



TUGAS AKHIR – SS141501

**BOOSTING *NEURAL NETWORK* DAN BOOSTING
CART PADA KLASIFIKASI JENIS KELAMIN
DENGAN METODE PENGUKURAN LINIER PADA
FORAMEN MENTAL MENGGUNAKAN
*PANORAMIC RADIOGRAPH***

**NUR LAILATUL FAIZAH
NRP. 1313 100 103**

**Dosen Pembimbing
Dr. Kartika Fithriasari, M.Si**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**



FINAL PROJECT – SS141501

**BOOSTING NEURAL NETWORK AND BOOSTING
CART FOR GENDER CLASSIFICATION BY
MENTAL FORAMENT USING LINEAR
MEASUREMENTS METHOD ON PANORAMIC
RADIOGRAPH**

**NUR LAILATUL FAIZAH
NRP. 1313 100 103**

**Supervisor
Dr. Kartika Fithriasari, M.Si**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2017**

**LEMBAR PENGESAHAN
BOOSTING NEURAL NETWORK DAN BOOSTING
CART PADA KLASIFIKASI JENIS KELAMIN
DENGAN METODE PENGUKURAN LINIER PADA
FORAMEN MENTAL MENGGUNAKAN
PANORAMIC RADIOGRAPH**

TUGAS AKHIR

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

Oleh :

NUR LAILATUL FAIZAH
NRP. 1313 100 103

Disetujui oleh Pembimbing:
Dr. Kartika Fithriasari, M.Si
NIP. 196912121993032002

()

Mengetahui,
Kepala Departemen


Dr. Suhartono
NIP. 19710929 199512 1 001

SURABAYA, JULI 2017

**BOOSTING NEURAL NETWORK DAN BOOSTING
CART PADA KLASIFIKASI JENIS KELAMIN
DENGAN METODE PENGUKURAN LINIER PADA
FORAMEN MENTAL MENGGUNAKAN
PANORAMIC RADIOGRAPH**

Nama : Nur Lailatul Faizah
NRP : 1313100103
Jurusan : Statistika
Dosen Pembimbing : Dr. Kartika Fithriasari, M.Si

Abstrak

Fakta pengamatan dilapangan menunjukkan bahwa identifikasi korban meninggal massal menggunakan tulang tengkorak memiliki kontribusi yang tinggi dalam menentukan identitas seseorang, sehingga pada penelitian ini akan digunakan salah satu bagian dari tulang tengkorak yakni foramen mental yang terletak di mandibula untuk identifikasi jenis kelamin. Pada penelitian ini digunakan metode FFNN dan CART dalam mengklasifikasikan jenis kelamin pada foramen mental sebagai upaya untuk meningkatkan akurasi dengan proses boosting. Model terbaik FFNN untuk klasifikasi jenis kelamin yaitu dengan menggunakan 3 input, 1 hidden layer dengan jumlah neuron 6 unit, dan 1 output. Pohon optimal analisis CART yaitu dengan jumlah simpul terminal sebanyak 5 simpul dan tingkat kedalaman 5. Hasil penelitian ini menghasilkan bahwa klasifikasi dengan Boosting pada metode Classification and Regression Tree (CART) memiliki performa yang lebih baik dibandingkan dengan metode lainnya, dimana accuracy, sensitivity, dan specificity yang didapat adalah 91.67%, 100.00%, dan 83.33%.

Kata Kunci : Boosting, Classification and Regression Tree, Feedforward Neural Network, Foramen mental, Mandibula, Panoramic Radiograph.

(Halaman ini sengaja dikosongkan)

BOOSTING NEURAL NETWORK AND BOOSTING CART FOR GENDER CLASSIFICATION WITH LINEAR MEASUREMENT METHOD ON MENTAL FORAMENT USING PANORAMIC RADIOGRAPH

Name : Nur Lailatul Faizah
NRP : 1313 100 103
Department : Statistics
Supervisor : Dr. Kartika Fithriasari, M.Si

Abstract

Based on experience in field, mass dead victims identified using skeletal bone has high contribution in determining a person identity, so this research is using a part of skeletal bone which is foramen mentale on mandible for gender identification. This research using FFNN Method and CART Method in gender classification on foramen mentale to improve the accuration with boosting process. The best FFNN model to gender classification is using 3 inputs, 1 hidden layer with 6 neuron units, and 1 output. CART analysis optimal tree with 5 terminal nodes and depth level 5. This research results show that classification using Boosting with Classification and Regression Tree (CART) method has a better performance compared to other methods with accuracy, sensitivity, and specitivity percentage 91.67%, 100.00% and 83.33%.

***Keyword: Boosting, Classification and Regression Tree,
Feedforward Neural Network, Foramen mentale,
Mandible, panoramic radiograph.***

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puja dan puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa atas segala berkah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul:

“Boosting *Neural network* dan Boosting CART pada Klasifikasi Jenis Kelamin dengan Metode Pengukuran Linier pada Foramen Mental menggunakan *Panoramic Radiograph*”

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini tidak terlepas dari bantuan maupun dukungan dari berbagai pihak yang telah terlibat baik secara langsung maupun tidak. Oleh karena itu, pada kesempatan ini penulis menyampaikan rasa terima kasih kepada:

1. Bapak Dr. Suhartono selaku Kepala Departemen Statistika ITS dan Bapak Dr. Sutikno, S.Si, M.Si selaku Ketua Program Studi Sarjana Statistika ITS.
2. Ibu Dr. Kartika Fithriasari, M.Si selaku dosen pembimbing Tugas Akhir yang telah sabar dalam memberikan banyak masukan serta pengarahan yang sangat bermanfaat.
3. Bapak Dr. Bambang Widjanarko Otok, S.Si, M.Si dan Ibu Dr. Irhamah, S.Si, M.Si selaku dosen penguji yang telah memberikan saran, kritik, dan masukan untuk kesempurnaan Tugas Akhir ini.
4. Ayah dan Ibu tersayang serta Kakak dan Adik yang telah memberikan motivasi, bimbingan, dan doa kepada penulis sehingga bisa menyelesaikan Tugas Akhir ini.

Penulis menyadari Tugas Akhir ini masih belum sempurna, sehingga kritik dan saran yang membangun sangat diharapkan guna perbaikan di masa yang akan datang sehingga hasil dari Tugas Akhir ini memberikan manfaat bagi pembaca dan semua pihak.

