteristik dari klasifikasi pengamatan untuk setiap kelas variabel respon yang terbentuk. Penandaan label kelas pada simpul terminal berdasarkan aturan jumlah terbanyak seperti yang ditunjukkan dalam persamaan 2.5.

$$p(j_0 | t) = \max p(j | t) = \max \frac{N_j(t)}{N(t)} \quad (2.5)$$

Label kelas untuk simpul terminal t adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t yang paling kecil, yaitu sebesar $r(t) = 1 - \max p(j | t)$ (Breiman dkk, 1993).

2.3.2 Pemangkasan Pohon Klasifikasi Maksimal (*Pruning*)

Pemangkasan pohon klasifikasi atau yang biasa disebut dengan *pruning* perlu dilakukan karena semakin banyak pemilahan yang dilakukan mengakibatkan makin kecilnya tingkat kesalahan prediksi atau dengan kata lain nilai prediksi melebihi nilai yang sebenarnya (*overfitting*). Pemangkasan pohon dilakukan de-

ngan menentukan *cost complexity minimum* (Breiman dkk, 1993). Ukuran *cost complexity* terdapat pada persamaan 2.6.

$$R_\alpha(T) = R(T) + a \left| \tilde{T} \right| \quad (2.6)$$

dengan,

- $R_\alpha(T)$: Ukuran kompleksitas (*cost complexity measure*) suatu pohon T pada kompleksitas α
- $R(T)$: Ukuran kesalahan (*cost misclassification*) klasifikasi pohon T atau penduga pengganti (*resubstitution estimate*) pohon
- a : Parameter kompleksitas (*complexity parameter*) bagi penambah satu simpul terminal pada pohon T
- $\left| \tilde{T} \right|$: Banyaknya simpul terminal pada pohon T

Cost complexity pruning 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 $R_\alpha(T)$ yaitu $R_\alpha(T(\alpha)) = \min_{T < T_{\max}} R_\alpha(T)$.

Secara umum tahapan pada proses pemangkasan pohon (*pruning*) adalah sebagai berikut :

1. Melakukan pemangkasan pohon klasifikasi maksimal T_{\max} yang kemudian diambil t_L dan t_R di mana merupakan simpul anak kiri dan kanan dari T_{\max} yang diperoleh dari pemilahan simpul induk t .
2. Apabila didapatkan dua simpul anak dan simpul induknya memenuhi persamaan $R(t) = R(t_L) + R(t_R)$, maka simpul anak t_L dan t_R akan dipangkas. Hasilnya adalah pohon T_l yang memenuhi kriteria $R(T_l) = R(T_{\max})$.
3. Melakukan pengulangan mulai dari langkah pertama hingga terakhir sampai tidak ada lagi pemangkasan yang mungkin.

2.3.3 Penentuan Pohon Klasifikasi Optimal

Pohon klasifikasi yang berukuran sangat besar memberikan nilai penduga pengganti yang paling kecil, sehingga pohon tersebut cenderung dipilih untuk menduga nilai respon. Tetapi ukuran pohon yang besar bersifat *overfitting* sehingga menyebabkan nilai kompleksitas yang tinggi. Hal ini dikarenakan struktur datanya yang digambarkan cenderung kompleks sehingga diperlukan pemilihan pohon yang optimum dengan ukuran yang lebih sederhana dan memberikan nilai penduga pengganti yang kecil. Data sampel akan digunakan untuk mendapatkan nilai pengganti paling kecil dari pohon klasifikasi yang dipilih. Terdapat dua jenis penduga pengganti yaitu penduga validasi silang lipat V (*cross validation V-fold estimate*) dan uji sampel penduga (*test sample estimate*). Penduga pengganti yang sering digunakan jika ukuran pengamatan besar adalah *test sample estimate*. Prosedur ini diterapkan dengan membagi sampel L menjadi dua bagian, yaitu L_1 (*training sampel*) dan L_2 (*testing sampel*). Pengamatan dalam L_1 digunakan untuk membentuk pohon T , sedangkan pengamatan dalam L_2 digunakan untuk menduga $R(T)$. N_1 merupakan jumlah pengamatan L_1 dan N_2 jumlah pengamatan L_2 . $X(\cdot)$ bernilai 0 jika pernyataan dalam tanda kurung salah dan bernilai 1 jika pernyataan dalam tanda kurung besar. Penduga sampel uji dapat ditunjukkan dalam persamaan 2.7.

$$R^{ts}(T_t) = \frac{1}{N_2} \sum_{(x_n, j_n) \in L_2} X(d(x_n \neq j_n)) \quad (2.7)$$

dengan,

$R^{ts}(T_t)$: Total proporsi kesalahan *test sample estimate*

N_2 : Jumlah pengamatan dari data *training* L_2

Dalam hal ini ingin menduga proporsi kesalahan yang dihasilkan dari proses pembentukan pohon klasifikasi, sehingga pohon klasifikasi optimal yang dipilih adalah pohon T_t yang memiliki nilai penduga sampel uji minimum atau $R^{ts}(T^*) = \min_t R^{ts}(T_t)$. Sedangkan penduga pengganti yang sering digunakan apabila pengamatan yang ada tidak cukup besar adalah *Cross Validation V-Fold Estimate*. Pengamatan dalam L dibagi secara acak menjadi

V bagian yang saling lepas dengan ukuran kurang lebih sama besar disetiap kelasnya. *Training sampel* ke- v dengan $v=1,2,\dots,V$ digunakan untuk membentuk pohon $T_k^{(v)}$. Misalkan $d^{(v)}(x)$ adalah hasil pengklasifikasian maka penduga sampel uji untuk $R^{ts}(T_k^{(v)})$ adalah sebagai berikut.

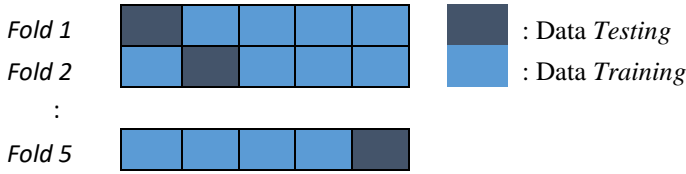
$$R^{ts}(T_k) = \frac{1}{N_2} \sum_{(x_n, j_n) \in L_2} X(d(x_n \neq j_n)) \quad (2.8)$$

Dengan menggunakan pengamatan induk L untuk membentuk deretan pohon (T_k) maka penduga validasi silang lipat V untuk $T_k^{(v)}$ adalah sebagai berikut. Pohon klasifikasi optimum dipilih T_k^* dengan $R^{cv}(T_k^*) = \min_k R^{cv}(T_k)$.

$$R^{cv}(T_k) = \frac{1}{V} \sum_{v=1}^V R^{ts}(T_k^{(v)}) \quad (2.9)$$

2.3 *K-Fold Cross Validation*

K-Fold Cross Validation merupakan metode yang menggunakan sebuah kombinasi dari beberapa tes untuk mendapatkan hasil estimasi yang stabil dari kesalahan model. Metode ini membagi dataset secara random menjadi sejumlah k bagian dengan ukuran yang sama untuk setiap bagian. Satu bagian membentuk set validasi *testing*, bagian lainnya membentuk set validasi *training*. Proses ini diulangi untuk setiap bagian data sebanyak k kali. Dan masing-masing bagian menggunakan data ke- k sebagai data *testing* dan sisanya dijadikan sebagai data *training*. Nilai k yang paling umum digunakan adalah 5 dan 10, karena merupakan nilai yang paling memadai untuk mendapatkan perkiraan kesalahan terbaik (Mitchell, 1997). Berikut ini merupakan ilustrasi dalam pengambilan data *testing* dan *training* menggunakan metode *k-fold cross validation*.



Gambar 2.2 Ilustrasi *K-Fold Cross Validation*

2.4 Bootstrap

Bootstrap adalah metode berbasis *resampling* data sampel yang digunakan untuk mengestimasi suatu distribusi populasi yang tidak diketahui dengan menggunakan distribusi empiris yang diperoleh dari proses pengambilan sampel ulang dari sampel asli dengan ukuran sama dengan ukuran sampel asli dan dilakukan dengan syarat pengembalian (*with replacement*) (Efron dan Tibshirani, 1993). Metode *bootstrap* merupakan salah satu pilihan yang dapat digunakan untuk menghasilkan model yang lebih baik. Kedudukan sampel asli dalam metode *bootstrap* dipandang sebagai populasi. Metode *bootstrap* ini biasa disebut dengan *resampling bootstrap*. (Berrar, 2018). Metode ini dapat digunakan untuk ukuran data sampel yang relatif kecil ($n < 30$).

Berikut ini merupakan algoritma *bootstrap* sebagai berikut.

1. Mengambil sampel berukuran n , yaitu $S : x_1, x_2, x_3, \dots, x_n$.
2. Mengambil sampel kembali dari S dengan pengembalian berukuran n dan didapat nilai statistik $\hat{\theta}_i$ untuk sampel S_i .
3. Melakukan pengambilan sampel ulang pada langkah 2 sebanyak B kali.
4. Menentukan nilai statistik dengan *bootstrap*.

$$\hat{\theta}_b = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i \quad (2.10)$$

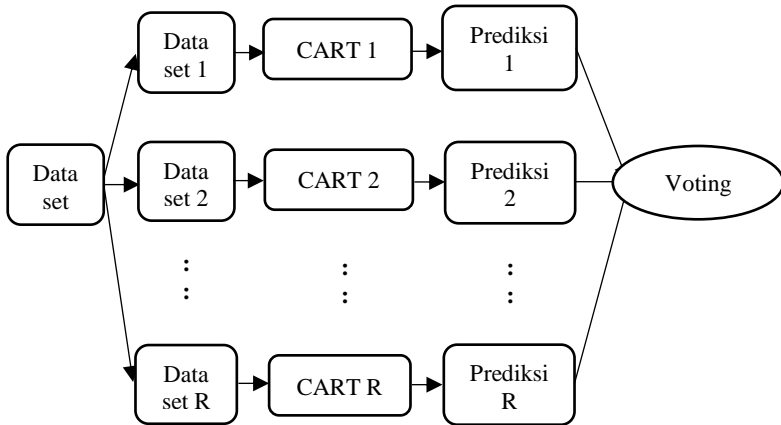
$$\hat{\sigma}_{\theta_b}^2 = \frac{\sum_{i=1}^B \left(\hat{\theta}_i - \hat{\theta}_b \right)^2}{(B-1)} \quad (2.11)$$

2.4.1 *Bootstrap Aggregating*

Bootstrap aggregating atau yang biasa disebut *bagging* merupakan metode pengembangan dari metode *classification and regression tree* (CART), yaitu dengan penerapan metode *bootstrap*. Metode *bagging* pertama kali digunakan oleh Breiman (1996) sebagai alat dalam pembentukan *classifier* yang lebih stabil dengan memodifikasi klasifikasi dengan menggabungkan klasifikasi pada data *training* yang dihasilkan secara acak. Pada beberapa kasus *bagging* pada *dataset* real mampu meningkatkan akurasi dari algoritma *machine learning* yang digunakan. *Bagging* juga bertujuan untuk mengurangi variansi dan menghindari terjadinya *overfitting*. Meskipun *bagging* biasa diterapkan pada metode *decision tree*, metode ini juga dapat digunakan untuk semua jenis metode (Breiman, 1996). Berikut ini merupakan algoritma dari *bootstrap aggregating* adalah sebagai berikut.

1. Sebuah *dataset* \mathcal{L} yang terdiri dari $\{(y_i, x_i), i=1,2,\dots,n\}$. Kemudian dilakukan replikasi *bootstrap* pada data tersebut sehingga didapatkan $\mathcal{L}_i^* = (y_i^*, x_i^*), i = 1, 2, \dots, n$.
2. Replikasi *bootstrap* dilakukan sebanyak B kali, sehingga diperoleh $\{\mathcal{L}^{(B)}\}$ dari \mathcal{L} . $\{\mathcal{L}^{(B)}\}$ adalah *resampling* dengan pengembalian.

Penentuan banyaknya replikasi B sangat bervariasi, karena besar kecilnya B dapat memberikan hasil yang berbeda pada setiap tahapan analisis. Sutton (2005) merekomendasikan replikasi sebanyak 25 atau lima puluh kali, namun Hastie, Tibshirani, dan Friedman (2001) menyatakan bahwa peningkatan akurasi akan terjadi jika banyaknya replikasi ditingkatkan dari lima puluh ke seratus kali, sedangkan jika banyaknya ditingkatkan hingga lebih dari seratus kali akan menghasilkan akurasi yang tidak lebih besar dari akurasi replikasi seratus kali. Sementara itu, Efron dan Tibshirani (1993) merekomendasikan nilai B yang kecil, misalnya 25 kali. Berikut ini merupakan ilustrasi dari metode Bagging CART.



Gambar 2.3 Ilustrasi Bagging CART

Setelah dilakukan proses *bootstrap* sebanyak B kali pembentukan pohon klasifikasi dan prediksi data *testing* pada masing-masing pohon klasifikasi, langkah selanjutnya adalah dilakukan prediksi gabungan (*aggregating*) berdasarkan B dugaan pada data *testing* dengan aturan *majority vote* (suara terbanyak). Sebagai contoh, misalkan dilakukan proses *bootstrap* sebanyak 25 kali. Sehingga diperoleh 25 model pohon klasifikasi yang akan dilakukan prediksi. Berdasarkan hasil prediksi 25 pohon klasifikasi, data *testing* pertama yang diprediksi masuk ke kelas positif kanker payudara sebanyak 21 sedangkan yang diprediksi masuk kelas negatif kanker payudara sebanyak 4. Hasil akhir pendugaan berdasarkan suara terbanyak, sehingga prediksi akhir data *testing* pertama masuk ke dalam kelas positif kanker payudara. Hasil dari prediksi gabungan yang telah diperoleh selanjutnya digunakan untuk menguji ketepatan klasifikasi pada penerapan metode Bagging CART.

2.5 Evaluasi Ketepatan Klasifikasi

Evaluasi ketepatan klasifikasi yang digunakan pada penelitian ini dapat ditinjau dari *accuracy*, *sensitifity*, dan *specificity* (Johnson dan Winchern, 2007). *Accuracy*, *sensitivity*, dan *specificity* dapat ditentukan menggunakan nilai yang terdapat dalam *confusion matrix*. *Confusion matrix* adalah salah satu metode yang

dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya (Rosely, 2015). Tiap kolom pada *confusion matrix* adalah contoh kelas prediksi, sedangkan tiap baris mewakili kejadian di kelas yang sebenarnya (Gorunescu, 2011). *Confusion matrix* ditunjukkan pada Tabel 2.1.

Tabel 2.1 *Confusion Matrix*

Aktual	Prediksi	
	Positif = kelas 1	Negatif = kelas 0
Positif = kelas 1	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
Negatif = kelas 0	FP (<i>False Positive</i>)	TN (<i>True Negative</i>)

dengan,

TP : Jumlah kasus positif yang diklasifikasikan sebagai positif

FP : Jumlah kasus negatif yang diklasifikasikan sebagai positif

TN : Jumlah kasus positif yang diklasifikasikan sebagai negatif

FN : Jumlah kasus negatif yang diklasifikasikan sebagai negatif

Kinerja ketepatan klasifikasi dapat dievaluasi dengan memperhatikan beberapa ukuran berikut ini.

1. *Accuracy* (%)

Accuracy merupakan ukuran ketepatan klasifikasi yang diperoleh. Persamaan yang dapat digunakan untuk menghitung akurasi adalah sebagai berikut.

$$Accuracy (\%) = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.12)$$

2. *Sensitivity* (%)

Sensitivity merupakan ukuran ketepatan dari suatu kejadian yang diinginkan. Persamaan yang dapat digunakan untuk menghitung sensitivitas adalah sebagai berikut.

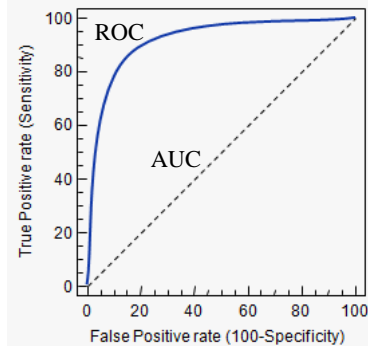
$$Sensitivity (\%) = \frac{TP}{TP + FN} \quad (2.13)$$

3. *Specificity* (%)

Specificity merupakan ukuran yang menyatakan persentase kejadian-kejadian yang tidak diinginkan. Persamaan yang dapat digunakan untuk menghitung spesifisitas adalah sebagai berikut.

$$\text{Specificity (\%)} = \frac{TN}{FP + TN} \quad (2.14)$$

Selain ukuran kebaikan yang telah disebutkan di atas sebagai evaluasi ketepatan klasifikasi dapat digunakan juga nilai AUC.



Gambar 2.4 Hubungan Kurva ROC Dengan AUC

AUC (*Area Under ROC Curve*) merupakan satu dari beberapa kriteria yang dapat digunakan sebagai evaluasi kualitas dari hasil klasifikasi (Huang, J. dan Ling, C. X., 2005). Berikut ini persamaan yang dapat digunakan untuk menghitung nilai AUC.

$$AUC = \frac{1}{2}(\text{sensitiftiy} + \text{specificity}) \quad (2.15)$$

Di dalam klasifikasi data, nilai AUC dapat dibagi menjadi beberapa kelompok sebagai berikut (Gorunescu, 2011).

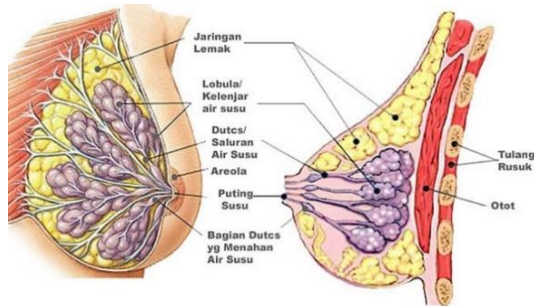
Tabel 2.2 Kategori Pengklasifikasian Model Berdasarkan Nilai AUC

Nilai AUC	Kategori Klasifikasi
0,90 – 1,00	<i>Excellent Classification</i>
0,80 – 0,90	<i>Good Classification</i>
0,70 – 0,80	<i>Fair Classification</i>
0,60 – 0,70	<i>Poor Classification</i>
0,50 – 0,60	<i>Failure</i>

2.6 Kanker Payudara

Kanker payudara (*carcinoma mammae*) adalah tumor ganas yang terbentuk dari sel-sel payudara yang tumbuh dan berkembang

biak tanpa terkontrol sehingga dapat menyebar di antara jaringan-jaringan atau organ-organ yang berada di sekitar payudara atau bahkan dapat menyebar ke bagian tubuh lainnya (KemenKes RI, 2016). Payudara wanita terdiri dari lobulus (kelenjar susu), duktus (saluran susu), lemak dan jaringan ikat, pembuluh darah dan *limfe*. Sebagian besar kanker payudara bermula pada sel-sel yang melapisi duktus (kanker duktal), beberapa bermula di lobulus (kanker lobulat), serta sebagian kecil bermula di jaringan lain (Ellis dkk, 2003).



Gambar 2.5 Anatomi Payudara Manusia

Penyebab kanker payudara sampai saat ini belum dapat diketahui secara pasti, diduga penyebab kanker payudara adalah multifaktor. Beberapa studi telah membahas mengenai faktor-faktor resiko yang diduga memicu terjadinya penyakit kanker payudara. Menurut Almutlaq, dkk (2017) resiko terjadinya kanker payudara di usia tua lebih tinggi dibandingkan dengan usia muda, sehingga semakin bertambahnya usia maka akan semakin tinggi resiko seseorang terkena kanker payudara. Sedangkan menurut Hinyard, dkk (2017) berdasarkan penelitian yang pernah dilakukan status pernikahan sangat terkait dengan peningkatan kesehatan dan umur panjang. Menikah telah terbukti berhubungan positif dengan kelangsungan hidup pada pasien dengan berbagai jenis keganasan. Namun sedikit yang diketahui tentang hubungan antara status perkawinan dan kanker payudara pada wanita muda. Wanita yang belum menikah 1,18 kali lebih mungkin didiagnosis kanker payudara daripada wanita yang sudah menikah.

2.6.1 Gejala Klinis

Terdapat beberapa gejala-gejala klinis kanker payudara yang dapat diperhatikan di mana gejala tersebut timbul melalui perubahan yang dapat dilihat sebagai kelainan yang terjadi dan perlu mendapat perhatian. Gejala-gejala tersebut diantaranya berawal dari perubahan yang terjadi pada bentuk dan ukuran payudara, teraba benjolan di payudara dan bawah ketiak, terjadi penebalan kulit di sekitar payudara, keluarnya cairan putih dari puting susu, kulit di sekitar payudara memiliki motif seperti kulit jeruk, penarikan puting susu ke dalam, luka pada payudara yang tidak kunjung sembuh dan pembuluh darah payudara yang terlihat lebih jelas dan lebar (*International Agency for Research on Cancer*, 2016).

2.6.2 Faktor Predisposisi

Hasil studi lain yang menggambarkan pengaruh kualitas tidur pada penderita kanker payudara diantaranya rasa terbatuk akibat radioterapi yang terjadi pada 85% dari 73 pasien yang diteliti menyebabkan mereka harus terbangun sebanyak 15 kali (Dhruva dkk, 2012). Kemudian 67% dari 252 mengalami kualitas tidur buruk akibat kelelahan, nyeri dan stres psikologis (Romito dkk, 2014). Secara umum, komponen-komponen yang berkontribusi dalam penilaian terhadap buruknya kualitas tidur seseorang diantaranya latensi tidur, durasi tidur, efisiensi kebiasaan tidur dan disfungsi siang hari. Pada komponen latensi tidur atau waktu yang dibutuhkan untuk memuai tertidur rata-rata membutuhkan waktu sekitar 60 menit. Hal tersebut dapat dikategorikan buruk karena durasi normal yang dibutuhkan untuk mulai tertidur pada dewasa sekitar 2 menit (Budhrani dkk, 2014). Begitu juga dengan penelitian yang dilakukan oleh Araujo, dkk (2014) bahwa penderita kanker payudara membutuhkan waktu untuk mulai tertidur sekitar 28 menit. Kemudian menurut Liu, dkk (2012) bahwa penderita kanker payudara menghabiskan waktu 2 jam untuk tertidur kembali ketika terbangun dari tidurnya di malam hari dan rata-rata total waktu tidur sekitar < 7 jam. Selain itu menurut Lowery, dkk (2014) bahwa gejala insomnia yang paling umum terjadi pada penderita

kanker payudara adalah beberapa kali terbangun di malam hari dan susah untuk tidur kembali.

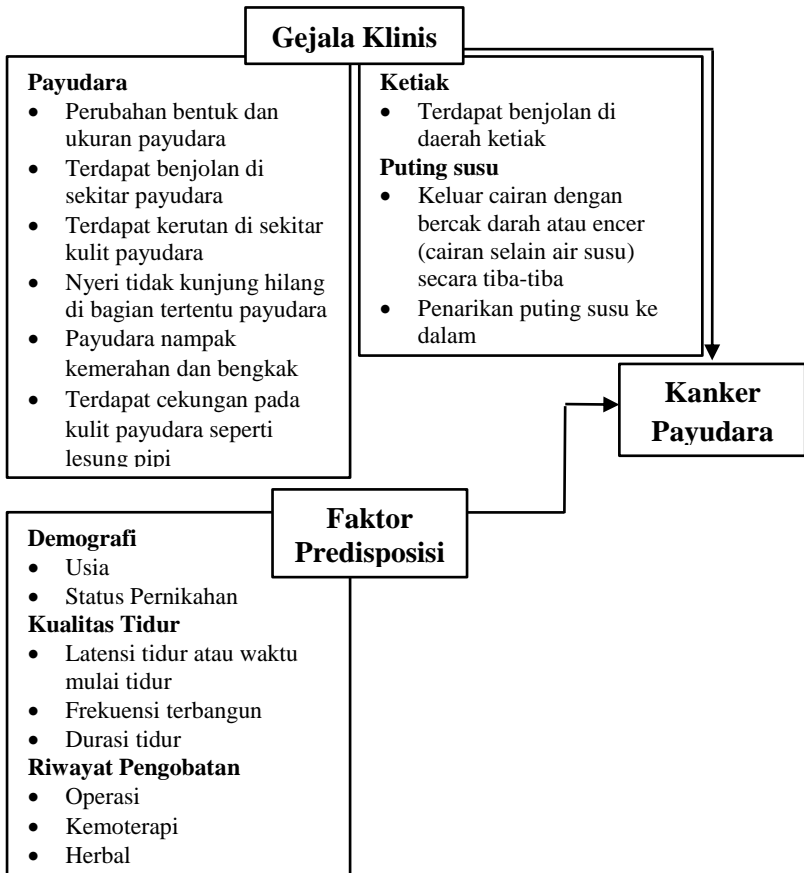
Adapun riwayat pengobatan diduga juga memiliki pengaruh yang kuat dengan penderita kanker. Pengobatan komplementer dan alternatif umumnya digunakan oleh penderita kanker payudara sebagai mengatasi mekanisme untuk memerangi penyakit itu. Alasan memberi pasien untuk menggunakan pengobatan komplementer dan alternatif yang terutama untuk membantu dalam penyembuhan kekuatan batin tubuh, untuk menyembuhkan kanker, dan untuk mengurangi stress (Shaharudin dkk, 2011). Pengobatan komplementer dan alternatif merupakan pengobatan yang mengacu pada praktik kesehatan yang tidak biasanya ditawarkan di fasilitas medis konvensional. Di Malaysia, pengobatan komplementer dan alternatif sangat populer di kalangan kanker dan pasien dengan tipe yang lain (Tasaki dkk, 2000). Salah satu bentuk pengobatan komplementer dan alternatif yang digunakan di seluruh dunia dalam berbagai bentuk adalah pengobatan tradisional. Menurut WHO, pengobatan tradisional didefinisikan sebagai sejumlah pengetahuan, keterampilan, dan praktik berdasarkan teori, keyakinan dan pengalaman adat untuk budaya yang berbeda, apakah dijelaskan atau tidak, digunakan dalam pemeliharaan kesehatan serta pencegahan, diagnosis, perbaikan atau perawatan fisik dan penyakit mental (Siti dkk, 2009).

BAB III

METODOLOGI PENELITIAN

3.1 Kerangka Konsep

Kerangka konsep dalam penelitian ini ditunjukkan dalam Gambar 3.1. Adapun dasaran dari pembuatan kerangka konsep ini merujuk pada gejala-gejala yang nampak yang diduga berpengaruh terhadap terjadinya kanker payudara.



Gambar 3.1 Kerangka Konsep Penelitian

3.2 Desain Penelitian

Desain penelitian yang digunakan pada Tugas Akhir ini merupakan penelitian berbasis *case-control* di mana dilakukan perbandingan dan identifikasi antara kelompok kasus (*case*) dan kelompok kontrol (*control*). Adapun yang dimaksud dengan penelitian berbasis *case-control* yaitu menentukan objek yang akan diteliti (*outcome*) terlebih dahulu kemudian mengidentifikasi penyebab (faktor resiko) yang mana dapat diketahui dari data histori pasien melalui wawancara dari responden penelitian. Pada penelitian ini responden dengan status pemeriksaan positif kanker payudara berdasarkan hasil tes deteksi kanker payudara adalah sebagai kelompok *case*. Jenis tes deteksi kanker payudara yang dilakukan oleh responden yang memiliki status pemeriksaan positif kanker payudara salah satunya adalah ultrasonografi (USG) dan Aspirasi Jarum Halus (AJH) yang dapat dilihat pada Lampiran 8. Sedangkan responden dengan status pemeriksaan negatif kanker payudara adalah sebagai kelompok *control*. Responden dengan status pemeriksaan negatif kanker payudara ini merupakan responden yang belum melakukan tes deteksi kanker payudara.

Adapun kriteria yang dibutuhkan pada subjek penelitian Tugas Akhir ini ada 2, yaitu kriteria inklusi dan eksklusi. Kriteria inklusi adalah kriteria atau ciri-ciri yang harus terpenuhi setiap masing-masing anggota populasi yang akan dijadikan sampel dalam pengambilan data. Sedangkan kriteria eksklusi adalah kriteria atau ciri-ciri anggota populasi yang tidak bisa dijadikan sebagai sampel penelitian dalam pengambilan data. Pada penelitian ini yang termasuk ke dalam kriteria inklusi yaitu responden dengan jenis kelamin perempuan, hasil dari tes deteksi kanker payudara adalah positif dan negatif, dan melakukan pengobatan atau terapi selain medis di rumah terapi tauhid RSUD Nur Hidayah Bantul. Kemudian yang termasuk ke dalam kriteria eksklusi yaitu responden yang sebelumnya pernah datang untuk melakukan pengobatan atau terapi selain medis di rumah terapi tauhid RSUD Nur Hidayah Bantul pada bulan yang sama di hari yang berbeda saat pengambilan data.

3.3 Sumber Data

Data yang digunakan dalam penelitian Tugas Akhir ini adalah data primer sebanyak 49 observasi dengan 17 variabel yang diperoleh secara langsung dari hasil survei melalui media komunikasi berupa kuisioner yang diberikan kepada kelompok responden dengan status pemeriksaan positif kanker payudara (*case*) dan kelompok responden dengan status pemeriksaan negatif kanker payudara (*control*) di rumah terapi tauhid RSUD Nur Hidayah Bantul pada bulan April 2019.

3.4 Variabel Penelitian

Penjelasan untuk masing-masing variabel yang akan diteliti adalah sebagai berikut.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan	Kategori
X ₁	Usia	0 = < 45 Tahun 1 = ≥ 45 Tahun
X ₂	Status Pernikahan	0 = Belum menikah 1 = Sudah menikah
X ₃	Perubahan bentuk dan ukuran payudara	0 = Tidak; 1 = Ya
X ₄	Benjolan di sekitar payudara	0 = Tidak; 1 = Ya
X ₅	Kerutan di sekitar kulit payudara	0 = Tidak; 1 = Ya
X ₆	Nyeri di bagian tertentu payudara	0 = Tidak; 1 = Ya
X ₇	Payudara nampak kemerahan dan bengkak	0 = Tidak; 1 = Ya
X ₈	Keluar cairan dengan bercak darah atau encer (cairan selain air susu)	0 = Tidak; 1 = Ya
X ₉	Penarikan puting susu ke dalam	0 = Tidak; 1 = Ya
X ₁₀	Benjolan di daerah ketiak	0 = Tidak; 1 = Ya
X ₁₁	Cekungan pada kulit payudara	0 = Tidak; 1 = Ya
X ₁₂	Durasi tidur	0 = < 5 jam 1 = ≥ 5-6 jam 2 = ≥ 6-7 jam 3 = ≥ 7 jam

Tabel 3.1 Variabel Penelitian (Lanjutan)

Variabel	Keterangan	Kategori
X ₁₃	Waktu mulai tertidur (latensi tidur)	0 = > 60 menit 1 = 31 - ≤ 60 menit 2 = 16 - ≤ 30 menit 3 = ≤ 15 menit
X ₁₄	Frekuensi terbangun	0 = > 5 kali 1 = 3 - 4 kali 2 = 1 - 2 kali 3 = Tidak pernah
X ₁₅	Riwayat pengobatan operasi	0 = Tidak; 1 = Ada
X ₁₆	Riwayat pengobatan kemoterapi	0 = Tidak; 1 = Ada
X ₁₇	Riwayat pengobatan herbal	0 = Tidak; 1 = Ada
Y	Status pemeriksaan kanker payudara	0 = Negatif Kanker Payudara 1 = Positif Kanker Payudara

Berikut ini merupakan definisi operasional dari variabel penelitian.

1. Usia
Usia seorang penderita dan bukan penderita kanker payudara. Risiko terjadinya kanker payudara akan meningkat seiring dengan bertambahnya usia.
2. Status pernikahan
Status pernikahan merupakan status seseorang apakah sudah menikah atau belum menikah. Seseorang yang belum menikah pada umumnya rentan terhadap kanker payudara.
3. Perubahan bentuk dan ukuran payudara
Perubahan bentuk dan ukuran payudara pada orang yang terkena kanker akan berbeda antara sisi kanan dengan sisi kiri dalam hal bentuk maupun ukuran yang tidak lagi normal.
4. Terdapat benjolan di sekitar payudara
Seseorang yang memungkinkan terkena kanker payudara memiliki ciri benjolan terasa keras, tidak mudah dibedakan, benjolan tidak bergerak, hanya terdapat satu benjolan, tidak

ada benjolan yang sama di payudara sebelah dan benjolan disertai dengan keluarnya cairan.

5. Terdapat kerutan di sekitar kulit payudara
Pada umumnya permukaan kasar dan spesifik di daerah tertentu pada kulit payudara. Biasanya kerutan berwarna sedikit hitam dan tidak terlalu besar lingkungannya.
6. Nyeri tidak kunjung hilang di bagian tertentu payudara
Apabila seseorang mengalami nyeri di bagian tertentu pada payudaranya dan rasa nyeri tersebut tidak kunjung hilang ada kemungkinan seseorang itu terdeteksi kanker payudara.
7. Payudara nampak kemerahan dan bengkak
Payudara nampak kemerahan, bengkak atau terlihat gelap di salah satu sisi dan bagian tertentu. Kemerahan yang terjadi pada payudara biasanya akan membawa kehangatan jika disentuh dan pada daerah tertentu saja. Sedangkan bengkak ini cenderung lebih lunak sifatnya.
8. Keluar cairan bercak darah atau encer (cairan selain susu)
Cairan dari puting susu yang perlu mendapat perhatian khusus diantaranya warna kemerahan atau encer disertai warna merah, pink atau coklat; lengket dan warnanya bening atau berwarna coklat sampai hitam (mengkilat); keluar secara spontan tanpa harus memijat puting susu; sifatnya menetap; hanya keluar dari salah satu puting susu.
9. Penarikan puting susu ke dalam
Pada umumnya seorang wanita yang mengalami retraksi puting susu atau penarikan puting susu ke dalam merupakan keadaan normal. Namun apabila hal ini hanya terjadi pada satu bagian payudara saja maka perlu diwaspadai dan perlu dilakukan pemeriksaan.
10. Terdapat benjolan di daerah ketiak
Selain benjolan pada payudara, benjolan yang terjadi di bawah ketiak juga perlu diwaspadai. Hal ini berkaitan dengan kelenjar getah bening di ketiak. Seseorang yang terdeteksi kanker payudara, sel kanker akan menyebar ke daerah kelenjar getah bening di ketiak (di sisi yang sama

dengan payudara yang terdeteksi kanker) sehingga menyebabkan benjolan di daerah ketiak.

11. Terdapat cekungan pada kulit payudara seperti lesung pipi
Sel-sel kanker dapat menyebabkan penumpukan cairan getah bening di payudara. Hal ini menyebabkan pembengkakan dan terbentuklah cekungan seperti lesung pipi. Selain itu cekungan pada kulit payudara diduga terjadi karena penarikan oleh tumor ganas yang berada pada payudara.
12. Durasi tidur
Lamanya waktu tidur seseorang secara rata-rata per 24 jam per hari. Durasi tidur orang normal pada umumnya sekitar 7-8 jam dalam sehari, sedangkan yang terkena kanker umumnya kurang dari 7-8 jam dalam sehari.
13. Latensi tidur atau waktu mulai tertidur
Durasi mulai dari berangkat tidur hingga tertidur. Seseorang dengan kualitas tidur baik menghabiskan waktu kurang dari 15 menit untuk dapat tertidur, sedangkan lebih dari 20 menit menandakan level insomnia yaitu seseorang yang mengalami kesulitan tidur.
14. Frekuensi terbangun
Frekuensi seseorang terbangun di malam hari pada saat jam tidur dan kesulitan untuk kembali tidur.
15. Riwayat pengobatan
Seseorang yang memiliki kemungkinan terkena kanker dilihat dari gejala-gejala kanker akan melakukan berbagai pengobatan mulai dari operasi, kemoterapi maupun herbal.

3.5 Struktur Data

Berikut merupakan struktur data yang akan digunakan dalam penelitian ini.

Tabel 3.2 Struktur Data Penelitian

Variabel Prediktor (X)								Variabel Respon (Y)
X ₁	X ₂	X ₃	X ₁₂	X ₁₃	X ₁₄	
X ₁₁	X ₂₁	X ₃₁	X ₁₂₁	X ₁₃₁	X ₁₄₁	Y ₁
X ₁₂	X ₂₂	X ₃₂	X ₁₂₂	X ₁₃₂	X ₁₄₂	Y ₂
X ₁₃	X ₂₃	X ₃₃	X ₁₂₃	X ₁₃₃	X ₁₄₃	Y ₃
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
X _{1n}	X _{2n}	X _{3n}	X _{12n}	X _{13n}	X _{14n}	Y _n

3.6 Langkah Analisis

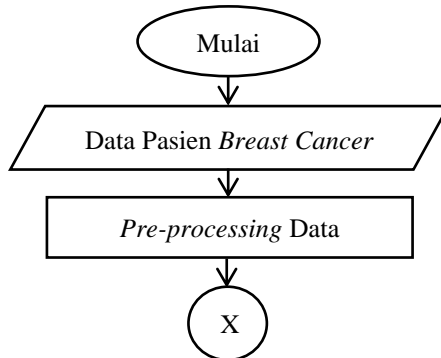
Langkah-langkah analisis yang akan dilakukan dalam penelitian ini didasarkan pada tujuan penelitian adalah sebagai berikut.

1. Melakukan pengumpulan data primer, yaitu data penderita dan bukan penderita kanker payudara yang diperoleh dari hasil survei di rumah terapi tauhid RSUD Nur Hidayah Bantul.
2. Melakukan *pre-processing* data, yaitu langkah awal sebelum pengolahan data lebih lanjut untuk mengidentifikasi data yang tidak lengkap, noisy maupun tidak konsisten.
3. Melakukan eksplorasi data untuk mengetahui deskripsi wanita dengan status pemeriksaan positif dan negatif kanker payudara dengan analisis statistika deskriptif.
4. Membagi data *training* dan *testing* menggunakan metode *5-fold cross validation*.
5. Melakukan analisis klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis dan faktor predisposisi dengan menggunakan metode *Classification Tree*. Berikut ini adalah tahapan dalam pengklasifikasian dengan metode *Classification Tree*.
 - a. Melakukan pembentukan pohon klasifikasi maksimal meliputi pemilahan (*splitter*) terbaik, penentuan simpul terminal (*terminal node*) dan penandaan label kelas.
 - b. Melakukan pemangkasan pohon klasifikasi dimulai dengan memangkas pohon klasifikasi maksimal sampai diperoleh ukuran pohon klasifikasi yang paling kecil

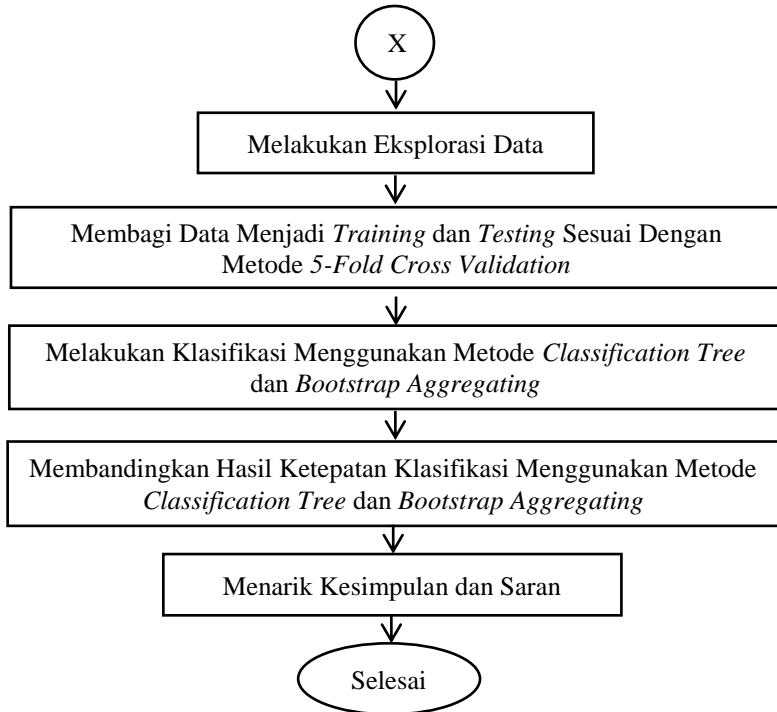
- dengan kriteria kompleksitas kesalahan (*cost complexity*) yang minimum.
- c. Melakukan pemilihan pohon terbaik dengan melakukan validasi model dengan memasukkan data *testing* pada pohon klasifikasi optimal.
 - d. Melakukan evaluasi ketepatan klasifikasi dengan meninjau *accuracy*, *sensitivity*, *specificity* dan nilai AUC.
6. Melakukan analisis klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis dan faktor predisposisi dengan menggunakan metode *Bootstrap Aggregating*.
 - a. Melakukan pemodelan *Bootstrap Aggregating* pada data *training* dengan 10, 15, 20, 25, 30, dan 35 replikasi *bootstrap*.
 - b. Melakukan pendugaan pengelompokan tiap data sampel dari 10, 15, 20, 25, 30, dan 35 replikasi yang telah dilakukan menggunakan *majority vote* untuk kasus klasifikasi.
 - c. Melakukan evaluasi ketepatan klasifikasi dengan meninjau *accuracy* dan AUC.
 7. Membandingkan hasil ketepatan klasifikasi menggunakan metode *Classification Tree* dan *Bootstrap Aggregating*.
 8. Menarik kesimpulan dan memberikan saran.

3.7 Diagram Alir

Berikut ini merupakan diagram alir dalam penelitian ini.



Gambar 3.2 Diagram Alir Penelitian



Gambar 3.3 Diagram Alir Penelitian (Lanjutan)

(Halaman ini sengaja dikosongkan)

BAB IV

ANALISIS DAN PEMBAHASAN

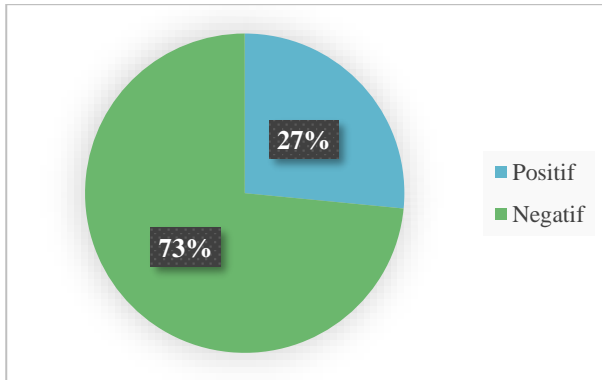
Pada bab ini akan ditampilkan hasil analisis dan pembahasan mengenai karakteristik dari suatu data, serta mendapatkan model klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara dalam mendiagnosa wanita yang memiliki status pemeriksaan positif maupun negatif kanker payudara menggunakan metode *Classification Tree* dan *Bootstrap Aggregating*, yang mana dari metode tersebut selanjutnya akan dilakukan perbandingan ketepatan klasifikasi di antara kedua metode tersebut dengan meninjau *accuracy*, *sensitivity*, *specificity*, dan nilai AUC.