Surabaya, Juli 2017

Penulis

(Halaman ini sengaja dikosongkan)

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
TITLE PAGE	iii
LEMBAR PENGESAHAN	v
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	5
1.3. Tujuan Penelitian.....	6
1.4. Manfaat.....	6
1.5. Batasan Masalah.....	7
BAB II TINJAUAN PUSTAKA	9
2.1. <i>Machine Learning</i>	9
2.2. Neural Network (NN).....	10
2.3. <i>Feedforward Neural network (FFNN)</i>	11
2.4. <i>Classification and Regression Tree (CART)</i>	16
2.5. Ilustrasi Pembentukan Pohon Klasifikasi CART.....	20
2.6. Boosting.....	29
2.7. Ilustrasi Boosting.....	32
2.8. Ukuran Ketepatan Klasifikasi.....	33
2.9. Panoramic radiograph.....	35
2.10. Foramen Mental.....	36
BAB III METODOLOGI PENELITIAN	37
3.1. Sumber Data.....	37
3.2. Variabel Penelitian.....	37
3.3. Definisi Operasional.....	37
3.4. Langkah Analisis.....	39
3.5. Diagram Alir.....	41

BAB IV ANALISIS DAN PEMBAHASAN.....	43
4.1. Deskripsi Data Hasil Pengukuran Linier pada Foramen Mental	43
4.2. Klasifikasi Jenis Kelamin Berdasarkan Foramen Mental	45
4.2.1. Klasifikasi dengan Metode <i>Feedforward</i> Neural Network (FFNN).....	45
4.2.2. Klasifikasi dengan Boosting pada Metode <i>Feedforward</i> Neural Network (FFNN)	48
4.2.3. Klasifikasi dengan Metode <i>Classification and Regression Tree</i> (CART)	49
4.2.4. Klasifikasi dengan Boosting pada Metode <i>Classification and Regression Tree</i> (CART)	59
4.5. Perbandingan Performa Antar Metode.....	61
BAB V KESIMPULAN DAN SARAN.....	63
5.1. Kesimpulan	63
5.2. Saran	63
DAFTAR PUSTAKA	65
LAMPIRAN	69
BIODATA PENULIS	78

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Jaringan Syaraf Tiruan <i>Backpropagation</i> dengan 1 Lapis Tersembunyi	12
Gambar 2.2 Pohon Klasifikasi Ilustrasi Hasil Pemilahan I.....	26
Gambar 2.3 Pohon Klasifikasi Hasil Pemilahan II	28
Gambar 2.4 Ilustrasi Boosting CART.....	32
Gambar 2.5 Ilustrasi Boosting CART.....	32
Gambar 2.6 (a) Posisi pengambilan radiografi panoramik, (b) Hasil radiografi panoramik.....	35
Gambar 2.7 (a) Tulang mandibula tampak depan, (b) Tulang mandibula tampak samping kiri, (c) Panoramic radiograph pada foramen mental.....	36
Gambar 3.1 Jarak A-B	38
Gambar 3.2 Jarak A-C	39
Gambar 3.3 Jarak D-B	39
Gambar 3.4 Langkah-langkah yang dilakukan dalam penelitian	41
Gambar 4.1 Perbedaan sebaran data antara jenis kelamin laki-laki dan jenis kelamin perempuan.....	43
Gambar 4.2 Sebaran data antar kategori jenis kelamin laki- laki dan perempuan pada setiap variabel prediktor	44
Gambar 4.3 <i>Topology</i> pohon klasifikasi maksimal.....	51
Gambar 4.4 Plot <i>relative cost</i>	52
Gambar 4.5 <i>Topology</i> pohon klasifikasi optimal	53

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 2.1 Variabel Penelitian Ilustrasi CART	21
Tabel 2.2 Contoh Data Sampel Ilustrasi CART	21
Tabel 2.3 Proporsi Pengamatan pada Respon Y	22
Tabel 2.4 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> $X_1=2,5$	23
Tabel 2.5 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> $X_1=3,5$	23
Tabel 2.6 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> $X_1=5$	24
Tabel 2.7 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> $X_2=5,5$	24
Tabel 2.8 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> $X_2=8$	25
Tabel 2.9 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> $X_2=11$	25
Tabel 2.10 Perbandingan <i>Goodness of Split</i> dari Pemilah 1	26
Tabel 2.11 Data Sampel Ilustrasi CART Pemilahan 2	27
Tabel 2.12 Perbandingan <i>Goodness of Split</i> dari Pemilah 2	27
Tabel 2.13 Pemberian Label pada Simpul Terminal 1	29
Tabel 2.14 <i>Confussion Matrix</i>	34
Tabel 3.1 Variabel yang digunakan dalam penelitian	37
Tabel 4.1 Performa metode FFNN pada setiap <i>folds</i>	46
Tabel 4.2 Performa metode FFNN pada setiap neuron pada <i>hidden layer</i>	47
Tabel 4.3 <i>Confoussion Matrix Boosting</i> FFNN untuk Data <i>Training</i>	48
Tabel 4.4 <i>Confoussion Matrix Boosting</i> FFNN untuk Data <i>Testing</i>	49
Tabel 4.5 Variabel terpenting dari pohon klasifikasi maksimal	50
Tabel 4.6 Urutan Pembentukan Pohon Klasifikasi (<i>Tree</i> <i>Sequence</i>)	52

Tabel 4.7	Pelabelan Kelas Simpul Terminal.....	56
Tabel 4.8	Karakteristik Jenis Kelamin Berdasarkan Simpul Terminal dengan Persentase Terbesar.....	56
Tabel 4.9	Ilustrasi Data Baru	57
Tabel 4.10	Hasil Klasifikasi Data Baru	58
Tabel 4.11	<i>Confussion Matrix</i> CART.....	59
Tabel 4.12	<i>Confoussion Matrix Boosting</i> CART untuk Data Training.....	60
Tabel 4.13	<i>Confoussion Matrix Boosting</i> CART untuk Data Testing.....	60
Tabel 4.14	Rangkuman performa dari metode FFNN, boosting FFNN, CART, dan boosting CART.....	61

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Hasil Pengukuran Jarak pada Foramen Mental.....	69
Lampiran 2. <i>Syntax Feedforward</i> Neural Network (FFNN)...	70
Lampiran 3. <i>Output Syntax Feedforward</i> Neural Network (FFNN)	71
Lampiran 4. Boosting FFNN	72
Lampiran 5. <i>Output</i> Boosting FFNN.....	73
Lampiran 6. <i>Syntax</i> untuk memilik <code>rpart.control</code> yang terbaik	73
Lampiran 7. <i>Output</i> untuk memilik <code>rpart.control</code> yang terbaik	73
Lampiran 8. <i>Syntax Boosting Classification and Regression Tree</i> (CART).....	74
Lampiran 9. <i>Output</i> CART Fold 1 (data training).....	75
Lampiran 10. <i>Output</i> CART Fold 1 (data testing).....	75
Lampiran 11. <i>Output</i> CART Fold 2 (data training).....	75
Lampiran 12. <i>Output</i> CART Fold 2 (data testing).....	76
Lampiran 13. <i>Output</i> CART Fold 3 (data training).....	76
Lampiran 14. <i>Output</i> CART Fold 3 (data testing).....	76
Lampiran 15. <i>Output</i> CART Fold 4 (data training).....	77
Lampiran 16. <i>Output</i> CART Fold 4 (data testing).....	77
Lampiran 17. <i>Output</i> CART Fold 5 (data training).....	77
Lampiran 18. Surat Pernyataan Legalitas Data	77

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1. Latar Belakang

Salah satu bagian ilmu dari *Artificial Intellegence* (AI) adalah *Machine Learning*, dimana pengertian *Machine Learning* adalah suatu disiplin ilmu yang berkonsentrasi dengan desain dan pengembangan algoritma yang didasarkan pada data input atau informasi yang didapatkan (Zhang, 2010). Proses *learning* akan menghasilkan suatu pola atau *pattern* yang dapat digunakan sebagai solusi dalam menghadapi permasalahan klasifikasi dan prediksi.