4.1 Karakteristik Data

Pada penelitian Tugas Akhir ini, data hasil tes deteksi kanker payudara yang diperoleh dari hasil survei disajikan dalam bentuk *pie chart*, *bar chart* dan *box plot* untuk menggambarkan karakteristik wanita dengan status pemeriksaan positif kanker payudara dan wanita dengan status pemeriksaan negatif kanker payudara berdasarkan gejala klinis, dan faktor predisposisi yang meliputi usia, status pernikahan, kualitas tidur dan riwayat pengobatan.

4.1.1 Hasil Tes Deteksi Terhadap Kanker Payudara

Karakteristik wanita berdasarkan hasil tes deteksi kanker payudara pada *event* terapi pengobatan kanker yang diadakan di rumah terapi tauhid RSUD Nur Hidayah Bantul pada Gambar 4.1 menunjukkan persentase banyaknya wanita dengan status pemeriksaan positif kanker payudara dan negatif kanker payudara.

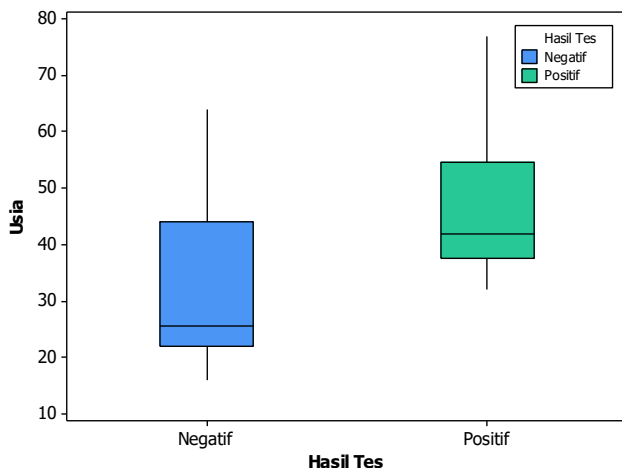


Gambar 4.1 Hasil Tes Deteksi Kanker Payudara

Berdasarkan Gambar 4.1 memberikan informasi bahwa dari 49 wanita yang datang pada event terapi pengobatan kanker sebesar 27% yaitu sebanyak 13 wanita diantaranya memiliki status pemeriksaan positif kanker payudara berdasarkan hasil tes deteksi kanker payudara yang pernah dilakukan sebelumnya. Adapun jenis tes deteksi kanker payudara yang pernah mereka lakukan diantaranya adalah Ultrasonografi (USG) dan Aspirasi Jarum Halus (AJH). Sedangkan sisanya sebesar 73% yaitu sebanyak 36 wanita memiliki status pemeriksaan negatif kanker payudara.

4.1.2 Usia Terhadap Status Pemeriksaan Kanker Payudara

Karakteristik wanita dengan status pemeriksaan positif dan negatif kanker payudara yang datang pada *event* terapi pengobatan kanker yang diadakan di rumah terapi tauhid RSUD Nur Hidayah Bantul berdasarkan usia dapat dilihat pada Gambar 4.2 sebagai berikut.

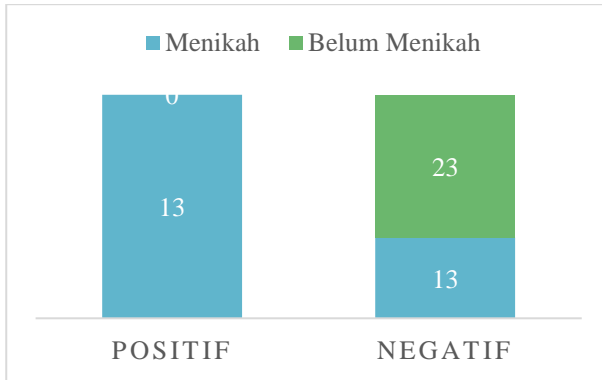


Gambar 4.2 Usia Terhadap Status Pemeriksaan Kanker Payudara

Berdasarkan Gambar 4.2 memberikan informasi bahwa 75% wanita dengan status pemeriksaan positif kanker payudara berada pada rentang usia antara 32 sampai 54 tahun dengan usia rata-rata 45 tahun. Sedangkan 75% wanita dengan status pemeriksaan negatif kanker payudara berada pada rentang usia antara 16 sampai 44 tahun dengan usia rata-rata 32 tahun. Hal ini menunjukkan bahwa semakin bertambahnya usia wanita memiliki resiko terkena kanker payudara lebih besar.

4.1.3 Status Pernikahan Terhadap Status Pemeriksaan Kanker Payudara

Karakteristik wanita dengan status pemeriksaan positif dan negatif kanker payudara yang datang pada *event* terapi pengobatan kanker yang diadakan di rumah terapi tauhid RSUD Nur Hidayah Bantul berdasarkan status pernikahan dapat dilihat pada Gambar 4.3 sebagai berikut.

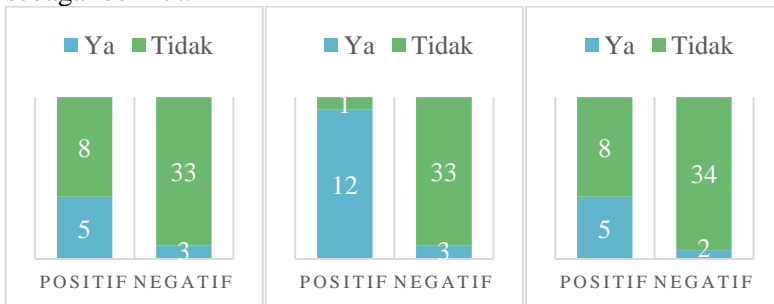


Gambar 4.3 Status Pernikahan Terhadap Status Kanker Payudara

Berdasarkan Gambar 4.3 memberikan informasi bahwa dari ke-13 wanita dengan status pemeriksaan positif kanker payudara adalah wanita yang sudah menikah. Sedangkan dari ke-36 wanita dengan status pemeriksaan negatif kanker payudara mayoritas merupakan wanita yang belum menikah yaitu sebanyak 23 wanita dan sisanya 13 wanita merupakan wanita yang sudah menikah.

4.1.4 Gejala Klinis Terhadap Status Pemeriksaan Kanker Payudara

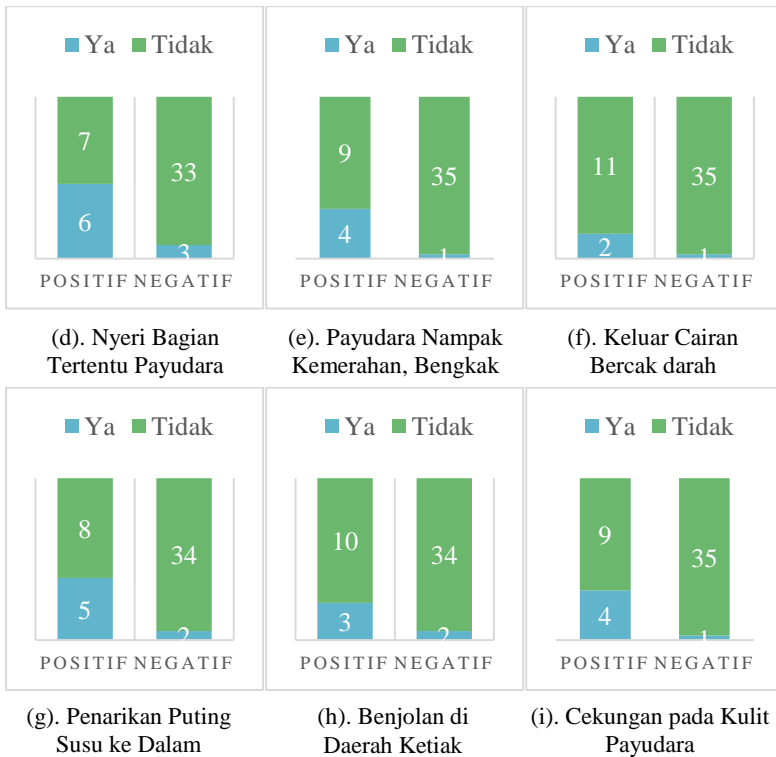
Karakteristik wanita dengan status pemeriksaan positif dan negatif kanker payudara yang datang pada *event* terapi pengobatan kanker yang diadakan di rumah terapi tauhid RSUD Nur Hidayah Bantul berdasarkan gejala klinis dapat dilihat pada Gambar 4.4 sebagai berikut.



(a). Perubahan Bentuk dan Ukuran Payudara

(b). Benjolan Sekitar Payudara

(c). Kerutan Sekitar Kulit Payudara



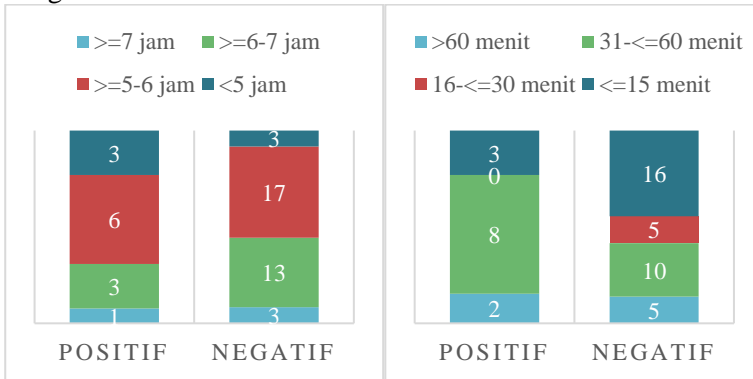
Gambar 4.4 Gejala Klinis Terhadap Status Kanker Payudara

Berdasarkan Gambar 4.4 memberikan informasi bahwa wanita dengan status pemeriksaan positif kanker payudara, menunjukkan 5 dari 13 wanita mengalami gejala klinis terjadi perubahan bentuk dan ukuran pada payudara, mengalami penarikan puting susu ke dalam dan terdapat kerutan di sekitar kulit payudara. Kemudian 12 dari 13 wanita terdapat benjolan di sekitar payudara, 6 dari 13 wanita mengalami nyeri di bagian tertentu payudara, 2 dari 13 wanita mengalami keluar cairan bercak darah dari payudara, 3 dari 13 wanita terdapat benjolan di daerah ketiak serta 4 dari 13 wanita terdapat cekungan pada kulit payudara dan payudara nampak kemerahan dan bengkak.

Sedangkan wanita dengan status pemeriksaan negatif kanker payudara mayoritas tidak mengalami gejala klinis seperti perubahan bentuk dan ukuran payudara, benjolan di sekitar payudara, ke-
rutan di sekitar kulit payudara, nyeri di bagian tertentu payudara, payudara nampak kemerahan dan bengkak, keluar cairan bercak darah, penarikan puting susu ke dalam, benjolan di daerah ketiak, maupun cekungan pada kulit payudara.

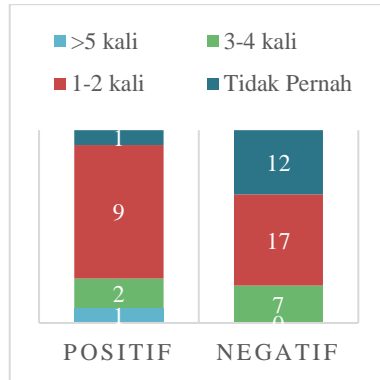
4.1.5 Kualitas Tidur Terhadap Status Pemeriksaan Kanker Payudara

Karakteristik wanita dengan status pemeriksaan positif dan negatif kanker payudara yang datang pada *event* terapi pengobatan kanker yang diadakan di rumah terapi tauhid RSUD Nur Hidayah Bantul berdasarkan kualitas tidur dapat dilihat pada Gambar 4.5 sebagai berikut.



(a). Durasi Tidur

(b). Waktu Mulai Tidur



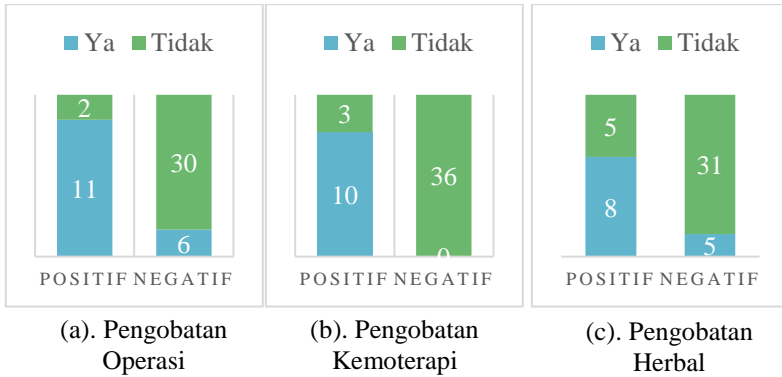
(c). Frekuensi Terbangun

Gambar 4.5 Kualitas Tidur Terhadap Status Kanker Payudara

Pada Gambar 4.5 memberikan informasi bahwa wanita dengan status pemeriksaan positif kanker payudara mayoritas memiliki durasi tidur antara $\geq 5 - 6$ jam, waktu mulai tidur antara $31 - \leq 60$ menit, dan frekuensi terbangun di malam hari antara 1 - 2 kali. Sedangkan wanita dengan status pemeriksaan negatif kanker payudara mayoritas memiliki durasi tidur antara $\geq 5 - 6$ jam, waktu mulai tidur antara ≤ 15 menit dan frekuensi terbangun di malam hari antara 1-2 kali.

4.1.6 Riwayat Pengobatan Terhadap Status Pemeriksaan Kanker Payudara

Karakteristik wanita dengan status positif dan negatif kanker payudara yang datang pada *event* terapi pengobatan kanker yang diadakan di rumah terapi tauhid RSUD Nur Hidayah Bantul berdasarkan riwayat pengobatan dapat dilihat pada Gambar 4.6 sebagai berikut.



Gambar 4.6 Riwayat Pengobatan Terhadap Status Kanker Payudara

Berdasarkan Gambar 4.6 memberikan informasi bahwa mayoritas wanita dengan status pemeriksaan positif kanker payudara berdasarkan hasil tes deteksi kanker payudara yang pernah dilakukan adalah memiliki riwayat pengobatan operasi sebanyak 11 dari 13 wanita, riwayat pengobatan kemoterapi sebanyak 10 dari 13 wanita dan riwayat pengobatan herbal sebanyak 8 dari 13 wanita. Sedangkan mayoritas wanita dengan status pemeriksaan negatif kanker payudara adalah tidak memiliki riwayat pengobatan operasi sebanyak 30 dari 36 wanita, riwayat pengobatan kemoterapi sebanyak 36 wanita dan riwayat pengobatan herbal sebanyak 31 dari 36 wanita.

4.2 Klasifikasi Hasil Tes Deteksi Kanker Payudara dengan Menggunakan Metode *Classification Tree*

Dilihat dari salah satu tujuan penelitian Tugas Akhir yaitu untuk mendapatkan model klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara dengan menggunakan metode *Classification Tree*. Adapun variabel respon yang digunakan pada penelitian ini berupa data kategorik yaitu wanita dengan status pemeriksaan positif kanker payudara dan wanita dengan status pemeriksaan negatif kanker payudara, sehingga metode analisis yang digunakan akan menghasilkan suatu pohon klasifikasi (*Classification Tree*).

Sesuai dengan prosedur algoritma *Classification Tree* yang telah dijelaskan di bab 2 tinjauan pustaka, maka terdapat 3 tahapan dalam klasifikasi menggunakan metode tersebut, yaitu mulai dari pembentukan pohon klasifikasi maksimal, pemangkasan pohon klasifikasi maksimal (*pruning*), dan penentuan pohon klasifikasi optimal.

Pada ilustrasi berikut, menggunakan semua variabel prediktor yaitu sebanyak 17 variabel yang mana di dalam pembentukan pohon klasifikasi dibutuhkan variabel-variabel yang berperan sebagai pemilah. Berikut merupakan banyaknya kemungkinan pemilah untuk membentuk pohon klasifikasi ditunjukkan pada Tabel 4.1.

Tabel 4.1 Banyaknya Kemungkinan Pemilah Variabel Prediktor

Variabel Prediktor	Skala Data	Jumlah Kategori	Kemungkinan Pemilah
Usia (X_1)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Status pernikahan (X_2)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Perubahan bentuk dan ukuran payudara (X_3)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Benjolan di sekitar payudara (X_4)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Kerutan di sekitar kulit payudara (X_5)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Nyeri di bagian tertentu payudara (X_6)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Payudara nampak kemerahan dan bengkak (X_7)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Keluar cairan dengan bercak darah (X_8)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Penarikan puting susu ke dalam (X_9)	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Benjolan di daerah ketiak (X_{10})	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Cekungan pada kulit payudara (X_{11})	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Durasi tidur (X_{12})	Ordinal	4	$4 - 1 = 3$ pemilah

Tabel 4.1 Banyaknya Kemungkinan Pemilah Variabel Prediktor (Lanjutan)

Variabel Prediktor	Skala Data	Jumlah Kategori	Kemungkinan Pemilah
Waktu mulai tidur (X_{13})	Ordinal	4	$4 - 1 = 3$ pemilah
Frekuensi terbangun (X_{14})	Ordinal	4	$4 - 1 = 3$ pemilah
Riwayat pengobatan operasi (X_{15})	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Riwayat pengobatan kemoterapi (X_{16})	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
Riwayat pengobatan herbal (X_{17})	Nominal	2	$2^{2-1} - 1 = 1$ pemilah

Berdasarkan 49 data observasi hasil survei yang digunakan dalam penelitian ini, terdapat 13 responden yang dikategorikan sebagai wanita dengan status pemeriksaan positif kanker payudara, sedangkan sisanya 36 responden yang dikategorikan sebagai wanita dengan status pemeriksaan negatif kanker payudara. Karena data penelitian berukuran kecil, maka di dalam penentuan pohon klasifikasi optimal digunakan metode *k-fold cross validation estimate* dimana jumlah *fold* yang digunakan adalah sebanyak *5-fold* ($k=5$), sehingga data akan dibagi menjadi 5 bagian dan dimasing-masing bagian berjumlah 9 dan 10 data.

4.2.1 Pembentukan Pohon Klasifikasi Maksimal

Pada mulanya dalam pembentukan pohon klasifikasi berupa pohon klasifikasi maksimal, yaitu pohon klasifikasi yang memiliki jumlah simpul terbanyak, setiap pemilah pada suatu simpul hanya bergantung pada nilai yang berasal dari suatu variabel prediktor, di mana setiap variabel prediktor mempunyai sejumlah kemungkinan pemilah, tergantung pada skala data variabel prediktor tersebut. Pada penelitian ini digunakan Indeks Gini untuk menentukan nilai *goodness of split* dari masing-masing pemilah. Pemilah yang akan terpilih adalah variabel pemilah yang memiliki nilai *goodness of split* tertinggi. Pemilah yang terpilih merupakan variabel terpenting atau utama dalam mengklasifikasikan data observasi. Hasil dari

pengolahan data didapatkan urutan kepentingan variabel prediktor yang berpengaruh terhadap kanker payudara dengan pembagian data *training* dan *testing* menggunakan *cross validation* untuk 5-fold ditunjukkan pada Tabel 4.2 sebagai berikut.

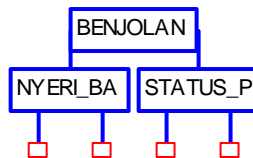
Tabel 4.2 Urutan Variabel Pembentuk Pohon Klasifikasi Maksimal

Variabel	Skor	
Benjolan di sekitar payudara (X_4)	100,00	
Riwayat pengobatan kemoterapi (X_{16})	88,63	
Status pernikahan (X_2)	88,24	
Riwayat pengobatan operasi (X_{15})	65,48	
Nyeri di bagian tertentu payudara (X_6)	45,71	
Riwayat pengobatan herbal (X_{17})	34,27	
Waktu mulai tertidur (X_{13})	6,82	
Usia (X_1)	0,00	
Perubahan bentuk dan ukuran payudara (X_3)	0,00	
Kerutan di sekitar kulit payudara (X_5)	0,00	
Payudara nampak kemerahan dan bengkak (X_7)	0,00	
Keluar cairan dengan bercak darah atau encer (X_8)	0,00	
Penarikan puting susu ke dalam (X_9)	0,00	
Benjolan di daerah ketiak (X_{10})	0,00	
Cekungan pada kulit payudara (X_{11})	0,00	
Durasi tidur (X_{12})	0,00	
Frekuensi terbangun (X_{14})	0,00	

Berdasarkan Tabel 4.2 memberikan informasi bahwa tidak semua variabel prediktor menjadi pembangun dalam pembentukan pohon klasifikasi. Namun berdasarkan skor yang dihasilkan diketahui bahwa variabel terpenting yang memiliki kontribusi terbesar adalah variabel benjolan sekitar payudara (X_4) karena mempunyai skor variabel tertinggi yakni 100. Artinya variabel benjolan sekitar

payudara (X_4) akan menjadi pemilah awal atau biasa disebut sebagai terminal induk (*parent nodes*). Variabel benjolan sekitar payudara (X_4) mampu menurunkan tingkat heterogenan paling besar, sehingga simpul yang dihasilkan lebih homogen. Selain itu, terdapat beberapa variabel lain yang juga berpengaruh besar dalam pembentukan pohon klasifikasi yaitu variabel riwayat pengobatan kemoterapi (X_{16}), status pernikahan (X_2), riwayat pengobatan operasi (X_{15}), nyeri di bagian tertentu payudara (X_6), riwayat pengobatan herbal (X_{17}), dan waktu mulai tidur (X_{13}).

Setelah diperoleh urutan variabel prediktor sebagai variabel pembentuk pohon klasifikasi maksimal maka selanjutnya dilakukan proses pemilahan dengan fungsi heterogenan indeks gini sampai terbentuk pohon klasifikasi maksimal. Pohon klasifikasi maksimal merupakan pohon klasifikasi yang menghasilkan simpul terminal dengan jumlah paling banyak atau maksimal dengan tingkat kedalaman (*depth*) tertinggi. Pada Gambar 4.7 ditampilkan hasil pohon klasifikasi maksimal dengan jumlah simpul terminal sebanyak 4 simpul terminal (*terminal node*) dan kedalaman (*depth*) pohon 2.

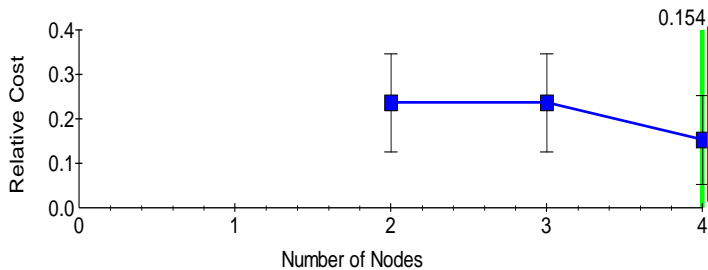


Gambar 4.7 Bentuk Pohon Klasifikasi Maksimal

Berdasarkan Gambar 4.7 memberikan informasi bahwa variabel benjolan sekitar payudara (X_4) merupakan variabel pemilah utama dan paling penting dalam menentukan klasifikasi hasil tes deteksi kanker payudara. Dikarenakan jumlah data yang sedikit yaitu sebanyak 49 data observasi menghasilkan pohon klasifikasi dengan jumlah simpul yang minimum dan tiga variabel penting sebagai pencari dari suatu pengklasifikasian yaitu variabel benjolan sekitar payudara (X_4), nyeri bagian tertentu payudara (X_6) dan status pernikahan (X_2).

4.2.2 Pemangkasan Pohon Klasifikasi Maksimal

Pemangkasan pohon klasifikasi maksimal merupakan tahapan yang dilakukan selanjutnya setelah pembentukan pohon klasifikasi maksimal. Tahapan pemangkasan pohon klasifikasi maksimal bertujuan untuk menghindari adanya kasus *overfitting* dimana suatu keadaan akurasi yang dihasilkan melebihi kenyataannya atau bisa juga *underfitting* di mana nilai akurasi yang dihasilkan kurang apabila dibandingkan dengan nilai aktual. Proses pemangkasan pohon dilakukan berdasarkan pendekatan *cost complexity minimum*. Pembagian data *training* dan *testing* menggunakan metode *k-fold cross validation* memberikan kesempatan yang sama bagi seluruh data untuk menjadi data *training* maupun *testing*. Data *training* digunakan sebagai pembangun model pohon klasifikasi, sedangkan data *testing* digunakan sebagai penduga kesalahan dalam klasifikasi atau yang biasa disebut dengan *misclassification error rate*. Berikut merupakan nilai *complexity parameter* pada pohon klasifikasi yang dihasilkan dari data *training* menggunakan *5-fold cross validation*.



Gambar 4.8 Plot *Relative Cost* Dan Banyaknya Simpul Terminal

Berdasarkan Gambar 4.8 menunjukkan bahwa nilai *relative cost* yang diperoleh pohon klasifikasi maksimal dengan jumlah simpul terminal 4 adalah sebesar 0,154 (garis merah). Nilai tersebut sama jika dibandingkan dengan nilai *relative cost* pada klasifikasi optimal yang memiliki simpul terminal sebanyak 4 simpul yaitu sebesar 0,154 (garis hijau). Nilai kompleksitas pohon klasifikasi optimal yang diperoleh sebesar 0,000 dan biaya kesalahan-

nya sebesar 0,154 +/- 0,100 atau antara 0,054 sampai 0,254. Hasil selengkapnya dapat dilihat pada Tabel 4.3 berikut ini. Karena nilai *relative cost* pohon klasifikasi optimal pada Gambar 4.8 sama dibandingkan dengan pohon klasifikasi maksimal sehingga pohon klasifikasi maksimal dipilih sebagai pohon yang layak untuk pohon klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara di RSUD Nur Hidayah Bantul.

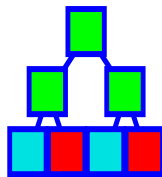
Tabel 4.3 Urutan Pembentukan Pohon Klasifikasi (*Tree Sequence*)

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Test Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Comple xity</i>
1**	4	0,154 +/- 0,100	0,000	0,000
2	3	0,237 +/- 0,110	0,077	0,038
3	2	0,237 +/- 0,110	0,160	0,042
4	1	1,000 +/- 0,863167E-04	1,000	0,420

** *pohon klasifikasi optimal*

4.2.3 Penentuan Pohon Klasifikasi Optimal

Setelah diperoleh pembentukan pohon klasifikasi secara iteratif menghasilkan pohon klasifikasi optimal dengan jumlah simpul terminal sebanyak 4 simpul dengan kedalaman 2. Gambar 4.9 menampilkan topologi pohon klasifikasi optimal, penentuan pohon klasifikasi optimal digunakan untuk mengetahui akurasi model klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara.

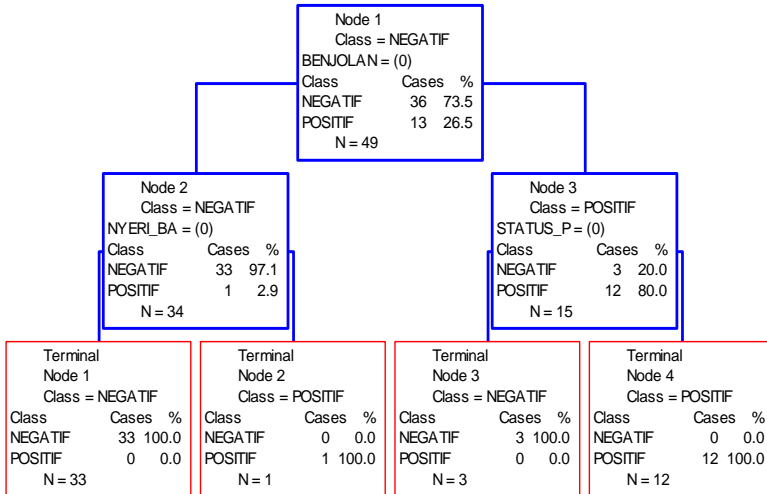


Gambar 4.9 Topologi Pohon Klasifikasi Optimal

Pada Gambar 4.9, terdapat warna simpul dengan degradasi warna antara merah dan biru yang mana kedua warna tersebut menunjukkan jumlah simpul terminal sebanyak 2 simpul. Perbe-

daan warna tersebut ada kaitannya dengan pemberian label kelas di masing-masing simpul terminal. Simpul terminal berwarna biru apabila suatu simpul diklasifikasikan ke dalam kelas 0 yaitu diklasifikasikan sebagai wanita dengan status pemeriksaan negatif kanker payudara. Jika warna simpul terminal semakin biru pekat menunjukkan bahwa wanita dengan status pemeriksaan negatif kanker payudara hasil klasifikasi tersebut termasuk kelas 0 dengan persentase mendekati 100 persen. Namun apabila warna simpul terminal semakin biru memudar menunjukkan bahwa wanita dengan status pemeriksaan negatif kanker payudara hasil klasifikasi tersebut termasuk kelas 0 dengan persentase menurun. Sedangkan simpul berwarna merah apabila suatu simpul diklasifikasikan ke dalam kelas 1 yaitu diklasifikasikan sebagai wanita dengan status positif kanker payudara. Jika warna simpul terminal semakin merah pekat menunjukkan bahwa wanita dengan status pemeriksaan positif kanker payudara hasil klasifikasi tersebut termasuk kelas 1 dengan persentase mendekati 100 persen. Begitu juga sebaliknya jika warna simpul terminal semakin merah memudar menunjukkan bahwa wanita dengan status pemeriksaan positif kanker payudara hasil klasifikasi tersebut termasuk kelas 1 dengan persentase menurun.

Berdasarkan penjelasan sebelumnya pada bagian 4.2.2 mengenai pemangkasan pohon klasifikasi maksimal, karena nilai *relative cost* pohon klasifikasi optimal sama dengan pohon klasifikasi maksimal sehingga pohon klasifikasi maksimal dipilih sebagai pohon yang layak untuk pohon klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara di RSUD Nur Hidayah Bantul. Berikut ini merupakan bentuk dari pohon klasifikasi optimal berdasarkan gejala klinis yang diduga mempengaruhi terjadinya kanker payudara.



Gambar 4.10 Bentuk Pohon Klasifikasi Optimal

Berdasarkan topologi pohon klasifikasi, pada bagian simpul terminal merupakan simpul akhir dalam hasil pemilahan yang dibedakan menjadi dua kategori dalam pelabelan kelas. Pelabelan kelas dilakukan untuk mengetahui karakteristik dari status pemeriksaan kanker payudara pada wanita. Apabila kelas dilabelkan sebagai 0 maka seorang wanita diklasifikasikan ke dalam status pemeriksaan negatif kanker payudara. Sedangkan jika kelas dilabelkan sebagai 1 maka seorang wanita diklasifikasikan ke dalam status pemeriksaan positif kanker payudara.

Di dalam proses pemilahan mengacu pada pohon klasifikasi optimal yang ditampilkan pada Gambar 4.6. Pemilahan diawali dengan memilah sebanyak 49 wanita berdasarkan gejala klinis dengan benjolan di sekitar payudara (X_4) dan nilai *threshold* 0 yang artinya tidak terdapat benjolan sekitar payudara. Setelah itu dari 49 data yang terdiri dari 13 data dengan label positif kanker payudara dan 36 data dengan label negatif kanker payudara tersebut dibagi menjadi 2 bagian yaitu simpul kiri (*node 2*) dan simpul kanan (*node 3*) yang masing-masing terdiri dari 34 dan 15 wanita dengan kondisi untuk simpul kiri adalah tidak terdapat benjolan di sekitar pa-

yudara dan diberi label kelas sebagai wanita dengan status pemeriksaan negatif kanker payudara. Namun untuk simpul kanan adalah terdapat benjolan di sekitar payudara dan diberi label kelas sebagai wanita dengan status pemeriksaan positif kanker payudara. Pemilahan tersebut terus berulang menjadi simpul anak baru (kiri dan kanan) sesuai prosedur *binary recursive partitioning* pada bab 2 tinjauan pustaka, sampai simpul tersebut telah dianggap memiliki anggota yang homogen atau jika simpul tersebut hanya memiliki 1 anggota pengamatan maka simpul akan menjadi simpul terminal dan tidak akan dilakukan pemilahan kembali dan untuk masing-masing simpul terminal memiliki karakteristik dan diprediksi sebagai kelas variabel respon tertentu (dalam hal ini status pemeriksaan kanker payudara) sesuai dengan label kelas yang diberikan. Pada penelitian ini, hasil dari pohon klasifikasi optimal menunjukkan bahwa pemilahan berhenti pada jumlah terminal simpul sebanyak 4 simpul dengan kedalaman pohon sebanyak 4 dapat dilihat pada Gambar 4.10.

Berikut ini penjelasan karakteristik simpul terminal yang mengklasifikasikan data hasil tes deteksi kanker payudara menjadi salah satu dari kedua status pemeriksaan kanker payudara.

1. Simpul terminal 1 terdiri dari 33 wanita yang diprediksi sebagai wanita dengan status pemeriksaan negatif kanker payudara. Karakteristik wanita pada simpul ini adalah wanita yang tidak memiliki gejala klinis berupa benjolan di sekitar payudara dan nyeri di bagian tertentu pada payudara.
2. Simpul terminal 2 terdiri dari 1 wanita yang diprediksi sebagai wanita dengan status pemeriksaan positif kanker payudara. Karakteristik wanita pada simpul ini adalah tidak memiliki gejala klinis berupa benjolan di sekitar payudara namun mengalami gejala klinis berupa nyeri di bagian tertentu pada payudara.
3. Simpul terminal 3 terdiri dari 3 wanita yang diprediksi sebagai wanita dengan status pemeriksaan negatif kanker payudara. Karakteristik wanita pada simpul ini adalah wanita

yang memiliki gejala klinis berupa benjolan di sekitar payudara dan memiliki status pernikahan belum menikah.

4. Simpul terminal 4 terdiri dari 12 wanita yang diprediksi sebagai wanita dengan status pemeriksaan positif kanker payudara. Karakteristik wanita pada simpul ini adalah wanita yang memiliki gejala klinis berupa terdapat benjolan di sekitar payudara dan memiliki status pernikahan sudah menikah.

Sesuai dengan penjelasan pada masing-masing simpul maka karakteristik wanita dengan status pemeriksaan positif dan negatif kanker payudara berdasarkan keempat simpul adalah sebagai berikut.

Tabel 4.4 Karakteristik Wanita Berdasarkan Simpul Terminal

Positif Kanker Payudara	Negatif Kanker Payudara
<ul style="list-style-type: none"> • Tidak terdapat benjolan di sekitar payudara • Terdapat benjolan di sekitar payudara • Mengalami nyeri di bagian tertentu pada payudara • Memiliki status pernikahan sudah menikah 	<ul style="list-style-type: none"> • Tidak terdapat benjolan di sekitar payudara • Terdapat benjolan di sekitar payudara • Tidak mengalami nyeri di bagian tertentu pada payudara • Memiliki status pernikahan belum menikah

4.2.4 Evaluasi Ketepatan Klasifikasi Menggunakan Metode *Classification Tree*

Berikut ini merupakan hasil evaluasi ketepatan klasifikasi pada data *training* yang dihasilkan dengan menggunakan metode *Classification Tree* ditunjukkan pada Tabel 4.5 sebagai berikut.

Tabel 4.5 Evaluasi Performansi Pada Data *Training*

Fold	Accuracy	Sensitivity	Specificity	AUC
1	100%	100%	100%	100%
2	87,18%	100%	82,76%	91,38%
3	100%	100%	100%	100%
4	100%	100%	100%	100%
5	100%	100%	100%	100%
Rata-Rata	97,44%	100%	96,55%	98,28%

Tabel 4.5 di atas menunjukkan bahwa pada data *training*, model yang terbaik ditunjukkan oleh *fold* 1, 3, 4 dan 5 karena keempatnya sama-sama memiliki nilai akurasi paling tinggi yaitu sebesar 100%. Sedangkan untuk rata-rata akurasi yang diperoleh dengan menggunakan *5-fold cross validation* sebesar 97,44%. Artinya bahwa pohon klasifikasi optimal mampu mengklasifikasikan hasil tes deteksi kanker payudara ke dalam kelas status pemeriksaan kanker payudara dengan tepat sebesar 97,44% dengan rata-rata nilai sensitivitas, spesifisitas, dan AUC adalah sebesar 100%, 96,55% dan 91,38%.

Pohon klasifikasi optimal yang terbentuk perlu divalidasi untuk mengetahui apakah pohon klasifikasi tersebut layak dan dapat digunakan untuk mengklasifikasikan data baru. Berikut ini merupakan *confusion matrix* data *testing* yang dihasilkan untuk *5-fold cross validation* ditampilkan pada Tabel 4.7 sebagai berikut.

Tabel 4.6 *Confusion Matrix Data Testing*

Aktual	Prediksi		Total
	Negatif=0	Positif=1	
Negatif=0	36	0	36
Positif=1	2	11	13
Total	38	11	49

Tabel 4.6 di atas menunjukkan bahwa pada data *testing* menggunakan *5-fold cross validation*, wanita dengan status pemeriksaan negatif kanker payudara yang tepat diklasifikasikan ke kelas negatif kanker payudara ada sebanyak 36 orang dan tidak terdapat orang yang salah diklasifikasikan ke kelas positif kanker payudara. Sedangkan wanita dengan status pemeriksaan positif kanker payudara yang tepat diklasifikasikan ke kelas positif kanker payudara ada sebanyak 11 orang dan terdapat 2 orang yang salah diklasifikasikan ke kelas negatif kanker payudara. Selanjutnya dilakukan perhitungan *total accuracy*, *sensitivity*, *specificity* dan nilai AUC untuk data *testing* sesuai dengan persamaan (2.12), (2.13), (2.14) dan (2.15) sebagai berikut.

$$TotalAccuracy = \frac{36 + 11}{36 + 0 + 2 + 11} \times 100 = 95,92\%$$

$$\text{Specificity} = \frac{36}{36+0} \times 100 = 100\%$$

$$\text{Sensitivity} = \frac{11}{2+11} \times 100 = 84,62\%$$

$$\text{AUC} = \frac{1}{2} \times (0,8462 + 1) \times 100 = 92,31\%$$

Berdasarkan perhitungan *total accuracy*, *sensitivity*, *specificity* dan nilai AUC yang diperoleh ketepatan klasifikasi pohon optimal data *testing* sebesar 95,92%. Artinya bahwa pohon klasifikasi optimal mampu mengklasifikasikan data baru hasil tes deteksi kanker payudara ke dalam kelas status pemeriksaan kanker payudara dengan tepat sebesar 95,92% dengan *specificity* sebesar 100%, *sensitivity* sebesar 84,62% dan AUC adalah sebesar 92,31%. Karena total akurasi yang sudah cukup tinggi, maka pohon klasifikasi optimal yang terbentuk sudah layak digunakan untuk pengklasifikasi hasil tes deteksi kanker payudara pada wanita.

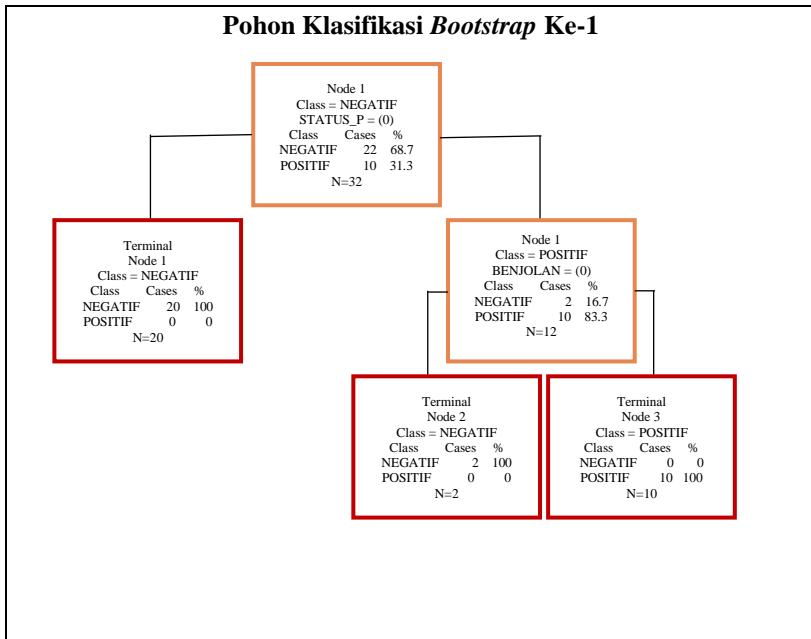
4.3 Klasifikasi Hasil Tes Deteksi Kanker Payudara dengan Menggunakan Metode *Bootstrap Aggregating*

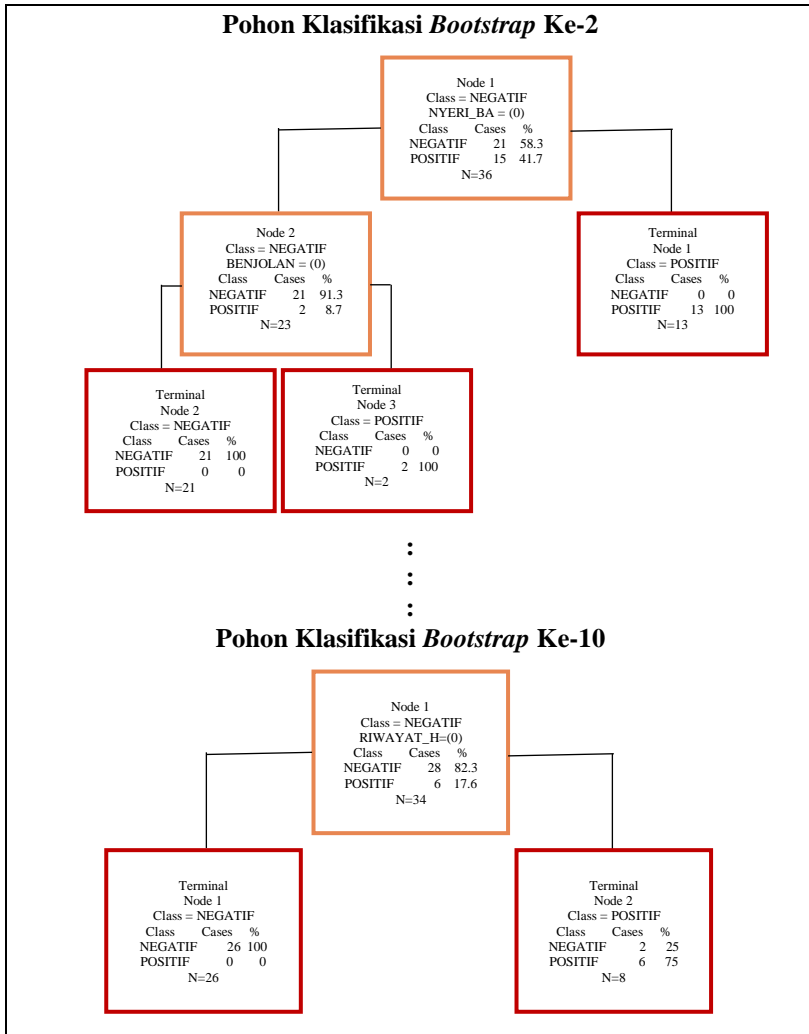
4.3.1 Pemodelan *Bootstrap Aggregating* Dengan 10, 15, 20, 25, 30, dan 35 Replikasi *Bootstrap*

Setelah dilakukan klasifikasi hasil tes deteksi kanker payudara dengan menggunakan metode *Classification Tree*, selanjutnya akan dilakukan klasifikasi dengan menggunakan metode *Bootstrap Aggregating* di mana dari data asli dilakukan pembagian menjadi data *training* dan *testing*. Kemudian pada data *training* dilakukan *bootstrap* sebanyak 10, 15, 20, 25, 30, dan 35 replikasi. Masing-masing *bootstrap* di setiap replikasi diambil data secara random dengan pengembalian sebanyak $n \leq N$ di mana n adalah banyaknya jumlah data hasil dari *bootstrap* sedangkan N adalah banyaknya jumlah data pada data asli. Hasil dari masing-masing *bootstrap* $\{B_1, B_2, B_3, \dots, B_n\}$ didapatkan model berupa pohon klasifikasi sebanyak replikasi *bootstrap* $\{M_1, M_2, M_3, \dots, M_n\}$. Kemudian dari model yang terbentuk dilakukan prediksi dengan menguji coba data *testing* ke dalam setiap model pohon klasifikasi sehingga diperoleh hasil ketepatan klasifikasi untuk masing-masing model.