Klasifikasi merupakan salah metode dalam statistika yang digunakan untuk memprediksi kelas dari sekumpulan data input. Metode *Machine Learning* yang biasa digunakan dalam klasifikasi adalah *Decision Tree*, *Support Vector Machine* (SVM), *Neural network*, dan *Naïve Bayes*. Metode klasifikasi tersebut memiliki kelebihan masing-masing sehingga penggunaan metode klasifikasi dapat disesuaikan dengan karakteristik data. Saat ini metode klasifikasi telah banyak diterapkan dalam berbagai aspek kehidupan untuk menyelesaikan suatu permasalahan, misalnya yaitu, dalam bidang kesehatan klasifikasi telah membantu dokter untuk mendiagnosa penyakit, dalam bidang keilmuan klasifikasi di gunakan untuk membantu peneliti dalam mengambil keputusan, dalam bidang multimedia klasifikasi digunakan untuk *face detection*, *speech recognition*, dan masih banyak lagi penerapan metode klasifikasi yang lainnya.

Dari beberapa manfaat klasifikasi yang telah dijelaskan diatas maka dalam penelitian ini metode klasifikasi akan digunakan untuk menyelesaikan permasalahan pada bidang kesehatan, yaitu kedokteran gigi forensik. Kasus yang di gunakan adalah klasifikasi jenis kelamin dengan metode pengukuran linier berdasarkan foramen mental menggunakan *panoramic radiograph*. Klasifikasi jenis kelamin berdasarkan foramen mental dapat digunakan untuk mengidentifikasi jenis kelamin ketika terjadi

bencana massal yang menimbulkan banyak korban jiwa, baik akibat bencana alam (*Natural Disaster*) maupun bencana yang disebabkan oleh manusia (*Man Made Disaster*).

Saat ini telah terjadi berbagai kejadian bencana yang menimbulkan banyak korban jiwa, terutama sejak kejadian Bom Bali I, kejadian tersebut membuat kegiatan identifikasi korban bencana massal (*Disaster Victim Identification*) menjadi kegiatan yang penting dan hampir selalu dilakukan ketika terjadi bencana yang menimbulkan korban jiwa dalam jumlah yang banyak. DVI adalah suatu definisi yang diberikan sebagai sebuah prosedur untuk mengidentifikasi korban mati akibat bencana massal secara ilmiah yang dapat dipertanggungjawabkan dan mengacu pada standar baku Interpol (Saparwoko, 2006). Tujuan utama dari pemeriksaan identifikasi pada kasus musibah bencana massal adalah agar dapat mengenali korban, misalnya golongan darah, DNA, jenis kelamin, dan identitas korban. Kejadian bencana massal tersebut akan menghasilkan keadaan jenazah yang mungkin dapat intak, separuh intak, membusuk, terpisah berfragmen-fragmen, terbakar menjadi abu, separuh terbakar, terkubur, ataupun kombinasi dari bermacam-macam keadaan (Blau, 2006).

Pada jenazah yang tidak dapat diidentifikasi secara visual yakni jenazah yang mengalami pembusukan lanjut, pemeriksaan identifikasi primer berdasarkan sidik jari akan sulit untuk dilakukan, sehingga dapat digantikan dengan pemeriksaan menggunakan tulang tengkorak karena tulang tengkorak bersifat lebih tahan lama terhadap proses pembusukan. Fakta pengalaman di lapangan menunjukkan bahwa identifikasi korban meninggal massal menggunakan tulang tengkorak memiliki kontribusi yang tinggi dalam menentukan identitas seseorang. Pada kasus Bom Bali I, korban yang teridentifikasi berdasarkan gigi geligi mencapai 56% dan pada kecelakaan lalu lintas bis terbakar di Situbondo mencapai 60% (Prawestiningtyas & Algozi, 2009).

Radiograph merupakan pemeriksaan penunjang yang sering dilakukan untuk membantu menegakkan suatu diagnosa. *Radiograph* pada kedokteran gigi dibagi menjadi dua yaitu intra oral dan

ekstra oral. *Radiograph* ekstra oral yang sering digunakan pada kedokteran gigi salah satunya adalah *panoramic radiograph* (White & Pharoah, 2009). Manfaat dari *panoramic radiograph* adalah untuk melihat adanya kelainan dan penyakit pada rongga mulut, serta untuk melihat dan mendeteksi manifestasi penyakit sistemik pada rongga mulut.

Foramen mental merupakan salah satu karakteristik mandibula yang dapat digunakan untuk mengestimasi usia dan menentukan jenis kelamin. Saat ini identifikasi dan penentuan jenis kelamin berdasarkan kerangka manusia telah menjadi salah satu tugas yang paling menantang bagi kedokteran gigi forensik. Perkembangan foramen mental dipengaruhi oleh berbagai faktor, yaitu usia, jenis kelamin, dan ras, oleh karena itu foramen mental merupakan salah satu karakteristik mandibula yang dapat di gunakan untuk estimasi usia dan penentuan jenis kelamin (Gupta & Soni, 2012).

Penentuan jenis kelamin pada foramen mental menggunakan *panoramic radiograph* dapat dilakukan dengan berbagai cara, salah satunya dengan cara pengukuran linear. Pengukuran linear pada foramen mental merupakan metode pengukuran dengan melihat besar jarak vertikal antara *superior border alveolar crest* hingga *inferior border* mandibula, jarak vertikal antara *superior border alveolar crest* dengan *superior border* foramen mental, dan jarak vertikal antara *inferior border* foramen mental dengan *inferior border* mandibula. Metode pengukuran linear memiliki beberapa kelebihan, salah satunya adalah metode ini lebih mudah diaplikasikan dalam interpretasi *radiograph* dalam mengidentifikasi jenis kelamin (Singal & Sharma, 2016).

Penelitian yang telah dilakukan oleh Chandra, Singh, Badni, Jaiswal, & Agnihotri (2013) dan Malik et. al (2016) menunjukkan bahwa jenis kelamin dapat ditentukan oleh foramen mental. Pada tahun 2016 Singal dan Sharma melakukan penelitian mengenai metode pengukuran linier dan didapatkan kesimpulan bahwa terdapat perbedaan yang signifikan dalam mengidentifikasi jenis kelamin dengan metode pengukuran linear. Penelitian lain yang berhubungan dengan pengukuran linear dilakukan oleh Rashid dan

Ali (2011) di Iraq pada 150 wanita dan 150 pria yang dibagi menjadi tiga kelompok rentang usia antara 20-29 tahun, 30-39 tahun dan 40-49 tahun, didapatkan jarak vertikal *superior border* foramen mental terhadap alveolar *crest* pada pria sebesar 17,4 mm dan pada wanita didapatkan jarak yang lebih kecil yaitu 16,35 mm. Sedangkan jarak vertikal *inferior border* foramen mental ke *inferior border* mandibula pada pria 10,6 mm dan 9,24 mm pada wanita.

Berdasarkan penelitian yang telah dilakukan oleh Ghouse et al (2016) didapat kesimpulan bahwa terdapat perbedaan signifikan antara jenis kelamin pria dan wanita pada setiap pengukuran jarak yang digunakan, sehingga pengukuran tersebut merupakan parameter yang sensitif dalam memprediksi jenis kelamin. Penelitian lain pada populasi India yang dilakukan oleh Thakur et. al (2014) menunjukkan bahwa terdapat perbedaan yang signifikan dalam membandingkan jarak mandibula pada laki-laki dan perempuan, hasil penelitian serupa juga dilakukan oleh Naroor et al. pada tahun 2015.