Model yang terpilih didapatkan dari pendugaan pengelompokan tiap data sampel dari 10, 15, 20, 25, 30, dan 35 replikasi yang telah dilakukan menggunakan *majority vote* untuk kasus klasifikasi. Jadi cara kerja dari *majority vote* ini yang pertama dapat dilihat dari model pohon klasifikasi yang terbentuk dari hasil *bootstrap*. Kemudian yang kedua dilihat dari hasil ketepatan klasifikasinya. Model yang terpilih merupakan model yang konsisten dalam hal pohon klasifikasi dan akurasi yang diperoleh dari model.

Sebagai ilustrasi untuk lebih jelasnya, berikut ini merupakan hasil pohon klasifikasi yang diperoleh dengan menggunakan metode *Bootstrap Aggregating* dengan 10 replikasi *bootstrap*. Pada kasus ini data dibagi menjadi dua yaitu data *training* dan *testing* di mana persentase data *testing* yang digunakan adalah 19% atau setara dengan 10 data. Sedangkan sisanya sebagai data *training*. Karena dilakukan 10 kali replikasi *bootstrap* jadi terdapat 10 model pohon klasifikasi yang terbentuk.





Gambar 4.11 Diagram Pohon Klasifikasi Hasil *Bootstrap*

Setelah dilakukan proses *bootstrap* sebanyak 10 kali pembentukan pohon klasifikasi dan prediksi data *testing* pada masing-masing pohon klasifikasi, langkah selanjutnya adalah dilakukan

prediksi gabungan (*aggregating*) berdasarkan 10 dugaan pada data *testing* dengan aturan *majority vote* (suara terbanyak). Sebagai contoh terdapat data *testing* baru yang akan diprediksi apakah masuk ke dalam kelas positif kanker payudara atau negatif kanker payudara. Berdasarkan hasil prediksi 10 pohon klasifikasi, data *testing* baru yang diprediksi masuk ke kelas positif kanker payudara sebanyak 7 sedangkan yang diprediksi masuk kelas negatif kanker payudara sebanyak 3. Hasil akhir pendugaan berdasarkan suara terbanyak, sehingga prediksi akhir data *testing* baru masuk ke dalam kelas positif kanker payudara. Hasil dari prediksi gabungan yang telah diperoleh selanjutnya digunakan untuk menguji ketepatan klasifikasi pada penerapan metode Bagging CART.

Proses di atas dilakukan dengan cara yang sama untuk replikasi *bootstrap* yang berbeda yaitu 15, 20, 25, 30, dan 35 replikasi. Kemudian masing-masing akan dilakukan perbandingan hasil ketepatan klasifikasinya yang menghasilkan tingkat akurasi dan AUC konsisten yang terpilih sebagai model yang terbaik yang dapat digunakan untuk mengklasifikasikan hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara.

4.3.2 Evaluasi Ketepatan Klasifikasi Menggunakan Metode *Bootstrap Aggregating* Dengan 10, 15, 20, 25, 30, dan 35 Replikasi *Bootstrap*

Berikut ini merupakan perbandingan ketepatan klasifikasi yang dihasilkan oleh pohon klasifikasi dengan menggunakan metode *Bootstrap Aggregating* dengan 10, 15, 20, 25, 30, dan 35 replikasi ditinjau dari akurasi dan AUC ditampilkan pada Tabel 4.7.

Tabel 4.7 Perbandingan Keباikan Hasil Klasifikasi Menggunakan Metode *Bootstrap Aggregating*

Replikasi	<i>Training</i>		<i>Testing</i>	
	<i>Accuracy</i>	<i>AUC</i>	<i>Accuracy</i>	<i>AUC</i>
10	97,14	98,15	80	88,89
15	97,14	98,15	80	88,89
20	97,22	98,45	80	88,89
25	97,22	98,45	80	88,89
30	97,22	98,45	80	88,89
35	97,22	98,45	80	88,89

Berdasarkan Tabel 4.7 menunjukkan bahwa dengan menggunakan metode *Bootstrap Aggregating* untuk 20 replikasi *bootstrap* sudah mampu dalam mengklasifikasikan hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara dengan tingkat akurasi dan AUC adalah sebesar 97,22% dan 98,45% untuk data *training*. Sedangkan untuk data *testing* adalah sebesar 80% dan 88,89%. Karena mulai dari replikasi 20 ke atas sudah menunjukkan tingkat akurasi dan AUC yang konsisten sehingga model terbaik yang dapat digunakan adalah *Bootstrap Aggregating* dengan 20 replikasi.

4.4 Perbandingan Ketepatan Klasifikasi Menggunakan Metode *Classification Tree* dan *Bootstrap Aggregating*

Pada tahapan akhir, setelah dilakukan klasifikasi hasil tes deteksi kanker payudara berdasarkan gejala klinis kanker payudara baik dengan menggunakan metode *Classification Tree* maupun *Bootstrap Aggregating* dapat dilakukan perbandingan ketepatan klasifikasi untuk menentukan model terbaik diantara kedua metode. Berikut ini ditampilkan tabel perbandingan ketepatan klasifikasi antara metode *Classification Tree* dengan *Bootstrap Aggregating* dengan menggunakan pembagian data *training* dan *testing* dimana persentase data *testing* adalah 19% atau sebanyak 10 data sedangkan sisanya 39 data sebagai data *training*.

Tabel 4.8 Perbandingan Kebaikan Hasil Klasifikasi Menggunakan Metode *Classification Tree* dan *Bootstrap Aggregating*

	<i>Classification Tree</i>		<i>Bagging 20 Replikasi</i>	
	<i>Accuracy</i>	<i>AUC</i>	<i>Accuracy</i>	<i>AUC</i>
<i>Training</i>	100%	100%	97,22%	98,45%
<i>Testing</i>	100%	100%	80%	88,89%

Berdasarkan Tabel 4.8 di atas memberikan informasi bahwa metode *Bootstrap Aggregating* tidak lebih baik dari metode *Classification Tree* karena baik untuk data *training* maupun *testing* menghasilkan ketepatan klasifikasi lebih tinggi dibandingkan dengan metode *Bootstrap Aggregating*. Namun dalam hal ini walaupun hasil dari *Bagging* tidak selalu meningkatkan hasil akurasi dari metode *Classification Tree*, akan tetapi dapat menghasilkan model yang konsisten. Dan tingginya akurasi sebesar 100% pada data *training* dan *testing* dengan menggunakan metode *Classification Tree* dikarenakan memang data penelitian yang sedikit sehingga mempengaruhi hasil ketepatan klasifikasi.

(Halaman ini sengaja dikosongkan)

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut.

1. Wanita dengan status pemeriksaan positif kanker payudara mayoritas adalah wanita yang sudah menikah, memiliki usia rata-rata 45 tahun yang mana dari ke-9 gejala klinis kanker payudara yaitu 12 dari 13 wanita mengalami benjolan di sekitar payudaranya, mayoritas dari mereka memiliki durasi tidur antara $\geq 5-6$ jam, waktu mulai tidur antara $31-\leq 60$ menit, frekuensi terbangun antara 1-2 kali dan mayoritas dari mereka memiliki riwayat pengobatan operasi, kemoterapi dan herbal. Sedangkan kelompok wanita dengan status pemeriksaan negatif kanker payudara mayoritas adalah wanita yang belum menikah, memiliki rata-rata usia 32 tahun yang mana mayoritas dari mereka tidak mengalami ke-9 gejala klinis, memiliki durasi tidur antara $\geq 5-6$ jam, waktu mulai tidur antara ≤ 15 menit, frekuensi terbangun antara 1-2 kali dan mayoritas dari mereka tidak memiliki riwayat pengobatan operasi, kemoterapi dan herbal.
2. Model pohon klasifikasi optimal yang terbentuk dengan menggunakan metode *Classification Tree* dan *5-fold cross validation* dalam pembagian data *training* dan *testing* adalah memiliki 4 simpul terminal dengan kedalaman 2. Variabel terpenting sebagai penentu status pemeriksaan kanker payudara berdasarkan hasil tes deteksi kanker payudara adalah variabel benjolan di sekitar payudara (X_4), status pernikahan (X_2), dan nyeri di bagian tertentu payudara (X_6). Model optimal tersebut mampu memprediksi data *testing* dengan tingkat *accuracy*, *sensitifity*, *specificity* dan AUC adalah sebesar 95,92%, 84,62%, 100% dan 92,31%. Sedangkan model terbaik yang terbentuk dengan menggunakan metode *Bootstrap Aggregating* adalah dengan 20 replikasi *bootstrap*. Model

tersebut mampu memprediksi data *testing* dengan tingkat *accuracy* dan AUC adalah sebesar 80% dan 88,89%.

3. Perbandingan ketepatan klasifikasi pohon yang diperoleh dengan menggunakan metode *Classification Tree* dan *Bootstrap Aggregating* menunjukkan bahwa metode *Bootstrap Aggregating* tidak lebih baik dari metode *Classification Tree* karena baik untuk data *training* maupun *testing* dengan metode *Classification Tree* menghasilkan ketepatan klasifikasi lebih tinggi dibandingkan dengan metode *Bootstrap Aggregating*. Namun dalam hal ini walaupun hasil dari Baging tidak selalu meningkatkan hasil akurasi dari metode *Classification Tree*, akan tetapi dapat menghasilkan model yang konsisten.

5.2 Saran

Saran yang dapat diberikan untuk penelitian selanjutnya antara lain adalah sebagai berikut.

1. Menambah data lebih banyak lagi disesuaikan dengan metode yang digunakan dan khususnya menambah data wanita dengan status pemeriksaan positif kanker payudara untuk membuat data menjadi *balance* atau dapat dilakukan penanganan terhadap data *imbalance* terlebih dahulu sebelum melakukan tahapan pengolahan lebih lanjut.
2. Melakukan perbandingan dengan metode klasifikasi yang lain seperti metode *random forest* untuk mendapatkan metode terbaik yang dapat digunakan dilihat dari ketepatan klasifikasi modelnya.

DAFTAR PUSTAKA

- Akman, T., Yavuzsen, T., Sevgen, Z., Eallidokuz, H., dan Yilmaz, A. U. (2015). Evaluation of Sleep Disorders in Cancer Patients Based on Pittsburgh Sleep Quality Index. *Eur Journal Cancer Care*, 24(4), 553–559.
- Almutlaq, B. A., Almuazzi, R. F., Almuhayfir, A. A., Alfouzan, A. M., Alshammari, B. T., AlAnzi, H. S., dan Ahmed, H. G. (2017). Breast Cancer in Saudi Arabia and Its Possible Risk Factors. *Journal of Cancer Policy*, 12, 83-89.
- Araújo, D. V. D., Costa, R. H. S., Justino, D. C. P., Batista, F. D. G. A., Andrade, F. B. D., dan Costa, I. D. C. C. (2014). Evaluation of Sleep Quality in Patients with Breast Cancer. *Journal of Sleep Disorders Treat Care*, 3-4.
- Bawazir, A., Bashateh, N., Jradi, H., dan Breik, A. B. (2018). Breast Cancer Screening Awareness and Practices Among Women Attending Primary Health Care Centers in The Ghail Bawazir District of Yemen. *Clinical Breast Cancer*, 19, e20-e29.
- Berrar, D. (2018). Introduction to the Non-Parametric Bootstrap. *Reference Module in Life Sciences*, 1, 766-773.
- Bray, F., Ferlay J., Soerjomataram, I., Siegel R. L., Torre, L. A., dan Jemal, A. (2018). Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *American Cancer Society*, 68, 394–424.
- Breiman, L., Friedman, J. H., Olshen, R. A., dan Stone, C. J. (1993). *Classification and Regression Tree*. New York: Chapman and Hall.
- Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24, 123-140.
- Budhrani, P. H., Lengacher, C. A., Kip, K., Tofthagen, C., dan Jim, H. (2014). An Integrative Review of Subjective and Objective Measures of Sleep Disturbances in Breast Cancer Survivors. *Clin Journal Oncology Nurs*, 19, No.2, 185-191.

- CART Reference Guide. (2000). *CART Users Guide*. San Diego: Salford System.
- Dhruva, A., Paul, S. M., Cooper, B. A., Lee, K., West, C., Aouizerat, B. E., Dunn, L. B., Swift, P. S., Wara, W., dan Miaskowski, C. (2012). A Longitudinal Study of Measures of Objective and Subjective Sleep Disturbance in Patients With Breast Cancer Before, During and After Radiation Therapy. *Journal Pain Symptom Manage*, 44, No.2, 215–228.
- Efron, B. dan Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Ellis, E. O., Schnitt, S. J., S-Garau, X., Bussolati, G., Tavassaoli, F. A., dan Eusebi, V. (2003). *Pathology and Genetic of Tumours of The Breast and Female Genital Organs / WHO Classification of Tumours*. Washington: IARC Press, P.10, 34-6.
- Gorunescu, F. (2011). *Data Mining Concepts, Model and Techniques*. Verlag Berlin Heidelberg: Intelligent Systems Reference Library-Springer, 12.
- Hartwell, L. H., dan Kastan, M. B. (1994). Cell Cycle Control and Cancer. *Science*, 266(5192), 1821–1828.
- Hastie, T., Tibshirani, R., dan Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer-Verlag.
- Hinyard, L., Wirth, L. S., Clancy, J. M., dan Schwartz, T. (2017). The effect of marital status on breast cancer-related outcomes in women under 65: A SEER database analysis. *The Breast*, 32, 13–17.
- Huang, J. dan Ling, C. X., (2005). Using AUC and Accuracy in Evaluating Learning Algorithms. *Transactions on Knowledge and Data Engineering*, 17(3), 299-310.
- International Agency for Research on Cancer (IARC)*. (2016). *Breast Cancer Screening*. Lyon: World Health Organization.

- Istiqomah, I. (2015). Penerapan Pohon Klasifikasi dan *Bootstrap Aggregating* dalam Klasifikasi Usia *Menarche* (Studi Kasus: SMPN Ragunan dan SMPN 1 Dramaga). Tugas Akhir Statistika Institut Pertanian Bogor.
- Johnson, R. A., dan Winchern, D. W. (2007). *Applied Multivariate Statistical Analysis (6th ed)*. New Jersey: Pearson Prentice Hall.
- KEMENKES RI. (2016). *Pusat Data dan Informasi*. Jakarta Selatan: Kementerian Kesehatan RI.
- Lewis, R. J. (2000). *An Introduction To Classification And Regression Trees (CART) Analysis*. California: Department of Emergency Medicine Harbor-UCLA Medical Center.
- Lim, G. C. C. (2002). Overview of cancer in Malaysia. *Japanese Journal Clinical Oncology*, 32(Supplement 1), S37–S42.
- Liu, L., Mills, P. J., Rissling, M., Fiorentino, L., Natarajan, L., Dimsdale, J. E., Sadler, G. R., Parker, B. A., dan Ancoli-Israel, S. (2012). Fatigue and Sleep Quality are Associated with Changes in Inflammatory Markers in Breast Cancer Patients Undergoing Chemotherapy. *Brain Behavior and Immunity*, 26, No.5, 706–713.
- Lowery, A. E. (2014). Insomnia Screening and Treatment in Cancer Care: Current Practices, Barriers and Future Directions, 2, 29–31.
- Lumachi, F., Ermani, M., Brandes, A., Boccagni, P., Polistina, F., Basso, S. M., dan D’Amico, D. (2002). Breast Complaints and Risk of Breast Cancer Population-based study of 2,879 self-selected women and long-term follow-up. *Biomedicine & Pharmacotherapy*, 56, No.2, 88–92.
- Mitchell, T. M., (1997). *Machine Learning*. New York : McGraw-Hill Science.
- Muhammad, M., Merriam, S., dan Suhami, N. (2012). Why Breast Cancer Patients Seek Traditional Healers. *International Journal of Breast Cancer*, 2012, 1-9.
- Muttaqin, M. J. (2015). Metode *Ensamble* pada CART untuk Perbaikan Klasifikasi Kemiskinan di Kabupaten Jombang.

- Tugas Akhir Statistika Institut Teknologi Sepuluh Nopember.
- Rosely, E. (2015). *Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining*. ITS-Surabaya: SESINDO.
- Romito, F., Cormio, C., De Padova, S., Lorusso, V., Berio, M. A., Fimiani, F., dan Mattioli, V. (2014). Patients Attitudes Towards Sleep Disturbances During Chemotherapy. *European Journal of Cancer Care*, 23, no.3, 385–393.
- Seltzer, M.H. (2004). Breast Complaints, Biopsies, and Cancer Correlated with Age in 10,000 Consecutive New Surgical Referrals. *The Breast Journal*, 10(2), 111-117.
- Shaharudin, S. H., Sulaiman, S., Emran, N. A., Shahril, M. R., dan Hussain, S. N. (2011). The Use of Complementary and Alternative Medicine Among Malay Breast Cancer Survivors. *Alternative Therapies Health Medicine*, 17(1), 50-56.
- Siti, Z. M., Tahir, A., Farah, A. I., Fazlin, S. M. A., Sondi, S., Azman, A. H., Maimunah, A. H., Haniza, M. A., Haslinda, M. D. S., Zulkarnain, A. K., Zakiah, I., dan Zaleha, W. C. W. (2009). Use of traditional and complementary medicine in Malaysia: A baseline study. *Complementary Therapies in Medicine*, 17, Issues 5-6, 292–299.
- Sutton, C. D. (2005). Classification and Regression Trees, Bagging, and Boosting. *Handbook of Statistics*, 24, 303-329.
- Tasaki, K., Maskarinec, G., Shumay, D. M., Tatsumura, Y., dan Kakai, H. (2002). Communication Between Physicians and Cancer Patients About Complementary and Alternative Medicine: Exploring Patients' Perspectives. *Psycho-Oncology*, 11, 212–220.
- Walpole, R. E. (1995). *Pengantar Metode Statistika*. Terjemahan oleh Ir. Bambang Sumantri. Edisi Ketiga. Jakarta: PT Gramedia Pustaka Utama.