Penelitian mengenai tingkat akurasi atau ketepatan dalam klasifikasi telah dilakukan oleh Setiawan et. al (2016) pada kasus Klasifikasi Diabetes Militus Tipe II yang menghasilkan akurasi sebesar 80.65% dengan menggunakan metode analisis diskriminan dan Hu et. al (2016) pada penentuan jenis kelamin menggunakan karakteristik non metrik mandibula di populasi korea yang menghasilkan akurasi sebesar 85.71% dengan menggunakan uji *Chi Square*. Penelitian mengenai *boosting* telah dilakukan oleh Fernanda dan Otok pada tahun 2013 yang menghasilkan tingkat akurasi sebesar 98.75% dengan menggunakan metode *boosting* CART.

Metode *Machine Learning* yang digunakan dalam penelitian ini adalah *Feedforward Neural network* (FFNN) dan *Classification and Regression Tree* (CART). Kedua metode ini digunakan karena memiliki beberapa kelebihan dibandingkan dengan metode *Machine Learning* yang lainnya. Kelebihan dari metode FFNN yaitu dapat digunakan untuk kasus dengan hubungan antara data

input dan *output* yang kompleks (Singh & Chauhan, 2009), sedangkan kelebihan dari metode CART yaitu hasil yang di dapatkan sangat mudah dalam interpretasi karena hasilnya di tampilkan dalam suatu diagram seperti pohon (Rokach & Maimon, 2015).

Tingkat akurasi dari suatu metode klasifikasi dapat di tingkatkan dengan tujuan memberikan hasil klasifikasi yang lebih baik. Salah satu cara yang dapat digunakan untuk meningkatkan tingkat akurasi dengan menggunakan metode *ensemble* yaitu *boosting*. *Adaboost* yang merupakan variasi dari *boosting* adalah salah satu metode *ensemble* yang digunakan untuk meningkatkan akurasi dari suatu metode klasifikasi. Pada tahun 2000 Breiman, et. al menerapkan *adaboost* pada multilayer *neural network* dan didapatkan kesimpulan bahwa *adaboost* mampu meningkatkan akurasi dari metode multilayer *neural network*. Pada tahun Tran et. al (2009) juga menerapkan metode *adaboost* pada *probabilistic neural network* untuk *Novel Intrusion Detection* dan hasilnya juga memperlihatkan bahwa *adaboost* mampu meningkatkan tingkat akurasi dari *probabilistic neural network*.

Berdasarkan beberapa penjelasan tersebut, maka pada penelitian ini akan digunakan metode *Feedforward Neural network* (FFNN) dan *Classification and Regression Tree* (CART) dalam mengklasifikasikan jenis kelamin pada foramen mental sebagai upaya untuk meningkatkan akurasi dengan proses *boosting*. Diharapkan penelitian ini dapat membantu instansi kesehatan khususnya bagi kedokteran gigi forensik untuk meningkatkan akurasi atau ketepatan dalam klasifikasi jenis kelamin pada foramen mental dengan metode pengukuran linier menggunakan *panoramic radiograph*.

1.2. Rumusan Masalah

Identifikasi dan penentuan lokasi foramen mental sangat penting dalam kedokteran gigi forensik. Pengetahuan foramen mental bermanfaat dalam hal pemberian anestesi lokal untuk tujuan pembedahan dan dalam perawatan endodontik. Perkembangan foramen mental dipengaruhi oleh berbagai faktor, yaitu

usia, jenis kelamin, dan ras, oleh karena itu foramen mental merupakan salah satu karakteristik mandibula yang dapat digunakan untuk estimasi usia dan penentuan jenis kelamin.

Saat ini tingkat akurasi atau ketepatan dalam klasifikasi jenis kelamin pada foramen mental menggunakan *panoramic radiograph* pada penelitian sebelumnya menghasilkan tingkat akurasi yang masih cukup rendah. Disisi lain, metode klasifikasi yang digunakan dalam penelitian sebelumnya adalah analisis diskriminan, dimana banyak penelitian yang telah membuktikan bahwa *neural network* menghasilkan ketepatan klasifikasi yang cenderung lebih tinggi dibandingkan dengan analisis diskriminan.

Oleh karena itu, permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana model terbaik untuk metode FFNN dan metode CART dalam klasifikasi jenis kelamin pada foramen mental dan bagaimana perbandingan performa antara metode FFNN dan metode CART dalam klasifikasi jenis kelamin pada foramen mental.

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan yang ingin dicapai dalam penelitian ini adalah mendapatkan model terbaik untuk metode FFNN dan metode CART dalam klasifikasi jenis kelamin pada foramen mental dan mendapatkan metode yang terbaik berdasarkan hasil perbandingan performa antara metode FFNN dan metode CART dalam klasifikasi jenis kelamin pada foramen mental.

1.4. Manfaat

Manfaat yang diharapkan dari penelitian ini adalah dapat memberikan gambaran kepada orang awam mengenai pentingnya identifikasi lokasi foramen mental dan peran foramen mental dalam klasifikasi jenis kelamin. Manfaat lain yang diharapkan dari penelitian ini adalah dapat memberikan pengetahuan tambahan mengenai performa metode *Feedforward Neural network* (FFNN) dan *Classification and Regression Tree* (CART) dalam klasifikasi,

serta hasil dari penerapan metode *adaboost* dalam upaya untuk meningkatkan akurasi dalam klasifikasi.

1.5. Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Metode *Machine Learning* yang digunakan hanya *Feedforward Neural network* (FFNN) dan *Classification and Regression Tree* (CART).
2. Kriteria umur yang digunakan dalam penelitian sebelumnya yaitu orang yang berusia lebih dari 21 tahun.
3. Klasifikasi yang dilakukan hanya fokus pada klasifikasi jenis kelamin saja.

(Halaman ini sengaja dikosongkan)

BAB II

TINJAUAN PUSTAKA

2.1. Machine Learning

Machine Learning merupakan suatu proses yang dilakukan untuk membangun suatu sistem komputer yang secara otomatis dapat meningkatkan suatu model berdasarkan pengalaman atau informasi sebelumnya, dan menerapkan pada suatu proses *learning* (Zhang, 2010). Tujuan utama dari proses *learning* ini adalah untuk menciptakan suatu algoritma yang dapat mengenal pola atau bentuk yang kompleks dan membuat suatu keputusan yang cerdas yang didasarkan pada data atau informasi yang ada. Tiga area yang menjadi area penelitian utama pada *machine learning* ialah *Task-oriented Studies*, *Cognitive Simulation*, dan *Theoretical Analysis*.

Machine learning menggunakan algoritma yang mampu menyerupai sistem klasifikasi yang ada pada dunia tumbuhan. Metode-metode yang digunakan dalam *Machine Learning* adalah sebagai berikut (Zhang, 2010).

1. *Supervised learning*: membangkitkan sebuah fungsi yang dapat memetakan data input sesuai keinginan dari data output.
2. *Unsupervised learning*: model dibuat dan diatur berdasarkan data input.
3. *Semi-supervised learning*: gabungan dari *supervised* dan *unsupervised learning* yang digunakan untuk membangkitkan fungsi yang tepat.
4. *Reinforcement learning*: pembelajaran tentang bagaimana perlakuan diberikan, setiap perlakuan mempunyai pengaruh dan terdapat *feedback* untuk menuntun algoritma dari pembelajaran tersebut.
5. *Transduction*: percobaan prediksi data output baru di dasarkan pada pembelajaran input-output dan tes input.
6. *Learning to learn*: pembelajaran guna untuk bisa induktif berdasarkan pengalaman sebelumnya.

Diantara metode yang telah dijelaskan, metode yang paling sering digunakan dalam masalah klasifikasi atau masalah pengenalan pola data adalah metode *supervised learning*. Dengan metode *supervised learning* dapat dibuat suatu prediksi nilai fungsi untuk semua nilai input yang memungkinkan setelah mengalami sejumlah data latihan. Untuk mencapai tujuan ini, mesin harus dapat melakukan proses generalisasi dari data latihan yang diberikan, untuk memprediksikan nilai output dari input yang belum pernah diberikan sebelumnya dengan cara yang masuk akal. Beberapa contoh algoritma *supervised learning* adalah *neural network* (Fernanda & Otok, 2012).

2.2. Neural Network (NN)

Neural network (NN) atau jaringan syaraf tiruan adalah sistem pemroses informasi yang memiliki karakteristik mirip dengan jaringan syaraf biologi (Siang, 2009). NN dibentuk sebagai generalisasi model matematika dari jaringan syaraf biologi, dengan asumsi sebagai berikut (Fausett, 1994).

- a. Pemrosesan informasi terjadi pada elemen sederhana (neuron).
- b. Sinyal dikirimkan diantara neuron-neuron melalui penghubung-penghubung.
- c. Penghubung antar neuron memiliki bobot yang akan memperkuat atau memperlemah sinyal.
- d. Untuk menentukan *output*, setiap neuron menggunakan fungsi aktivasi yang dikenakan pada jumlahan input yang diterima.

Sebelum melakukan pemodelan dengan *Neural network*, diperlukan data *preprocessing* yang bertujuan untuk mengkonversi data input ke dalam rentang tertentu berdasarkan nilai asli dari data input. Tahap *preprocessing* berperan penting dalam NN yaitu untuk menghindari adanya masalah dalam proses komputasi, untuk memenuhi persyaratan algoritma yang digunakan, serta untuk memberikan fasilitas pembelajaran yang lebih baik sehingga diharapkan mempercepat proses pembelajaran *network*.

Neural network ditentukan oleh tiga hal yaitu pola hubungan antar neuron (arsitektur jaringan), metode untuk menentukan bobot penghubung (algoritma) dan fungsi aktivasi yang digunakan. Arsitektur NN terdiri dari tiga jenis lapisan (*layer*), yaitu lapisan *input* (*input layer*), lapisan tersembunyi (*hidden layer*) serta lapisan *output* (*output layer*). Lapisan *input* berfungsi sebagai tempat untuk *input* data, dimana *input* data tersebut diproses lebih lanjut pada lapisan tersembunyi, selanjutnya lapisan *output* menampilkan hasil dari proses tersebut. Terdapat beberapa arsitektur jaringan dalam *neural network* antara lain (Haykin, 1999).

1. Jaringan Lapisan Tunggal (*Single Layer Networks*)

Jaringan ini hanya menerima *input* kemudian secara langsung akan mengolahnya menjadi *output* tanpa harus melalui lapisan lain (lapisan tersembunyi). Ciri-ciri lapisan ini adalah hanya terdiri dari lapisan *input* dan lapisan *output*.

2. Jaringan Banyak Lapisan (*Multilayer Networks*)

Jaringan ini tersusun atas satu lapisan *input*, satu atau lebih lapisan tersembunyi dan satu lapisan *output*. Dalam setiap lapisan terdapat satu atau lebih unit neuron yang menerima *input* dari neuron-neuron pada lapisan sebelumnya kemudian meneruskan hasil pemrosesan ke neuron-neuron pada lapisan sesudahnya.

3. Jaringan *Recurrent* (*Recurrent Networks*)

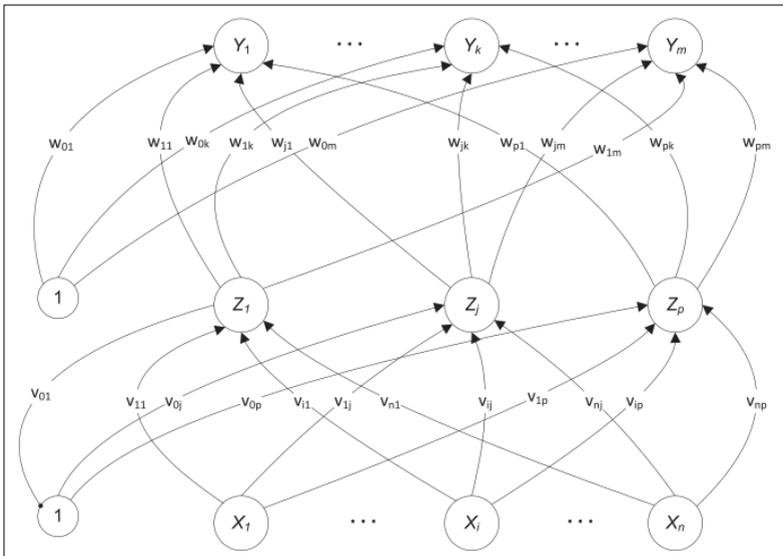
Pada jaringan ini neuron-neuron dapat saling dihubungkan. Sebagai contoh, suatu *Competitive Layer* terdiri dari suatu lapisan dimana setiap neuron mengirimkan kembali sinyal output yang dihasilkan sebagai input pada neuron lainnya.

2.3. Feedforward Neural network (FFNN)

Feedforward Neural network (FFNN) atau *Multilayer Perceptron* merupakan *neural network* yang tersusun dari beberapa *layer*. Struktur FFNN terdiri dari satu *input layer*, satu atau lebih *hidden layer*, dan satu *output layer*. FFNN sukses dalam menyelesaikan beberapa permasalahan yang sulit dipecahkan. Algoritma yang sering digunakan dalam FFNN untuk menyelesaikan masalah adalah *backpropagation* (Fernanda & Otok, 2012).

Fungsi aktivasi merupakan fungsi matematis yang berguna untuk membatasi dan menentukan jangkauan output suatu *neuron*. Fungsi aktivasi untuk jaringan syaraf tiruan *backpropagation* harus memiliki beberapa karakteristik penting, yaitu kontinyu, dapat dideferensialkan, dan monoton tanpa penurunan. Pada umumnya fungsi aktivasi biasa digunakan untuk mencari nilai asimtot maksimum dan minimum. Dalam penelitian ini fungsi aktivasi yang digunakan adalah fungsi sigmoid. Fungsi output dari fungsi sigmoid memiliki nilai pada interval 0 sampai 1, dengan rumus sebagai berikut.

$$y = f(x) = \frac{1}{1 + e^{-ax}} \quad (2.1)$$



Gambar 2.1 Jaringan Syaraf Tiruan Backpropagation dengan 1 Lapis Tersembunyi

Notasi-notasi yang digunakan pada algoritma *Backpropagation* adalah sebagai berikut (Fausett, 1994).

X : vektor input untuk proses pelatihan (*training*).