LAMPIRAN

Lampiran 1. Data Asli Penderita dan Bukan Penderita Kanker Payudara

No	Y	X1	X2	X3	X4	X5	X6	X7	X8
1	Positif	39	Menikah	Ya	Ya	Tidak	Tidak	Tidak	Tidak
2	Positif	32	Menikah	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
3	Positif	38	Menikah	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
4	Positif	38	Menikah	Tidak	Ya	Ya	Tidak	Tidak	Tidak
5	Positif	42	Menikah	Tidak	Ya	Ya	Ya	Tidak	Ya
6	Positif	52	Menikah	Ya	Ya	Ya	Ya	Ya	Ya
7	Positif	42	Menikah	Ya	Ya	Ya	Ya	Ya	Tidak
8	Positif	54	Menikah	Ya	Ya	Tidak	Ya	Ya	Tidak
9	Positif	37	Menikah	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
10	Positif	57	Menikah	Tidak	Ya	Tidak	Tidak	Tidak	Tidak
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
45	Negatif	17	Belum Menikah	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
46	Negatif	16	Belum Menikah	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
47	Negatif	17	Belum Menikah	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
48	Negatif	41	Menikah	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
49	Negatif	22	Belum Menikah	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak

X9	X10	X11	X12	X13	X14	X15	X16	X17
Tidak	Tidak	Tidak	$\geq 5-6$	$31-\leq 60$	1-2 kali	Ya	Ya	Ya
Tidak	Tidak	Tidak	$\geq 5-6$	$31-\leq 60$	1-2 kali	Ya	Ya	Tidak
Tidak	Tidak	Tidak	$\geq 5-6$	$31-\leq 60$	3-4 kali	Ya	Ya	Tidak
Tidak	Tidak	Tidak	$\geq 6-7$	≤ 15	1-2 kali	Ya	Ya	Tidak
Ya	Tidak	Tidak	$\geq 5-6$	$31-\leq 60$	1-2 kali	Ya	Ya	Ya
Ya	Ya	Ya	< 5	≤ 15	> 5 kali	Ya	Ya	Tidak
Ya	Ya	Ya	$\geq 5-6$	> 60	3-4 kali	Ya	Ya	Ya
Tidak	Tidak	Ya	< 5	$31-\leq 60$	1-2 kali	Ya	Ya	Ya
Tidak	Tidak	Tidak	< 5	> 60	1-2 kali	Tidak	Tidak	Tidak
Tidak	Tidak	Tidak	$\geq 6-7$	$31-\leq 60$	Tidak pernah	Tidak	Tidak	Ya
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
Tidak	Tidak	Tidak	$\geq 5-6$	$16-\leq 30$	1-2 kali	Tidak	Tidak	Tidak
Tidak	Tidak	Tidak	$\geq 5-6$	≤ 15	1-2 kali	Tidak	Tidak	Tidak
Tidak	Tidak	Tidak	$\geq 6-7$	≤ 15	1-2 kali	Tidak	Tidak	Tidak
Tidak	Tidak	Tidak	$\geq 5-6$	$16-\leq 30$	1-2 kali	Ya	Tidak	Tidak
Tidak	Tidak	Tidak	$\geq 6-7$	≤ 15	1-2 kali	Tidak	Tidak	Tidak

Lampiran 2. Data Hasil Pengkodean

No	Y	X1	X2	X3	X4	X5	X6	X7	X8
1	1	0	1	1	1	0	0	0	0
2	1	0	1	0	1	0	0	0	0
3	1	0	1	0	1	0	0	0	0
4	1	0	1	0	1	1	0	0	0
5	1	0	1	0	1	1	1	0	1
6	1	1	1	1	1	1	1	1	1
7	1	0	1	1	1	1	1	1	0
8	1	1	1	1	1	0	1	1	0
9	1	0	1	0	1	0	0	0	0
10	1	1	1	0	1	0	0	0	0
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
45	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0
48	0	0	1	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0

X9	X10	X11	X12	X13	X14	X15	X16	X17
0	0	0	1	1	2	1	1	1
0	0	0	1	1	2	1	1	0
0	0	0	1	1	1	1	1	0
0	0	0	2	3	2	1	1	0
1	0	0	1	1	2	1	1	1
1	1	1	0	3	0	1	1	0
1	1	1	1	0	1	1	1	1
0	0	1	0	1	2	1	1	1
0	0	0	0	0	2	0	0	0
0	0	0	2	1	3	0	0	1
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
0	0	0	1	2	2	0	0	0
0	0	0	1	3	2	0	0	0
0	0	0	2	3	2	0	0	0
0	0	0	1	2	2	1	0	0
0	0	0	2	3	2	0	0	0

Lampiran 3. Crosstab Variabel Penelitian
Output 3.1 Variabel Usia

Crosstab

		X1		Total	
		0	1		
Y	0	Count	29	7	36
		Expected Count	27.2	8.8	36.0
	1	Count	8	5	13
		Expected Count	9.8	3.2	13.0
Total		Count	37	12	49
		Expected Count	37.0	12.0	49.0

Output 3.2 Variabel Status Pernikahan

Crosstab

		X2		Total	
		1	2		
Y	0	Count	23	13	36
		Expected Count	16.9	19.1	36.0
	1	Count	0	13	13
		Expected Count	6.1	6.9	13.0
Total		Count	23	26	49
		Expected Count	23.0	26.0	49.0

Output 3.3 Variabel Perubahan Bentuk dan Ukuran Payudara

Crosstab

		X3		Total	
		1	2		
Y	0	Count	3	33	36
		Expected Count	5.9	30.1	36.0
	1	Count	5	8	13
		Expected Count	2.1	10.9	13.0
Total		Count	8	41	49
		Expected Count	8.0	41.0	49.0

Output 3.4 Variabel Benjolan Disekitar Payudara

Crosstab

		X4		Total	
		1	2		
Y_1	0	Count	3	33	36
		Expected Count	11.0	25.0	36.0
	1	Count	12	1	13
		Expected Count	4.0	9.0	13.0
Total		Count	15	34	49
		Expected Count	15.0	34.0	49.0

Output 3.5 Variabel Kerutan Disekitar Payudara**Crosstab**

		X5		Total	
		1	2		
Y	0	Count	2	34	36
		Expected Count	5.1	30.9	36.0
	1	Count	5	8	13
		Expected Count	1.9	11.1	13.0
Total		Count	7	42	49
		Expected Count	7.0	42.0	49.0

Output 3.6 Variabel Nyeri Dibagian Tertentu Payudara**Crosstab**

		X6		Total	
		1	2		
Y	0	Count	3	33	36
		Expected Count	6.6	29.4	36.0
	1	Count	6	7	13
		Expected Count	2.4	10.6	13.0
Total		Count	9	40	49
		Expected Count	9.0	40.0	49.0

Output 3.7 Variabel Payudara Nampak Kemerahan dan Bengkak**Crosstab**

		X7		Total	
		1	2		
Y	0	Count	1	35	36
		Expected Count	3.7	32.3	36.0
	1	Count	4	9	13
		Expected Count	1.3	11.7	13.0
Total		Count	5	44	49
		Expected Count	5.0	44.0	49.0

Output 3.8 Variabel Keluar Cairan dengan Bercak Darah atau Encer**Crosstab**

		X8		Total	
		1	2		
Y	0	Count	1	35	36
		Expected Count	2.2	33.8	36.0
	1	Count	2	11	13
		Expected Count	.8	12.2	13.0
Total		Count	3	46	49
		Expected Count	3.0	46.0	49.0

Output 3.9 Variabel Penarikan Puting Susu Kedalam**Crosstab**

		X9		Total	
		1	2		
Y	0	Count	2	34	36
		Expected Count	5.1	30.9	36.0
	1	Count	5	8	13
		Expected Count	1.9	11.1	13.0
Total		Count	7	42	49
		Expected Count	7.0	42.0	49.0

Output 3.10 Variabel Benjolan Didaerah Ketiak**Crosstab**

		X10		Total	
		1	2		
Y	0	Count	2	34	36
		Expected Count	3.7	32.3	36.0
	1	Count	3	10	13
		Expected Count	1.3	11.7	13.0
Total		Count	5	44	49
		Expected Count	5.0	44.0	49.0

Output 3.11 Variabel Cekungan Pada Kulit Payudara**Crosstab**

		X11		Total	
		1	2		
Y	0	Count	1	35	36
		Expected Count	3.7	32.3	36.0
	1	Count	4	9	13
		Expected Count	1.3	11.7	13.0
Total		Count	5	44	49
		Expected Count	5.0	44.0	49.0

Output 3.12 Variabel Durasi Tidur**Crosstab**

		X12				Total	
		1	2	3	4		
Y	0	Count	3	13	17	3	36
		Expected Count	2.9	11.8	16.9	4.4	36.0
	1	Count	1	3	6	3	13
		Expected Count	1.1	4.2	6.1	1.6	13.0
Total		Count	4	16	23	6	49
		Expected Count	4.0	16.0	23.0	6.0	49.0

Output 3.13 Variabel Latensi Tidur**Crosstab**

		X13				Total	
		1	2	3	4		
Y	0	Count	5	10	5	16	36
		Expected Count	5.1	13.2	3.7	14.0	36.0
	1	Count	2	8	0	3	13
		Expected Count	1.9	4.8	1.3	5.0	13.0
Total		Count	7	18	5	19	49
		Expected Count	7.0	18.0	5.0	19.0	49.0

Output 3.14 Variabel Frekuensi Terbangun**Crosstab**

		X14				Total	
		1	2	3	4		
Y	0	Count	0	7	17	12	36
		Expected Count	.7	6.6	19.1	9.6	36.0
	1	Count	1	2	9	1	13
		Expected Count	.3	2.4	6.9	3.4	13.0
Total		Count	1	9	26	13	49
		Expected Count	1.0	9.0	26.0	13.0	49.0

Output 3.15 Variabel Riwayat Pengobatan Operasi**Crosstab**

		X15		Total	
		1	2		
Y	0	Count	6	30	36
		Expected Count	12.5	23.5	36.0
	1	Count	11	2	13
		Expected Count	4.5	8.5	13.0
Total		Count	17	32	49
		Expected Count	17.0	32.0	49.0

Output 3.16 Variabel Riwayat Pengobatan Kemoterapi**Crosstab**

		X16		Total	
		1	2		
Y	0	Count	0	36	36
		Expected Count	7.3	28.7	36.0
	1	Count	10	3	13
		Expected Count	2.7	10.3	13.0
Total		Count	10	39	49
		Expected Count	10.0	39.0	49.0

Output 3.17 Variabel Riwayat Pengobatan Herbal Crosstab

		X17		Total	
		1	2		
Y	0	Count	5	31	36
		Expected Count	9.6	26.4	36.0
	1	Count	8	5	13
		Expected Count	3.4	9.6	13.0
Total	Count	13	36	49	
	Expected Count	13.0	36.0	49.0	

Lampiran 4. Informasi Pemilahan Simpul Pohon Klasifikasi Optimal

```

=====
NODE INFORMATION
=====
          *****
          *                Node 1: BENJOLAN                *
          *                N: 49                          *
          *****

          *****
          *                Node 2                *          *                Node 3                *
          *                N: 34                *          *                N: 15                *
          *****

Node 1 was split on BENJOLAN
A case goes left if BENJOLAN = (0)
Improvement = 0.353          Complexity Threshold = 0.420

Node      Cases  Wgt Counts      Cost Class
  1         49     49.00      0.500 NEGATIF
  2         34     34.00      0.077 NEGATIF
  3         15     15.00      0.083 POSITIF

          Weighted Counts
Class      Top      Left      Right
NEGATIF    36.00    33.00     3.00
POSITIF    13.00     1.00    12.00

          Within Node Probabilities
Class      Top      Left      Right
NEGATIF    0.500    0.923    0.083
POSITIF    0.500    0.077    0.917

          Surrogate          Split          Assoc.          Improve.
  1 RIWAYATO s 0          Split          0.761          0.313
  2 RIWAYAT_ s 0          Split          0.516          0.231
  3 STATUS_P s 0          Split          0.475          0.235
  4 RIWAYATI s 0          Split          0.424          0.121
  5 NYERI_BA s 0          Split          0.381          0.090

          Competitor          Split          Improve.
  1 RIWAYATO 0          Split          0.313
  2 STATUS_P 0          Split          0.235
  3 RIWAYAT_ 0          Split          0.231

```

```

4 RIWAYAT1      0      0.121
5 NYERI_BA      0      0.090

*****
*           Node 2: NYERI_BA           *
*           N: 34                       *
*****

=====
=           Terminal Node 1           =
=           N: 33                     =
=====

=====
=           Terminal Node 2           =
=           N: 1                      =
=====

Node 2 was split on NYERI_BA
A case goes left if NYERI_BA = (0)
Improvement = 0.071      Complexity Threshold = 0.038

Node      Cases  Wgt Counts      Cost Class
  2         34      34.00      0.077 NEGATIF
 -1         33      33.00      0.000 NEGATIF
 -2          1       1.00      0.000 POSITIF

                Weighted Counts
Class      Top      Left      Right
NEGATIF    33.00    33.00     0.00
POSITIF     1.00     0.00     1.00

                Within Node Probabilities
Class      Top      Left      Right
NEGATIF    0.923    1.000     0.000
POSITIF     0.077    0.000     1.000

Competitor      Split      Improve.
1 RIWAYAT1      0      0.031
2 RIWAYAT      0      0.018
3 USIA_45      0      0.016
4 LATENSI      0,2,3    0.012
5 STATUS_P      0      0.008

*****
*           Node 3: STATUS_P           *
*           N: 15                       *
*****

=====
=           Terminal Node 3           =
=           N: 3                      =
=====

=====
=           Terminal Node 4           =
=           N: 12                     =
=====

Node 3 was split on STATUS_P
A case goes left if STATUS_P = (0)
Improvement = 0.076      Complexity Threshold = 0.042

Node      Cases  Wgt Counts      Cost Class
  3         15      15.00      0.083 POSITIF
 -3          3       3.00      0.000 NEGATIF
 -4         12      12.00      0.000 POSITIF

                Weighted Counts
Class      Top      Left      Right
NEGATIF    3.00     3.00     0.00
POSITIF    12.00    0.00    12.00

```

		Within Node Probabilities				
Class		Top	Left	Right		
NEGATIF		0.083	1.000	0.000		
POSITIF		0.917	0.000	1.000		
Surrogate		Split			Assoc.	Improve.
1	LATENSI_ s 2				0.333	0.024
Competitor		Split				Improve.
1	LATENSI_ 2					0.024
2	RIWAYAT_ 0					0.022
3	RIWAYAT0 0					0.022
4	NYERI_BA 1					0.008
5	FREKUENS 1,3					0.005

Lampiran 5. Informasi Proporsi Setiap Kelas Pada Setiap Simpul Pohon Klasifikasi Optimal

```

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

[Breiman adjusted cost, lambda = 0.399]

```

Node	Class	Wgt	Count	N	Prob	Cost	Parent Complexity
1	NEGATIF	33.00	33	33	0.458	0.000	0.038 [0.143]
	NEGATIF	33.00	33	33	1.000		
	POSITIF	0.00	0	0	0.000		
2	POSITIF	1.00	1	1	0.038	0.000	0.038 [0.281]
	NEGATIF	0.00	0	0	0.000		
	POSITIF	1.00	1	1	1.000		
3	NEGATIF	3.00	3	3	0.042	0.000	0.042 [0.279]
	NEGATIF	3.00	3	3	1.000		
	POSITIF	0.00	0	0	0.000		
4	POSITIF	12.00	12	12	0.462	0.000	0.042 [0.143]
	NEGATIF	0.00	0	0	0.000		
	POSITIF	12.00	12	12	1.000		

Lampiran 6. Informasi Pemilahan Simpul Pohon Klasifikasi Menggunakan Metode *Bootstrap Aggregating* Untuk 10 Replikasi

```

=====
INITIAL BUILD RUN TO PRECEDE 10 BOOTSTRAP CYCLES
=====
39 Observations in the learning sample.
Tree constructed with complexity parameter = 0.000

```

```

=====
TREE SEQUENCE
=====
Terminal Resubstitution Complexity
Tree   Nodes Relative Cost   Parameter
-----
  1     4       0.000       0.000
  2     3       0.037       0.019
  3     2       0.120       0.042
  4     1       1.000       0.440

Initial misclassification cost = 0.500
Initial class assignment = NEGATIF

=====
CLASSIFICATION TREE DIAGRAM
=====

          |
          |-----1-----|
          |                 |
    -----2-----|         |-----3-----|
    |                 |         |                 |
    |                 |         |                 |

Terminal Regions
1                 2                 3                 4

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

*****
*           Node 1: BENJOLAN           *
*           N: 39                       *
*****

*****
*           Node 2           *           *           Node 3           *
*           N: 27           *           *           N: 12           *
*****

Node 1 was split on BENJOLAN
A case goes left if BENJOLAN = (0)
Improvement = 0.388           Complexity Threshold = 0.440

Node   Cases   Wgt Counts   Cost   Class
  1       39       39.00       0.500 NEGATIF
  2       27       27.00       0.080 NEGATIF
  3       12       12.00       0.039 POSITIF

Weighted Counts
Class   Top      Left      Right
NEGATIF 27.00     26.00     1.00
POSITIF 12.00      1.00     11.00

Within Node Probabilities
Class   Top      Left      Right
NEGATIF 0.500     0.920     0.039
POSITIF 0.500     0.080     0.961

Surrogate      Split      Assoc.      Improve.
  1 RIWAYAT0 s 0      0.786      0.300
  2 RIWAYAT_ s 0      0.583      0.262
  3 STATUS_P s 0      0.563      0.271
  4 KERUTAN s 0      0.398      0.103
  5 RIWAYAT1 s 0      0.398      0.123

```


Competitor	Split	Improve.
1 RIWAYAT0	0	0.300
2 STATUS_P	0	0.271
3 RIWAYAT_	0	0.262
4 NYERI_BA	0	0.136
5 RIWAYAT1	0	0.123

* Node 2: NYERI_BA *
* N: 27 *

Terminal Node 1	Terminal Node 2
N: 26	N: 1

Node 2 was split on NYERI_BA
A case goes left if NYERI_BA = (0)
Improvement = 0.077 Complexity Threshold = 0.042

Node	Cases	Wgt	Counts	Cost	Class
2	27		27.00	0.080	NEGATIF
-1	26		26.00	0.000	NEGATIF
-2	1		1.00	0.000	POSITIF

Weighted Counts

Class	Top	Left	Right
NEGATIF	26.00	26.00	0.00
POSITIF	1.00	0.00	1.00

Within Node Probabilities

Class	Top	Left	Right
NEGATIF	0.920	1.000	0.000
POSITIF	0.080	0.000	1.000

Surrogate	Split	Assoc.	Improve.
1 RIWAYAT1 s 0		0.111	0.037

Competitor	Split	Improve.
1 RIWAYAT1	0	0.037
2 RIWAYAT_	0	0.029
3 USIA_45	0	0.023
4 LATENSI	0,2,3	0.019
5 STATUS_P	0	0.012

* Node 3: STATUS_P *
* N: 12 *

Terminal Node 3	Terminal Node 4
N: 1	N: 11

Node 3 was split on STATUS_P
A case goes left if STATUS_P = (0)
Improvement = 0.036 Complexity Threshold = 0.019

Node	Cases	Wgt	Counts	Cost	Class
3	12		12.00	0.039	POSITIF
-3	1		1.00	0.000	NEGATIF
-4	11		11.00	0.000	POSITIF

Weighted Counts			
Class	Top	Left	Right
NEGATIF	1.00	1.00	0.00
POSITIF	11.00	0.00	11.00

Within Node Probabilities			
Class	Top	Left	Right
NEGATIF	0.039	1.000	0.000
POSITIF	0.961	0.000	1.000

Competitor	Split	Improve.
1 LATENSI_	0,2	0.005
2 RIWAYAT_	0	0.005
3 RIWAYAT0	0	0.005
4 NYERI_BA	1	0.002
5 DURASI_T	1,3	0.002

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	27.00	0.00	27.00
POSITIF	0.00	12.00	12.00

PRED. TOT.	27.00	12.00	39.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.308	0.692	
TOT. CORRECT	1.000		

SENSITIVITY:	1.000	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.000
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

VARIABLE IMPORTANCE

=====

	Relative Importance	Number Of Categories	Minimum Category
BENJOLAN	100.000	2	0
STATUS_P	79.191	2	0
RIWAYAT0	77.378	2	0
RIWAYAT_	67.477	2	0
RIWAYAT1	41.387	2	0
KERUTAN_	26.493	2	0
NYERI_BA	19.782	2	0

=====

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	9.00	0.00	9.00
POSITIF	0.00	1.00	1.00

PRED. TOT.	9.00	1.00	10.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.100	0.900	
TOT. CORRECT	1.000		

SENSITIVITY:	1.000	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.000

```
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"
-----

=====
BOOTSTRAP BUILD 1 OF 10
=====
32 Observations in the learning sample.
Tree constructed with complexity parameter = 0.000

=====
TREE SEQUENCE
=====
      Terminal Resubstitution Complexity
Tree  Nodes Relative Cost  Parameter
-----
1      3       0.000      0.000
2      2       0.091      0.045
3      1       1.000      0.455

Initial misclassification cost = 0.500
Initial class assignment = NEGATIF

=====
CLASSIFICATION TREE DIAGRAM
=====
-----1-----
|
|
|
|
|
|
|
|
|
|
|-----2-----|
|
|
|
|
|
|
|
|
|
|
|

Terminal Regions

1                    2                    3

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

*****
*           Node 1: STATUS_P           *
*           N: 32                       *
*****

=====
*           *
= Terminal Node 1 = * Node 2 *
= N: 20           = * N: 12       *
=====
*           *

Node 1 was split on STATUS_P
A case goes left if STATUS_P = (0)
Improvement = 0.417      Complexity Threshold = 0.455

Node      Cases  Wgt Counts      Cost Class
  1         32    32.00      0.500 NEGATIF
 -1         20    20.00      0.000 NEGATIF
  2         12    12.00      0.083 POSITIF

      Weighted Counts
Class     Top      Left      Right
NEGATIF  22.00    20.00    2.00
POSITIF  10.00     0.00    10.00

      Within Node Probabilities
Class     Top      Left      Right
NEGATIF  0.500    1.000    0.083
```