- t : vektor target output. $t=(t_1, \dots, t_k, \dots, t_m)$
 δ : bobot koreksi error yang telah disesuaikan untuk w_{jk} yang berkaitan dengan error pada unit keluaran Y_k . Simbol ini juga merupakan informasi tentang error pada unit Y_k yang di sebarakan kembali pada unit tersembunyi yang berhubungan dengan unit keluaran Y_k .
 δ_j : bobot koreksi error yang telah disesuaikan untuk v_{ij} yang berkaitan dengan penyebaran kembali informasi error dari layer ouput ke unit tersembunyi Z_j .
 α : *learning rate*.
 X_i : unit input i . Untuk sebuah unit masukan, signal input dan signal output adalah sama dengan X_i .
 v_{oj} : bias pada unit tersembunyi (hidden unit) ke j .
 Z_j : unit tersembunyi pada j . Unit input ke Z_j dinotasikan dengan z_in_j , dengan rumus.

$$z_in_j = v_{oj} + \sum_i^n x_i v_{ij} \quad (2.2)$$

Signal output dari Z_j dinotasikan dengan.

$$z_j = f(z_in_j) \quad (2.3)$$

- w_{ok} : bias pada output unit ke k
 Y_k : ouput unit ke k . Input ke Y_k dinotasikan y_in_k , dengan rumus.

$$y_in_k = w_{ok} + \sum_j^p z_j w_{jk} \quad (2.4)$$

Sinyal output dari Y_k dinotasikan dengan

$$y_k = f(y_in_k) \quad (2.5)$$

Pada jaringan *backpropagation* terdapat beberapa alternatif untuk melakukan pembaharuan bobot, diantaranya adalah pembaharuan bobot standard, pembaharuan bobot dengan momentum, dan pembaharuan bobot dengan deltabar-delta. Pada pembaharuan bobot dengan momentum, perubahan bobot berada pada kombinasi gradien sekarang dan gradien sebelumnya. Dalam hal ini, digunakan laju pembelajaran yang kecil untuk menghindari

gangguan pembelajaran ketika sepasang pola pembelajarannya yang tidak biasa diberikan.

Algoritma pelatihan *backpropagation* terdiri dari dua proses, yaitu *feedforward* dan *backpropagation* dari erornya. Algoritma yang digunakan untuk pelatihan *backpropagation* adalah sebagai berikut.

Langkah 0. Inisialisasi bobot (ambil bobot dengan nilai random yang cukup kecil).

Fase I: Propagasi maju (feedforward)

Langkah 1. Masing-masing input unit ($X_i, i=1, \dots, n$) mengirim sinyal input x_i dan menyebarkannya ke unit-unit pada layer berikutnya (unit-unit pada *hidden layer*).

Langkah 2. Masing-masing *hidden unit* ($Z_j, j=1, \dots, p$) menjumlahkan semua sinyal input yang masing-masing telah dikalikan dengan faktor penimbang atau bobot koneksinya serta ditambah dengan bias:

$$z_in_j = v_{0j} + \sum_{i=1}^n x_i v_{ij} \quad (2.6)$$

Menghitung sinyal output sesuai dengan fungsi aktivasi yang digunakan $z_j=f(z_in_j)$, kemudian mengirim sinyal ini ke semua unit pada layer berikutnya (*output unit*).

Langkah 3. Masing-masing unit output ($Y_k, k=1, \dots, m$) menjumlahkan semua sinyal input yang masing-masing telah dikalikan dengan faktor penimbang atau bobot koneksinya serta ditambah dengan bias:

$$y_in_k = w_{0k} + \sum_{j=1}^n z_j w_{jk} \quad (2.7)$$

Menghitung sinyal output sesuai dengan fungsi aktivasi yang digunakan $y_k=f(y_in_k)$.

Fase II: Backpropagation dari Error

Langkah 4. Menghitung informasi *error* pada masing-masing output unit ($Y_k, k=1, \dots, m$).

$$\delta_k = (t_k - y_k) f'(y_{in_k}) \quad (2.8)$$

Menghitung perbaikan bobot (yang kemudian akan digunakan untuk memperbaiki w_{jk}), dengan laju percepatan α .

$$\Delta w_{jk} = \alpha \cdot \delta_k \cdot z_j \quad (2.9)$$

Menghitung nilai korelasi biasnya.

$$\Delta w_{0k} = \alpha \delta_k \quad (2.10)$$

Dan mengirim δ_k ke semua unit *layer* sebelumnya.

Langkah 5. Setiap unit-unit pada *hidden layer* ($Z_j, j=1, \dots, p$) menjumlahkan δ_k masukannya (dari unit-unit pada *layer* di atasnya).

$$\delta_{in_j} = \sum_{k=1}^m \delta_k w_{jk} \quad (2.11)$$

Kalikan dengan turunan dari fungsi aktivasinya untuk menghitung *error*nya:

$$\delta_j = \delta_{in_j} f'(z_{in_j}) \quad (2.12)$$

Menghitung koreksi bobotnya (digunakan untuk memperbaiki v_{ij})

$$\Delta v_{ij} = \alpha \delta_j x_i \quad (2.13)$$

Menghitung koreksi biasnya (untuk memperbaiki v_{0j})

$$\Delta v_{0j} = \alpha \delta_j \quad (2.14)$$

Fase III: Memperbaiki bobot dan bias:

Langkah 6. Setiap *output unit* ($Y_k, k=1, \dots, m$) memperbaiki bias dan bobotnya ($j=0, \dots, p$):

$$w_{jk}(\text{new}) = w_{jk}(\text{old}) + \Delta w_{jk} \quad (2.15)$$

Setiap *hidden unit* ($Z_j, j=1, \dots, p$) memperbaiki bias dan bobotnya ($i=0, \dots, n$)

$$v_{ij}(\text{new}) = v_{ij}(\text{old}) + \Delta v_{ij} \quad (2.16)$$

Langkah 7. Uji kondisi berhenti.

*Ulangi langkah 2 – 7 sampai kondisi akhir iterasi dipenuhi atau *stopping error* dicapai.

*Untuk masing-masing pasangan data pelatihan (*training data*) lakukan langkah 2 – 6.

Pemilihan bobot akan sangat mempengaruhi jaringan FFNN dalam mencapai global minimum (atau lokal saja) terhadap nilai *error*, dan cepat tidaknya proses pelatihan menuju kondisi konvergen. Apabila nilai bobot awal terlalu besar, maka input ke lapisan tersembunyi atau lapisan output akan jatuh pada daerah dimana turunan fungsi sigmoidnya akan sangat kecil. Sedangkan jika nilai bobot awal terlalu kecil, maka input ke setiap lapisan tersembunyi atau lapisan output akan sangat kecil. Pemilihan bobot awal dapat dilakukan dengan inisialisasi secara random dengan nilai antara -0.5 sampai 0.5 (atau -1 sampai 1, atau interval lainnya).

2.4. Classification and Regression Tree (CART)

CART merupakan salah satu metode *machine learning* dimana metode eksplorasi data dilakukan dengan teknik keputusan. CART termasuk metode statistika non parametrik yang dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen, Charles J. Stone (1984) untuk kasus klasifikasi, baik untuk variabel respon yang bertipe kategorik maupun kontinyu. Jika variabel respon yang digunakan bertipe kontinyu, maka CART akan menghasilkan pohon regresi, sedangkan jika variabel respon yang digunakan bertipe kategorik, maka CART akan menghasilkan pohon klasifikasi. Pohon klasifikasi adalah hasil dari suatu rangkaian perintah pertanyaan dan tipe pertanyaan yang diajukan pada setiap langkah dalam urutan tergantung pada jawaban dari pertanyaan sebelumnya. Rangkaian urutan perintah pertanyaan berakhir pada prediksi dari suatu kelas (Izenman, 2008).