POSITIF	0.500	0.000	0.917		
Surrogate		Split		Assoc.	Improve.
1 USIA_45	s 0			0.840	0.366
2 BENJOLAN	s 0			0.800	0.417
3 CEKUNGAN	s 0			0.680	0.333
4 RIWAYAT	s 0			0.680	0.254
5 PAYUDARA	s 0			0.680	0.333
Competitor		Split			Improve.
1 BENJOLAN	0				0.417
2 USIA_45	0				0.366
3 PAYUDARA	0				0.333
4 CEKUNGAN	0				0.333
5 RIWAYAT0	0				0.333

* Node 2: BENJOLAN *					
* N: 12 *					

=====			=====		
=	Terminal Node 2	=	=	Terminal Node 3	=
=	N: 2	=	=	N: 10	=
=====			=====		
Node 2 was split on BENJOLAN					
A case goes left if BENJOLAN = (0)					
Improvement = 0.083 Complexity Threshold = 0.045					
Node	Cases	Wgt	Counts	Cost	Class
2	12		12.00	0.083	POSITIF
-2	2		2.00	0.000	NEGATIF
-3	10		10.00	0.000	POSITIF
Weighted Counts					
Class	Top		Left		Right
NEGATIF	2.00		2.00		0.00
POSITIF	10.00		0.00		10.00
Within Node Probabilities					
Class	Top		Left		Right
NEGATIF	0.083		1.000		0.000
POSITIF	0.917		0.000		1.000
Competitor		Split			Improve.
1 FREKUENS	1				0.036
2 NYERI_BA	0				0.021
3 PAYUDARA	0				0.021
4 CEKUNGAN	0				0.021
5 RIWAYAT0	0				0.021
=====					
LEARNING SAMPLE CLASSIFICATION TABLE					
=====					
Actual Class	Predicted Class			Actual Total	
	NEGATIF	POSITIF			
-----	-----			-----	
NEGATIF	22.00	0.00		22.00	
POSITIF	0.00	10.00		10.00	
-----	-----			-----	
PRED. TOT.	22.00	10.00		32.00	
CORRECT	1.000	1.000			
SUCCESS IND.	0.313	0.688			
TOT. CORRECT	1.000				

SENSITIVITY: 1.000 SPECIFICITY: 1.000
 FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.000
 REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

=====
 CASE-BY-CASE SAMPLE CLASSIFICATION TABLE
 =====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	9.00	0.00	9.00
POSITIF	0.00	1.00	1.00
PRED. TOT.	9.00	1.00	10.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.100	0.900	
TOT. CORRECT	1.000		

SENSITIVITY: 1.000 SPECIFICITY: 1.000
 FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.000
 REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

=====
 BOOTSTRAP BUILD 2 OF 10
 =====

36 Observations in the learning sample.
 Tree constructed with complexity parameter = 0.000

=====
 TREE SEQUENCE
 =====

Tree	Terminal Nodes	Resubstitution Relative Cost	Complexity Parameter
1	3	0.000	0.000
2	2	0.133	0.067
3	1	1.000	0.433

Initial misclassification cost = 0.500
 Initial class assignment = NEGATIF

=====
 CLASSIFICATION TREE DIAGRAM
 =====



Terminal Regions

1 2 3

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

```

*****
*           Node 1: NYERI_BA           *
*           N: 36                       *
*****
  
```

```

*****
*           Node 2           *
*           N: 23           *
*****
=====
=           Terminal Node 3   =
=           N: 13           =
=====

```

Node 1 was split on NYERI_BA
A case goes left if NYERI_BA = (0)
Improvement = 0.382 Complexity Threshold = 0.433

Node	Cases	Wgt	Counts	Cost	Class
1	36		36.00	0.500	NEGATIF
2	23		23.00	0.118	NEGATIF
-3	13		13.00	0.000	POSITIF

Weighted Counts			
Class	Top	Left	Right
NEGATIF	21.00	21.00	0.00
POSITIF	15.00	2.00	13.00

Within Node Probabilities			
Class	Top	Left	Right
NEGATIF	0.500	0.882	0.000
POSITIF	0.500	0.118	1.000

Surrogate	Split	Assoc.	Improve.
1 RIWAYAT_ s 0		0.890	0.298
2 USIA 45 s 0		0.769	0.333
3 RIWAYAT0 s 0		0.692	0.214
4 CEKUNGAN s 0		0.692	0.214
5 PAYUDARA s 0		0.692	0.214

Competitor	Split	Improve.
1 STATUS_P 0		0.375
2 USIA 45 0		0.333
3 RIWAYAT_ 0		0.298
4 BENJOLAN 0		0.289
5 PAYUDARA 0		0.214


```

*****
*           Node 2: BENJOLAN   *
*           N: 23           *
*****
=====
=           Terminal Node 1   =
=           N: 21           =
=====
=====
=           Terminal Node 2   =
=           N: 2           =
=====

```

Node 2 was split on BENJOLAN
A case goes left if BENJOLAN = (0)
Improvement = 0.118 Complexity Threshold = 0.067

Node	Cases	Wgt	Counts	Cost	Class
2	23		23.00	0.118	NEGATIF
-1	21		21.00	0.000	NEGATIF
-2	2		2.00	0.000	POSITIF

Weighted Counts			
Class	Top	Left	Right
NEGATIF	21.00	21.00	0.00
POSITIF	2.00	0.00	2.00

Within Node Probabilities			
Class	Top	Left	Right
NEGATIF	0.882	1.000	0.000

POSITIF 0.118 0.000 1.000

Surrogate	Split	Assoc.	Improve.
1 USIA_45 s 0		0.500	0.055
2 DURASI_T s 1,2,3		0.143	0.028
3 LATENSI_ s 1,2,3		0.143	0.028
Competitor	Split		Improve.
1 USIA_45 0			0.055
2 STATUS_P 0			0.049
3 DURASI_T 1,2,3			0.028
4 LATENSI_ 1,2,3			0.028
5 RIWAYATI 0			0.016

=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	21.00	0.00	21.00
POSITIF	0.00	15.00	15.00
PRED. TOT.	21.00	15.00	36.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.417	0.583	
TOT. CORRECT	1.000		

SENSITIVITY: 1.000 SPECIFICITY: 1.000
FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.000
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

=====
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE
=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00
PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		

SENSITIVITY: 0.778 SPECIFICITY: 1.000
FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.667
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

=====
BOOTSTRAP BUILD 3 OF 10
=====

34 Observations in the learning sample.
Tree constructed with complexity parameter = 0.000

=====
TREE SEQUENCE
=====

```

Terminal Resubstitution Complexity
Tree  Nodes Relative Cost  Parameter
-----
1      2      0.074      0.000
2      1      1.000      0.463

Initial misclassification cost = 0.500
Initial class assignment = NEGATIF

=====
CLASSIFICATION TREE DIAGRAM
=====

-----|-----
|-----1-----|

Terminal Regions

1-----2

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

*****
*           Node 1: RIWAYAT1           *
*           N: 34                       *
*****

=====
=           Terminal Node 1           =           Terminal Node 2           =
=           N: 25                       =           N: 9                       =
=====

Node 1 was split on RIWAYAT1
A case goes left if RIWAYAT1 = (0)
Improvement = 0.431           Complexity Threshold = 0.463

Node      Cases  Wgt Counts      Cost  Class
-----
1         34     34.00           0.500 NEGATIF
-1        25     25.00           0.000 NEGATIF
-2         9      9.00           0.069 POSITIF

Class      Top      Left      Right
NEGATIF    27.00    25.00     2.00
POSITIF     7.00     0.00     7.00

Class      Top      Left      Right
NEGATIF    0.500    1.000    0.069
POSITIF    0.500    0.000    0.931

Surrogate      Split      Assoc.      Improve.
1 NYERI_BA    s 0      0.766      0.375
2 RIWAYAT_    s 0      0.726      0.279
3 LATENSI    s 0,2,3  0.686      0.203
4 STATUS_P    s 0      0.640      0.250
5 USIA_45    s 0      0.606      0.251

Competitor      Split      Improve.
1 NYERI_BA      0      0.375
2 RIWAYAT_      0      0.279
3 USIA_45       0      0.251
4 STATUS_P      0      0.250

```


Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	25.00	2.00	27.00
POSITIF	0.00	7.00	7.00
PRED. TOT.	25.00	9.00	34.00
CORRECT	0.926	1.000	
SUCCESS IND.	0.132	0.794	
TOT. CORRECT	0.941		
SENSITIVITY:		0.926	SPECIFICITY: 1.000
FALSE REFERENCE:		0.000	FALSE RESPONSE: 0.222
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00
PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		
SENSITIVITY:		0.778	SPECIFICITY: 1.000
FALSE REFERENCE:		0.000	FALSE RESPONSE: 0.667
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

BOOTSTRAP BUILD 4 OF 10

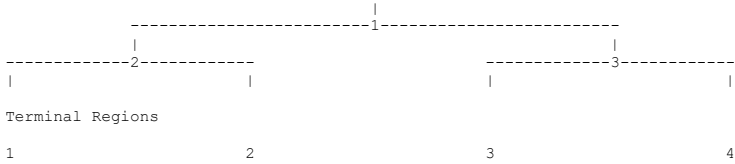
37 Observations in the learning sample.
Tree constructed with complexity parameter = 0.000

TREE SEQUENCE

Tree	Nodes	Relative Cost	Parameter
1	4	0.000	0.000
2	3	0.037	0.019
3	2	0.137	0.050
4	1	1.000	0.431

Initial misclassification cost = 0.500
Initial class assignment = NEGATIF

=====
 CLASSIFICATION TREE DIAGRAM
 =====



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

```

*****
*           Node 1: NYERI_BA           *
*           N: 37                       *
*****

*****
*           Node 2           *           *           Node 3           *
*           N: 27           *           *           N: 10           *
*****
  
```

Node 1 was split on NYERI_BA
 A case goes left if NYERI_BA = (0)
 Improvement = 0.374 Complexity Threshold = 0.431

Node	Cases	Wgt	Counts	Cost	Class
1	37		37.00	0.500	NEGATIF
2	27		27.00	0.094	NEGATIF
3	10		10.00	0.040	POSITIF

Class	Weighted Counts		
	Top	Left	Right
NEGATIF	27.00	26.00	1.00
POSITIF	10.00	1.00	9.00

Class	Within Node Probabilities		
	Top	Left	Right
NEGATIF	0.500	0.906	0.040
POSITIF	0.500	0.094	0.960

Surrogate	Split	Assoc.	Improve.
1 RIWAYAT_ s 0		0.802	0.283
2 RIWAYAT0 s 0		0.747	0.269
3 BENJOLAN s 0		0.680	0.299
4 RIWAYAT1 s 0		0.668	0.283
5 STATUS_P s 0		0.538	0.271

Competitor	Split	Improve.
1 BENJOLAN 0		0.299
2 RIWAYAT_ 0		0.283
3 RIWAYAT1 0		0.283
4 STATUS_P 0		0.271
5 RIWAYAT0 0		0.269

```

*****
*           Node 2: BENJOLAN           *
*           N: 27                       *
*****
  
```

```

=====
=          Terminal Node 1          =          Terminal Node 2          =
=          N: 26                    =          N: 1                    =
=====

```

Node 2 was split on BENJOLAN
A case goes left if BENJOLAN = (0)
Improvement = 0.091 Complexity Threshold = 0.050

Node	Cases	Wgt Counts	Cost	Class
2	27	27.00	0.094	NEGATIF
-1	26	26.00	0.000	NEGATIF
-2	1	1.00	0.000	POSITIF

Class	Weighted Counts		
	Top	Left	Right
NEGATIF	26.00	26.00	0.00
POSITIF	1.00	0.00	1.00

Class	Within Node Probabilities		
	Top	Left	Right
NEGATIF	0.906	1.000	0.000
POSITIF	0.094	0.000	1.000

Competitor	Split	Improve.
1 RIWAYATI	0	0.038
2 USIA 45	0	0.031
3 FREKUENS	0,1,2	0.026
4 STATUS_P	0	0.016
5 LATENSI_	0,2,3	0.016


```

*****
*          Node 3: STATUS_P          *
*          N: 10                    *
*****

```

```

=====
=          Terminal Node 3          =          Terminal Node 4          =
=          N: 1                    =          N: 9                    =
=====

```

Node 3 was split on STATUS_P
A case goes left if STATUS_P = (0)
Improvement = 0.036 Complexity Threshold = 0.019

Node	Cases	Wgt Counts	Cost	Class
3	10	10.00	0.040	POSITIF
-3	1	1.00	0.000	NEGATIF
-4	9	9.00	0.000	POSITIF

Class	Weighted Counts		
	Top	Left	Right
NEGATIF	1.00	1.00	0.00
POSITIF	9.00	0.00	9.00

Class	Within Node Probabilities		
	Top	Left	Right
NEGATIF	0.040	1.000	0.000
POSITIF	0.960	0.000	1.000

Surrogate	Split	Assoc.	Improve.
1 RIWAYAT_ s 0		1.000	0.036

Competitor	Split	Improve.
1 RIWAYAT	0	0.036
2 LATENSI	0,2	0.004
3 RIWAYAT0	0	0.004
4 KERUTAN	0	0.002
5 PAYUDARA	0	0.002


```

=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====

```

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	27.00	0.00	27.00
POSITIF	0.00	10.00	10.00

PRED. TOT.	27.00	10.00	37.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.270	0.730	
TOT. CORRECT	1.000		


```

SENSITIVITY:          1.000    SPECIFICITY:          1.000
FALSE REFERENCE:      0.000    FALSE RESPONSE:       0.000
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"
-----

```

```

=====
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE
=====

```

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	9.00	0.00	9.00
POSITIF	0.00	1.00	1.00

PRED. TOT.	9.00	1.00	10.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.100	0.900	
TOT. CORRECT	1.000		


```

SENSITIVITY:          1.000    SPECIFICITY:          1.000
FALSE REFERENCE:      0.000    FALSE RESPONSE:       0.000
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"
-----

```

```

=====
BOOTSTRAP BUILD 5 OF 10
=====
36 Observations in the learning sample.
Tree constructed with complexity parameter = 0.000

=====
TREE SEQUENCE
=====

```

Tree	Nodes	Relative Cost	Complexity Parameter
1	2	0.000	0.000
2	1	1.000	0.500


```

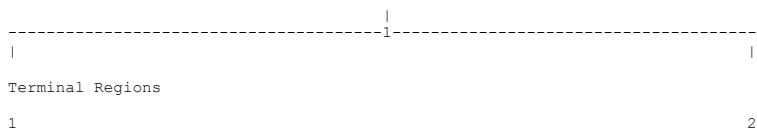
Initial misclassification cost = 0.500
Initial class assignment = NEGATIF

```

```

=====
CLASSIFICATION TREE DIAGRAM
=====

```



```

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

```

```

*****
*           Node 1: BENJOLAN           *
*           N: 36                       *
*****

```

```

=====
=           Terminal Node 1           =           Terminal Node 2           =
=           N: 29                     =           N: 7                       =
=====

```

```

Node 1 was split on BENJOLAN
A case goes left if BENJOLAN = (0)
Improvement = 0.500           Complexity Threshold = 0.500

```

Node	Cases	Wgt	Counts	Cost	Class
1	36		36.00	0.500	NEGATIF
-1	29		29.00	0.000	NEGATIF
-2	7		7.00	0.000	POSITIF

Weighted Counts			
Class	Top	Left	Right
NEGATIF	29.00	29.00	0.00
POSITIF	7.00	0.00	7.00

Within Node Probabilities			
Class	Top	Left	Right
NEGATIF	0.500	1.000	0.000
POSITIF	0.500	0.000	1.000

Surrogate	Split	Assoc.	Improve.
1 NYERI_BA s 0		0.857	0.375
2 RIWAYATO s 0		0.857	0.375
3 RIWAYATI s 0		0.754	0.284
4 RIWAYAT_ s 0		0.754	0.284
5 STATUS_P s 0		0.655	0.244

Competitor	Split	Improve.
1 NYERI_BA 0		0.375
2 RIWAYATO 0		0.375
3 RIWAYAT_ 0		0.284
4 RIWAYATI 0		0.284
5 STATUS_P 0		0.244

```

=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====

```

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	29.00	0.00	29.00
POSITIF	0.00	7.00	7.00

PRED. TOT.	29.00	7.00	36.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.194	0.806	
TOT. CORRECT	1.000		
SENSITIVITY:	1.000	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.000
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

=====			
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE			
=====			
Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00

PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		
SENSITIVITY:	0.778	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.667
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

=====			
BOOTSTRAP BUILD 6 OF 10			
=====			
31 Observations in the learning sample.			
Tree constructed with complexity parameter = 0.000			
=====			
TREE SEQUENCE			
=====			
Terminal Resubstitution Complexity			
Tree	Nodes	Relative Cost	Parameter

1	2	0.000	0.000
2	1	1.000	0.500
Initial misclassification cost = 0.500			
Initial class assignment = NEGATIF			
=====			
CLASSIFICATION TREE DIAGRAM			
=====			

-----1-----			
Terminal Regions			
1			2

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

```
*****
*           Node 1: BENJOLAN           *
*           N: 31                       *
*****
```

```
=====
=           Terminal Node 1           =           Terminal Node 2           =
=           N: 24                     =           N: 7                       =
=====
```

```
Node 1 was split on BENJOLAN
A case goes left if BENJOLAN = (0)
Improvement = 0.500           Complexity Threshold = 0.500
```

Node	Cases	Wgt	Counts	Cost	Class
1	31		31.00	0.500	NEGATIF
-1	24		24.00	0.000	NEGATIF
-2	7		7.00	0.000	POSITIF

Class	Weighted Counts		
	Top	Left	Right
NEGATIF	24.00	24.00	0.00
POSITIF	7.00	0.00	7.00

Class	Within Node Probabilities		
	Top	Left	Right
NEGATIF	0.500	1.000	0.000
POSITIF	0.500	0.000	1.000

Surrogate	Split	Assoc.	Improve.
1 STATUS_P s 0		0.792	0.328
2 DURASI_T s 1,2,3		0.732	0.268
3 USIA_45 s 0		0.631	0.208
4 KERUTAN_ s 0		0.571	0.200
5 NYERI_BA s 0		0.571	0.200

Competitor	Split	Improve.
1 STATUS_P 0		0.328
2 DURASI_T 1,2,3		0.268
3 USIA_45 0		0.208
4 KERUTAN_ 0		0.200
5 NYERI_BA 0		0.200

```
=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	24.00	0.00	24.00
POSITIF	0.00	7.00	7.00
PRED. TOT.	24.00	7.00	31.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.226	0.774	
TOT. CORRECT	1.000		

SENSITIVITY:	1.000	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.000
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

=====			
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE			
=====			
Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
-----	-----	-----	-----
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00
-----	-----	-----	-----
PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		
SENSITIVITY:	0.778	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.667
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

=====			
BOOTSTRAP BUILD 7 OF 10			
=====			
37 Observations in the learning sample.			
Tree constructed with complexity parameter = 0.000			
=====			
TREE SEQUENCE			
=====			
Tree	Terminal Nodes	Resubstitution Relative Cost	Complexity Parameter
-----	-----	-----	-----
1	3	0.000	0.000
2	2	0.077	0.038
3	1	1.000	0.462
Initial misclassification cost = 0.500			
Initial class assignment = NEGATIF			
=====			
CLASSIFICATION TREE DIAGRAM			
=====			
<pre> -----1----- -----2----- </pre>			
Terminal Regions			
1	2	3	
=====			
NODE INFORMATION			
=====			

*	Node 1: NYERI_BA	*	
*	N: 37	*	

```

*****
*           Node 2           *
*           N: 25           *
*****
=====
=           Terminal Node 3   =
=           N: 12           =
=====

```

Node 1 was split on NYERI_BA
A case goes left if NYERI_BA = (0)
Improvement = 0.429 Complexity Threshold = 0.462

Node	Cases	Wgt Counts	Cost	Class
1	37	37.00	0.500	NEGATIF
2	25	25.00	0.071	NEGATIF
-3	12	12.00	0.000	POSITIF

Weighted Counts			
Class	Top	Left	Right
NEGATIF	24.00	24.00	0.00
POSITIF	13.00	1.00	12.00

Within Node Probabilities			
Class	Top	Left	Right
NEGATIF	0.500	0.929	0.000
POSITIF	0.500	0.071	1.000

Surrogate	Split	Assoc.	Improve.
1 RIWAYAT_ s 0		0.910	0.353
2 RIWAYAT0 s 0		0.750	0.265
3 PENARIKA s 0		0.750	0.265
4 KERUTAN_ s 0		0.705	0.228
5 BENJOLAN_ s 0		0.667	0.313

Competitor	Split	Improve.
1 RIWAYAT_ 0		0.353
2 BENJOLAN_ 0		0.313
3 STATUS_P 0		0.300
4 PENARIKA 0		0.265
5 RIWAYAT0 0		0.265


```

*****
*           Node 2: BENJOLAN   *
*           N: 25           *
*****

```

```

=====
=           Terminal Node 1   =
=           N: 24           =
=====
=====
=           Terminal Node 2   =
=           N: 1           =
=====

```

Node 2 was split on BENJOLAN
A case goes left if BENJOLAN = (0)
Improvement = 0.071 Complexity Threshold = 0.038

Node	Cases	Wgt Counts	Cost	Class
2	25	25.00	0.071	NEGATIF
-1	24	24.00	0.000	NEGATIF
-2	1	1.00	0.000	POSITIF

Weighted Counts			
Class	Top	Left	Right
NEGATIF	24.00	24.00	0.00
POSITIF	1.00	0.00	1.00

Within Node Probabilities			
Class	Top	Left	Right
NEGATIF	0.929	1.000	0.000

POSITIF	0.071	0.000	1.000	
Competitor			Split	Improve.
1 FREKUENS	0,1,2			0.031
2 RIWAYAT1	0			0.031
3 USIA_45	0			0.019
4 LATENSI	0,2,3			0.015
5 STATUS_P	0			0.013

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	24.00	0.00	24.00
POSITIF	0.00	13.00	13.00
PRED. TOT.	24.00	13.00	37.00
CORRECT	1.000	1.000	
SUCCESS IND.	0.351	0.649	
TOT. CORRECT	1.000		

SENSITIVITY: 1.000 SPECIFICITY: 1.000
 FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.000
 REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

=====

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00
PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		

SENSITIVITY: 0.778 SPECIFICITY: 1.000
 FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.667
 REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

=====

BOOTSTRAP BUILD 8 OF 10

=====

33 Observations in the learning sample.
 Tree constructed with complexity parameter = 0.000

=====

TREE SEQUENCE

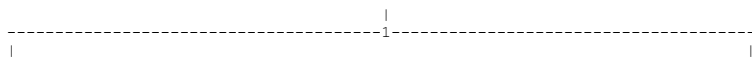
=====

Tree	Terminal Nodes	Resubstitution Relative Cost	Complexity Parameter
1	2	0.040	0.000
2	1	1.000	0.480

Initial misclassification cost = 0.500

Initial class assignment = NEGATIF

=====
 CLASSIFICATION TREE DIAGRAM
 =====



Terminal Regions

1 2

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

```

*****
*           Node 1: BENJOLAN           *
*           N: 33                       *
*****

```

```

=====
=           Terminal Node 1           =           Terminal Node 2           =
=           N: 24                     =           N: 9                     =
=====

```

Node 1 was split on BENJOLAN

A case goes left if BENJOLAN = (0)

Improvement = 0.462

Complexity Threshold = 0.480

Node	Cases	Wgt	Counts	Cost	Class
1	33		33.00	0.500	NEGATIF
-1	24		24.00	0.000	NEGATIF
-2	9		9.00	0.038	POSITIF

Class	Weighted Counts		
	Top	Left	Right
NEGATIF	25.00	24.00	1.00
POSITIF	8.00	0.00	8.00

Class	Within Node Probabilities		
	Top	Left	Right
NEGATIF	0.500	1.000	0.038
POSITIF	0.500	0.000	0.962

Surrogate	Split	Assoc.	Improve.
1 NYERI_BA s 0		0.740	0.264
2 STATUS_P s 0		0.708	0.306
3 CEKUNGAN s 0		0.698	0.300
4 RIWAYATO s 0		0.698	0.300
5 PAYUDARA s 0		0.698	0.300

Competitor	Split	Improve.
1 STATUS_P 0		0.306
2 PAYUDARA 0		0.300
3 CEKUNGAN 0		0.300
4 RIWAYATO 0		0.300
5 NYERI_BA 0		0.264

```
=====
LEARNING SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	24.00	1.00	25.00
POSITIF	0.00	8.00	8.00

PRED. TOT.	24.00	9.00	33.00
CORRECT	0.960	1.000	
SUCCESS IND.	0.202	0.758	
TOT. CORRECT	0.970		

SENSITIVITY:	0.960	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.111
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

```
=====
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE
=====
```

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00

PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		

SENSITIVITY:	0.778	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.667
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

```
=====
BOOTSTRAP BUILD 9 OF 10
=====
```

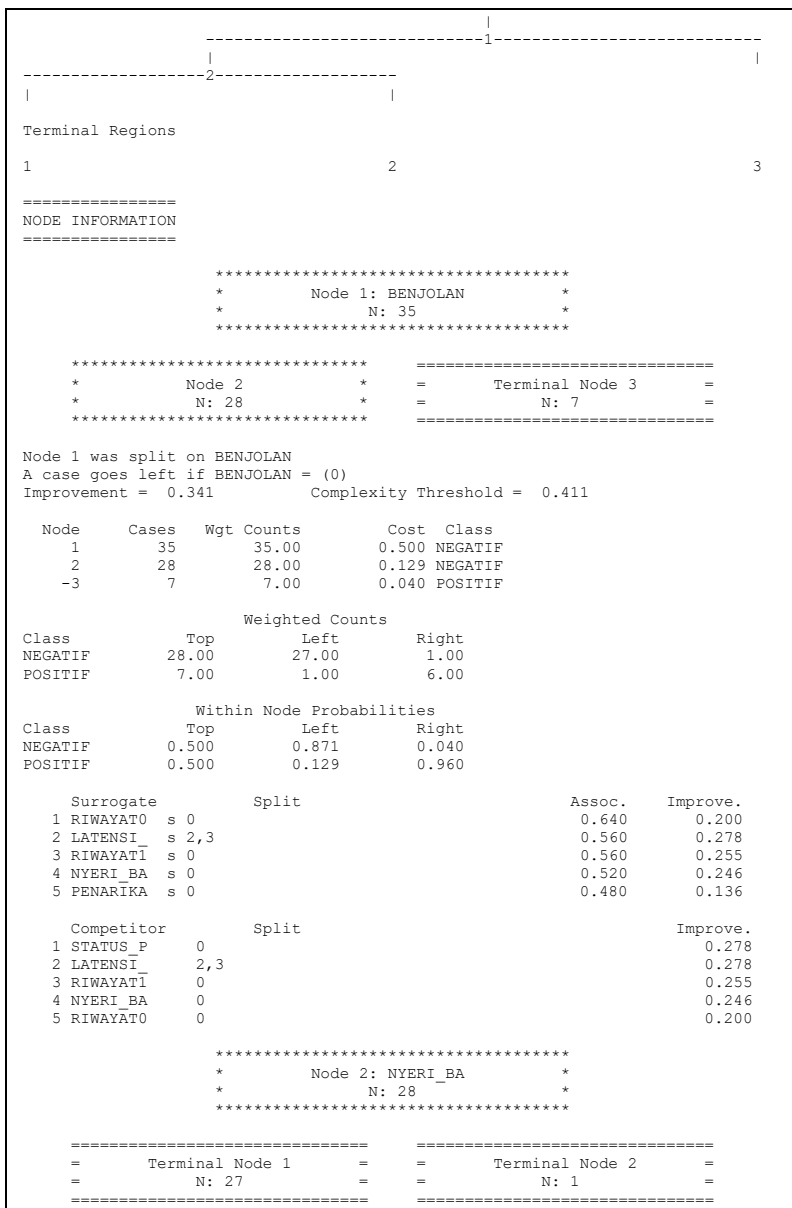
35 Observations in the learning sample.
Tree constructed with complexity parameter = 0.000

```
=====
TREE SEQUENCE
=====
```

Tree	Terminal Resubstitution		Complexity Parameter
	Nodes	Relative Cost	
1	3	0.036	0.000
2	2	0.179	0.071
3	1	1.000	0.411

Initial misclassification cost = 0.500
Initial class assignment = NEGATIF

```
=====
CLASSIFICATION TREE DIAGRAM
=====
```



Node 2 was split on NYERI_BA
 A case goes left if NYERI_BA = (0)
 Improvement = 0.124 Complexity Threshold = 0.071

Node	Cases	Wgt	Counts	Cost	Class
2	28		28.00	0.129	NEGATIF
-1	27		27.00	0.000	NEGATIF
-2	1		1.00	0.000	POSITIF

Weighted Counts				
Class	Top	Left	Right	
NEGATIF	27.00	27.00	0.00	
POSITIF	1.00	0.00	1.00	

Within Node Probabilities				
Class	Top	Left	Right	
NEGATIF	0.871	1.000	0.000	
POSITIF	0.129	0.000	1.000	

Surrogate		Split	Assoc.	Improve.
1	RIWAYAT1	s 0	0.250	0.063
2	USIA_45	s 0	0.250	0.063

Competitor		Split	Improve.
1	RIWAYAT1	0	0.063
2	USIA_45	0	0.063
3	RIWAYAT_	0	0.053
4	LATENSI_	0,2,3	0.045
5	STATUS_P	0	0.029

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	27.00	1.00	28.00
POSITIF	0.00	7.00	7.00

PRED. TOT.	27.00	8.00	35.00
CORRECT	0.964	1.000	
SUCCESS IND.	0.164	0.800	
TOT. CORRECT	0.971		

SENSITIVITY:	0.964	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.125
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

=====

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00

PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		

SENSITIVITY: 0.778 SPECIFICITY: 1.000
 FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.667
 REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

=====

BOOTSTRAP BUILD 10 OF 10

=====

34 Observations in the learning sample.
 Tree constructed with complexity parameter = 0.000

=====

TREE SEQUENCE

=====

Terminal Resubstitution Complexity

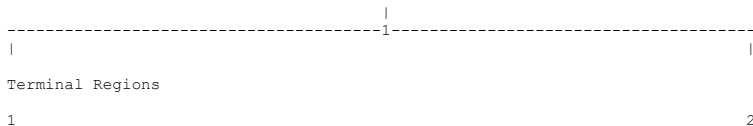
Tree	Nodes	Relative Cost	Parameter
1	2	0.071	0.000
2	1	1.000	0.464

Initial misclassification cost = 0.500
 Initial class assignment = NEGATIF

=====

CLASSIFICATION TREE DIAGRAM

=====



=====

NODE INFORMATION

=====

```

*****
*           Node 1: RIWAYAT1          *
*           N: 34                      *
*****
    
```

```

=====
=           Terminal Node 1           = =           Terminal Node 2           =
=           N: 26                      = =           N: 8                      =
=====
    
```

Node 1 was split on RIWAYAT1
 A case goes left if RIWAYAT1 = (0)
 Improvement = 0.433 Complexity Threshold = 0.464

Node	Cases	Wgt	Counts	Cost	Class
1	34		34.00	0.500	NEGATIF
-1	26		26.00	0.000	NEGATIF
-2	8		8.00	0.067	POSITIF

Class	Weighted Counts		
	Top	Left	Right
NEGATIF	28.00	26.00	2.00
POSITIF	6.00	0.00	6.00

Within Node Probabilities			
Class	Top	Left	Right
NEGATIF	0.500	1.000	0.067
POSITIF	0.500	0.000	0.933

Surrogate	Split	Assoc.	Improve.
1 LATENSI	s 0,2,3	0.846	0.375
2 BENJOLAN	s 0	0.782	0.324
3 USIA_45	s 0	0.731	0.348
4 STATUS_P	s 0	0.692	0.278
5 NYERI_EA	s 0	0.603	0.218

Competitor	Split	Improve.
1 LATENSI	0,2,3	0.375
2 USIA_45	0	0.348
3 BENJOLAN	0	0.324
4 STATUS_P	0	0.278
5 NYERI_EA	0	0.218

=====

LEARNING SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	26.00	2.00	28.00
POSITIF	0.00	6.00	6.00

PRED. TOT.	26.00	8.00	34.00
CORRECT	0.929	1.000	
SUCCESS IND.	0.105	0.824	
TOT. CORRECT	0.941		

SENSITIVITY:	0.929	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.250
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00

PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		

SENSITIVITY:	0.778	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.667
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

Lampiran 7. Ketepatan Klasifikasi Dengan Metode *Bootstrap Aggregating*

Output 7.1 Ketepatan Klasifikasi Hasil *Bootstrap* Dengan 10 dan 15 Replikasi

=====			
LEARNING SAMPLE CLASSIFICATION TABLE			
=====			
Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	27.00	1.00	28.00
POSITIF	0.00	7.00	7.00

PRED. TOT.	27.00	8.00	35.00
CORRECT	0.964	1.000	
SUCCESS IND.	0.164	0.800	
TOT. CORRECT	0.971		

SENSITIVITY:	0.964	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.125
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

=====			
CASE-BY-CASE SAMPLE CLASSIFICATION TABLE			
=====			
Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00

PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		

SENSITIVITY:	0.778	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.667
REFERENCE = "NEGATIF", RESPONSE = "POSITIF"			

Output 7.2 Ketepatan Klasifikasi Hasil *Bootstrap* Dengan 20, 25, 30 dan 35 Replikasi

=====			
LEARNING SAMPLE CLASSIFICATION TABLE			
=====			
Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	31.00	1.00	32.00
POSITIF	0.00	4.00	4.00

PRED. TOT.	31.00	5.00	36.00
CORRECT	0.969	1.000	
SUCCESS IND.	0.080	0.889	
TOT. CORRECT	0.972		

SENSITIVITY: 0.969 SPECIFICITY: 1.000
 FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.200
 REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

=====

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

=====

Actual Class	Predicted Class		Actual Total
	NEGATIF	POSITIF	
NEGATIF	7.00	2.00	9.00
POSITIF	0.00	1.00	1.00
PRED. TOT.	7.00	3.00	10.00
CORRECT	0.778	1.000	
SUCCESS IND.	-0.122	0.900	
TOT. CORRECT	0.800		

SENSITIVITY: 0.778 SPECIFICITY: 1.000
 FALSE REFERENCE: 0.000 FALSE RESPONSE: 0.667
 REFERENCE = "NEGATIF", RESPONSE = "POSITIF"

Lampiran 8. Salah Satu Hasil Tes Deteksi Kanker Payudara Menggunakan Ultrasonografi (USG)

RSUP Dr Sardjito
 Jl. Kesehatan No. 1 Sekeloa Yogyakarta
 Telp. +62-274-367333
 Faks. +62-274-566369

HASIL PEMERIKSAAN
 Er-Fr Crb B2 / Her 2 Topo PSE

No. PA	: IH6-17-1085	Ruang / Poli	: Cendana 2 (Bedah)
No. Reg	: 01-04-01-201700346558-002	Kelas	: Kelas I Non AC
No. RM	: 01.82.94.26	Tanggal Periksa	: 04/12/2017 10:06
Nama	: Lathifatuz Zahroh	Tanggal Hasil	: 10/01/2018 14:00
Tanggal Lahir	: 01-08-1987	Topografi	: -
Jenis Kelamin	: Perempuan	Morfologi	: -
Rujukan	: Pelayanan Rawat Inap		

Makroskopik
 HC-17-15432

Mikroskopik
 ER : pulasan positif sedang sampai kuat pada 30-40% inti sel tumor
 FR : pulasan positif sedang sampai kuat pada 30-40% inti sel tumor
 HER 2 : pulasan positif ringan sampai kuat sebagian tidak melingkar pada 10-20% sel tumor (41)
 Ki67 : pulasan positif sedang sampai kuat pada 15% inti sel tumor

Kesimpulan
 ER, FR : + (positif)
 HER 2 : 1+ (negatif)
 Ki67 : indeks proliferasi 15%

Yogyakarta, 10 Januari 2018
 Pemeriksa *C. R.*

Ahli Rajal Patologi Anatomi
 dr. Ahmad Shozali, Sp-PA(K)

dr. Rita Daspaqa Sp.FS

CATATAN
 Jika sekiranya ada keraguan tentang hasil pemeriksaan diharapkan segera menghubungi
 Instalasi Patologi Anatomi RSUP Dr Sardjito

Lampiran 9. Surat Keterangan Pengambilan Data

SURAT KETERANGAN

Saya yang bertanda tangan di bawah ini menerangkan bahwa :

1. Mahasiswa Statistika FMKSD-ITS dengan identitas berikut :

Nama : Desintya Rachma Anggraeni Putri

NRP : 06211540000113

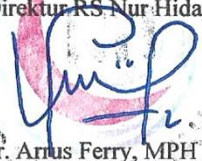
Telah mengambil data di instansi/perusahaan kami :

Nama Instansi : RS Nur Hidayah

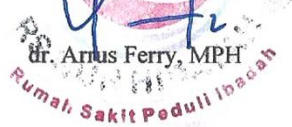
sejak tanggal 7 April 2019 sampai dengan 28 April 2019 untuk keperluan Tugas Akhir/ ~~Thesis~~ Semester Gasal/Genap* 2018/ 2019.

2. Tidak Keberatan/~~Keberatan~~* nama perusahaan dicantumkan dalam Tugas Akhir/ ~~Thesis~~ mahasiswa Statistika yang akan di simpan di Perpustakaan ITS dan dibaca di lingkungan ITS.
3. Tidak Keberatan/~~Keberatan~~* bahwa hasil analisis data dari perusahaan dipublikasikan dalam E journal ITS yaitu Jurnal Sains dan Seni ITS.

Bantul, 1 Juli 2019
Direktur RS Nur Hidayah



Dr. Arius Ferry, MPH



*(coret yang tidak perlu)

Lampiran 10. Surat Ijin Pengambilan Data Dari Pihak RSUD Nur Hidayah Bantul



Profesional ♦ Bersahabat ♦ Islami

Nomor : 03 /RSNH/KET/DIR/IV/2019
Perihal : Balasan Ijin Pengambilan Data Tugas Akhir
Lampiran : --

Kepada Yth.
Desintya Rachma Anggraeni Putri
Di tempat

Assalamu'alaikum Warohmatullahi Wabarokatuh

Puji syukur kehadiran Allah SWT atas limpahan rahmat dan karuniaNya. Shalawat dan salam semoga tercurah kepada Nabi Besar Muhammad SAW, keluarga, sahabat dan para pengikutnya. Semoga kemudahan dan kebaikan senantiasa diberikan pada usaha kita. Menindaklanjuti surat dari Institut Teknologi Sepuluh Nopember Fakultas Matematika Komputasi dan Sains Data Departemen Statistika dengan 25165/IT2.VI.9.2/TU.00.09/2019 perihal permohonan ijin Penelitian:

Nama : Desintya Rachma Anggraeni Putri
No. Mhs : 06211540000113 (1315 100 113)
Prodi : Sarjana (S1)
Perguruan Tinggi : Institut Teknologi Sepuluh Nopember

Perihal kegiatan:

Jenis kegiatan : Ijin Pengambilan Data Tugas Akhir
Lokasi Kegiatan : RS Nur Hidayah
Mata Kuliah : "Klasifikasi Hasil Tes Deteksi Kanker Payudara Berdasarkan Gejala Klinis Menggunakan Metode Classification Tress di RS Nur Hidayah Bantul

Waktu pelaksanaan : 06 – 07 April 2019
Pembimbing : Dr. Estianna Khoirunnisa (+62 822-2027-5909)

Maka dengan ini kami memberikan ijin kepada mahasiswa tersebut untuk melakukan Pengambilan Data Tugas Akhir di RS Nur Hidayah dengan ketentuan sebagai berikut:

1. Wajib menjaga tata tertib dan mentaati ketentuan-ketentuan yang berlaku selama Penelitian di lingkungan rumah sakit
2. Wajib memberikan laporan hasil penelitian berupa Hard Copy dan Soft Copy kepada Direktur c/q Penanggungjawab Diklat RS Nur Hidayah Bantul
3. Surat izin ini hanya diperlukan untuk kegiatan ilmiah
4. Surat izin ini dapat dibatalkan sewaktu-waktu apabila tidak dipenuhi ketentuan-ketentuan yang sudah disampaikan

Demikian surat ini kami sampaikan. Atas perhatian dan kerjasamanya kami ucapkan terima kasih.

Wassalamu'alaikum Warohmatullohi Wabarokatuh

Bantul, April 2019
Menyetujui
Direktur RS Nur Hidayah

Dr. Agus Ferry, MPH.
RUMAH SAKIT PECAUT

Tembusan:

1. Pembimbing Lapangan/ CI
2. Bagian Diklat
3. Yang bersangkutan

(Halaman ini sengaja dikosongkan)

BIODATA PENULIS



Penulis dengan nama lengkap Desintya Rachma Anggraeni Putri yang lebih dikenal dengan nama sapaan Sintya lahir di Nganjuk pada 9 Desember 1996. Penulis menempuh pendidikan formal mulai dari TK Bhayangkari 42 Kediri, SD Negeri Burengan 2 Kediri, SMP Negeri 1 Kediri, dan SMA Negeri 1 Kediri. Penulis diterima sebagai Mahasiswa Departemen Statistika ITS melalui jalur Mandiri pada tahun 2015.

Putri dari pasangan Bapak Bambang Murdianto dan Ibu Retno Wulandari ini aktif dalam organisasi dan panitia kegiatan seperti *Staff* Tim Penelitian dan Pengembangan HIMASTA ITS 2016/2017, Sekretaris Tim Penelitian dan Pengembangan HIMASTA-ITS 2017/2018, panitia Pekan Raya Statistika ITS 2016/2017 dan panitia di berbagai kegiatan yang diadakan oleh HIMASTA ITS. Selama menjalani masa perkuliahan, penulis pernah menjadi asisten dosen untuk mata kuliah pemrograman komputer dan juga mengikuti beberapa *project* sebagai *surveyor*. Selain itu penulis juga aktif mengikuti perlombaan yang diadakan oleh jurusan statistika di berbagai universitas yang ada di Indonesia. Pada bulan Agustus 2018 penulis berkesempatan untuk melakukan *internship program* di KPw Bank Indonesia Malang. Apabila pembaca ingin memberi kritik dan saran serta diskusi lebih lanjut mengenai Tugas Akhir ini, dapat menghubungi penulis melalui email cintyarachma@gmail.com.