Keunikan yang dimiliki pohon klasifikasi adalah terdapat suatu *root node* yang terdiri dari seluruh kumpulan *learning*. Dalam klasifikasi pohon dikenal istilah *node*, dimana *node* merupakan bagian (*subset*) dari suatu variabel. *Node* dibedakan

menjadi dua macam, yaitu *terminal node* dan *non terminal node*. *Non terminal (parent) node* adalah suatu node yang dapat membagi suatu *node* ke dalam dua *node* yang lain seperti pembagian dengan menggunakan kondisi Boolean dengan dikategorikan ke dalam *satisfied* (“yes”) atau *not satisfied* (“no”). Sedangkan *terminal node* adalah suatu node yang tidak membagi node ke node yang lain dan ditugaskan sebagai kelas label.

Berikut merupakan beberapa kelebihan dari metode CART menurut Rokach dan Maimon (2008).

1. Mampu menjelaskan dirinya sendiri (*self explanatory*) dan mudah untuk diikuti ketika proses penyusunan *decision tree*.
2. Cukup bagus untuk merepresentasikan klasifikasi bernilai kategorik.
3. Dapat mengatasi data-data yang memiliki *error*.
4. Dapat digunakan untuk data yang memiliki nilai *missing*.

Pembentukan pohon klasifikasi diawali dengan menentukan variabel dan nilai dari variabel tersebut untuk dijadikan pemilah tiap simpul. Sedangkan proses akhir pembentukan pohon klasifikasi adalah pemangkasan pohon (*prunning*). Berikut merupakan hal-hal yang harus diperhatikan dalam pembentukan pohon klasifikasi CART (Khasanah, 2015).

a. Pemilihan Pemilah

Dalam tahap pemilihan pemilah bertujuan untuk mengurangi tingkat keheterogenan pada simpul induk dan mendapatkan simpul anak dengan tingkat homogenitas yang tinggi. Data yang digunakan adalah data *learning L* yang kemudian dipilah berdasarkan aturan pemilahan menggunakan kriteria *goodness of split*. Ukuran heterogenitas dapat dinilai dengan indeks gini, persamaan fungsi indeks gini dituliskan sebagai berikut.

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j \quad (2.17)$$

dimana:

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

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

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

Ukuran heterogenitas yang paling sering digunakan adalah indeks Gini. Ukuran heterogenitas ini memiliki keunggulan lebih sederhana sehingga cepat simetri, serta mampu diterapkan untuk permasalahan dengan kelas yang banyak (Breiman, Friedman, Olshen, & Stone, 1984).

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

$$\phi(s,t) = \Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.18)$$

dimana:

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

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

P_L = proporsi pengamatan simpul kiri

P_R = proporsi pengamatan simpul kanan

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

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

Pemilah yang menghasilkan $\phi(s,t)$ lebih tinggi merupakan pemilah terbaik karena dapat mereduksi heterogenitas lebih tinggi. Pengembangan pohon dilakukan dengan pencarian pemilah yang mungkin pada simpul t_1 yang kemudian akan dipilah menjadi t_2 dan t_3 dan seterusnya.

b. Penentuan Terminal Node

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

1. Hanya terdapat satu pengamatan ($n=1$) pada setiap simpul anak atau adanya batasan minimum n dan jika pada simpul terdapat pengamatan berjumlah kurang dari 5 ($n_i < 5$).
2. Jika sudah mencapai batasan jumlah level yang telah ditentukan atau tingkat kedalaman (*depth*) dalam pohon maksimal.

c. Penandaan Label Kelas

Dalam penandaan label kelas pada *terminal nodes* perlu diberikan label kelas sehingga nantinya dapat diketahui karakteristik dari klasifikasi pengamatan untuk setiap kelas variabel respon yang terbentuk. Penentuan label kelas dapat dilakukan berdasarkan aturan jumlah terbanyak, yaitu jika.

$$p(j_0 | t) = \max_j p(j | t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.19)$$

dengan $N_j(t)$ merupakan jumlah total pengamatan dalam *terminal node t* dan $N(t) =$ jumlah total pengamatan dalam *terminal node t*. Label kelas untuk *terminal node t* adalah j_0 yang memberikan nilai dugaan akurasi pengklasifikasian pada simpul t terbesar.

d. Pemangkasan Pohon Klasifikasi

Pohon klasifikasi yang terbentuk pertama kali adalah pohon klasifikasi maksimal yakni dapat diindikasikan pohon yang terbentuk dari hasil pemilahan bisa berukuran sangat besar, karena aturan penghentian pohon hanya berdasarkan pada jumlah pengamatan pada simpul akhir. Banyak pemilah yang terbentuk pada pohon klasifikasi maksimal dapat mengakibatkan makin kecilnya tingkat kesalahan prediksi, oleh karena itu agar dapat mendapatkan pohon yang layak perlu dilakukan pemangkasan pohon (*prunning*) yaitu suatu penilaian ukuran pohon tanpa mengorbankan akurasi yang berarti melalui pengurangan simpul pohon sehingga dicapai ukuran pohon yang layak dan tidak terlalu melebar. Ukuran pohon yang layak dapat dilakukan pemangkasan pohon dengan ukuran *cost complexity*. Metode ini mengandalkan parameter complexity yang dinotasikan dengan α , yang secara perlahan nilainya meningkat selama proses pemangkasan.

Suatu simpul anak akan dipangkas jika hasil perubahan nilai kesalahan klasifikasi prediksi lebih kecil daripada α kali perubahan kompleksitas pohon. Pada sembarang pohon T yang merupakan sub pohon dari pohon klasifikasi maksimal (T_{max}) sehingga $T < T_{max}$, dengan nilai $\alpha \geq 0$, maka persamaan fungsi *cost complexity* adalah sebagai berikut.

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \quad (2.20)$$

dengan:

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

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

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

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

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

$$R_\alpha(T(\alpha)) = \min_{T < T_{max}} R_\alpha(T) \quad (2.21)$$

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

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

2.5. Ilustrasi Pembentukan Pohon Klasifikasi CART

Algoritma CART dalam pembentukan pohon klasifikasi dapat diilustrasikan dengan menggunakan 2 variabel prediktor

yaitu variabel X_1 (Jarak antara *superior border alveolar crest* terhadap *inferior border mandibula*), X_2 (Jarak antara *superior border alveolar crest* dengan *superior border foramen mental*), kemudian satu variabel respon yang terdiri dari 2 kelas yakni kelas 0 adalah jenis kelamin laki-laki dan kelas 2 adalah jenis kelamin perempuan. Variabel yang digunakan dalam ilustrasi CART dijelaskan dalam Tabel 2.4 sebagai berikut.

Tabel 2.1 Variabel Penelitian Ilustrasi CART

	Variabel	Skala	Keterangan
X_1	Jarak antara <i>superior border alveolar crest</i> terhadap <i>inferior border mandibula</i>	Rasio	Satuan <i>centimeter</i>
X_2	Jarak antara <i>superior border alveolar crest</i> dengan <i>superior border foramen mental</i>	Rasio	Satuan <i>centimeter</i>
Y	Jenis kelamin	Nominal	0 : laki-laki 1: perempuan

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

Tabel 2.2 Contoh Data Sampel Ilustrasi CART

Data Ke-	Y	X_1	X_2
1	0	3	5
2	1	6	10
3	0	2	6
4	0	4	10
5	0	6	12
6	1	2	6
7	0	4	5
8	1	3	10
9	1	3	12
10	1	4	6

Data sampel pengamatan tersebut akan membentuk simpul induk (*parent nodes*) yang kemudian akan memilah menjadi simpul anak (*child nodes*). Algoritma CART yang pertama adalah

pembentukan pohon klasifikasi, dengan pemilihan pemilah menggunakan indeks gini. Sebelum melakukan proses pemilihan pemilah dalam analisis CART, terlebih dahulu menghitung nilai keheterogenan simpul induk berdasarkan fungsi keheterogenan indeks gini yang telah didefinisikan pada persamaan (2.3). Pada saat memilah simpul induk menjadi dua simpul anak, dilakukan penentuan pemilah dari semua kemungkinan variabel dan nilai variabel (*threshold*). Sebelumnya mencari nilai *goodness of split* menggunakan persamaan (2.5) dari kemungkinan *threshold* pada setiap variabel yang dihasilkan kemudian dilakukan perbandingan dan memilih nilai *threshold* pada variabel yang memiliki *goodness of split* yang paling tinggi berdasarkan persamaan (2.6). *Threshold* variabel terpilih tersebut selanjutnya dijadikan sebagai pemilah terbaik untuk simpul awal. Berikut ini merupakan contoh perhitungan manual dari proses pemilihan pemilah terbaik.

a. Perhitungan Indeks Gini untuk Variabel Y

Tabel 2.3 Proporsi Pengamatan pada Respon Y

Variabel respon (Y)	$N_i(t)$	$p(i t)$
Kelas 0	5	$5/10 = 0,5$
Kelas 1	5	$5/10 = 0,5$
Total	10	1

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

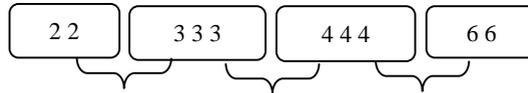
$$i(t) = p(0|t)p(1|t)]$$

$$i(t) = (0,5)(0,5)$$

$$= 0,25$$

b. Penentuan *Threshold* dan Perhitungan *Goodness of Split* Variabel X_1 (Jarak antara *superior border alveolar crest* terhadap *inferior border mandibula*)

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



Nilai median = 2,5 3,5 5

Threshold untuk variabel X_1 dengan nilai median 2,5 maka untuk data $X_1 \leq 2,5$ akan dipilah ke simpul kiri sedangkan data $X_1 > 2,5$ akan dipilah ke simpul kanan. Perhitungannya pada Tabel 2.4 sebagai berikut.

Tabel 2.4 Perhitungan *Goodness of Split* untuk *Threshold* $X_1=2,5$

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

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

Simpul kiri : $i(t_L) = (1/2)(1/2) = 0,25$

Simpul kanan : $i(t_R) = (4/8)(4/8) = 0,25$

Maka, perhitungan *goodness of split* untuk *threshold* $X_1=2,5$ adalah sebagai berikut.

$$\phi(s,t) = 0,25 - 0,2(0,25) - 0,8(0,25) = 0$$

Pada *threshold* dengan nilai median 3,5 pada data $X_1 \leq 3,5$ akan dipilah ke simpul kiri sedangkan data $X_1 > 3,5$ akan dipilah ke simpul kanan. Perhitungannya akan ditampilkan pada Tabel 2.5 sebagai berikut.

Tabel 2.5 Perhitungan *Goodness of Split* untuk *Threshold* $X_1=3,5$

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

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

Simpul kiri : $i(t_L) = (2/5)(3/5) = 0,24$

Simpul kanan : $i(t_R) = (3/5)(2/5) = 0,24$

Maka, perhitungan *goodness of split* untuk *threshold* $X_1=3,5$ adalah sebagai berikut.

$$\phi(s,t) = 0,25 - 0,5(0,24) - 0,5(0,24) = 0,01$$

Threshold untuk variabel X_1 dengan nilai median 5 maka untuk data $X_1 \leq 5$ akan dipilah ke simpul kiri sedangkan data $X_1 > 5$ akan dipilah ke simpul kanan. Perhitungannya pada Tabel 2.6 sebagai berikut.

Tabel 2.6 Perhitungan *Goodness of Split* untuk *Threshold* $X_1=5$

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

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

Simpul kiri : $i(t_L) = (4/8)(4/8) = 0,25$

Simpul kanan : $i(t_R) = (1/2)(1/2) = 0,25$

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

$$\phi(s,t) = 0,25 - 0,8(0,25) - 0,2(0,25) = 0$$

c. Penentuan *Threshold* dan Perhitungan *Goodness of Split* Variabel X_2 (Jarak antara *superior border alveolar crest* dengan *superior border foramen mental*)

Variabel X_2 memiliki skala pengukuran kontinyu, sehingga *threshold* pada variabel X_2 berupa nilai tengah (median). Diperoleh kemungkinan nilai median sebagai berikut.



Nilai median = 5,5 8 11

Threshold untuk variabel X_2 dengan nilai median 5,5 maka untuk data $X_2 \leq 5,5$ akan dipilah ke simpul kiri sedangkan data $X_2 > 5,5$ akan dipilah ke simpul kanan. Perhitungannya pada Tabel 2.7 sebagai berikut.

Tabel 2.7 Perhitungan *Goodness of Split* untuk *Threshold* $X_2=5,5$

Simpul kiri		Simpul kanan	
$N_L = 2$	$p_L = 0,2$	$N_R = 8$	$p_R = 0,8$
$N(0 t_L)=2$	$p(0 t_L)=2/2$	$N(0 t_R)=3$	$p(0 t_R)=3/8$
$N(1 t_L)=0$	$p(1 t_L)=0$	$N(1 t_R)=5$	$p(1 t_R)=5/8$

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

Simpul kiri : $i(t_L) = (2/2)(0) = 0$

Simpul kanan : $i(t_R) = (3/8)(5/8) = 0,234$

Maka, perhitungan *goodness of split* untuk *threshold* $X_2=5,5$ adalah sebagai berikut.

$$\phi(s,t) = 0,25 - 0,2(0) - 0,8(0,234) = 0,0625$$

Pada *threshold* dengan nilai median 8 pada data $X_2 \leq 8$ akan dipilah ke simpul kiri sedangkan data $X_2 > 8$ akan dipilah ke simpul kanan. Perhitungannya akan ditampilkan pada Tabel 2.8 sebagai berikut.

Tabel 2.8 Perhitungan *Goodness of Split* untuk *Threshold* $X_2=8$

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

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

Simpul kiri : $i(t_L) = (3/5)(2/5) = 0,24$

Simpul kanan : $i(t_R) = (2/5)(3/5) = 0,24$

Maka, perhitungan *goodness of split* untuk *threshold* $X_2=8$ adalah sebagai berikut.

$$\phi(s,t) = 0,25 - 0,5(0,24) - 0,5(0,24) = 0,01$$

Threshold untuk variabel X_2 dengan nilai median 11 maka untuk data $X_2 \leq 11$ akan dipilah ke simpul kiri sedangkan data $X_2 > 11$ akan dipilah ke simpul kanan. Perhitungannya pada Tabel 2.9 sebagai berikut.

Tabel 2.9 Perhitungan *Goodness of Split* untuk *Threshold* $X_2=11$

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

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

Simpul kiri : $i(t_L) = (4/8)(4/8) = 0,25$

Simpul kanan : $i(t_R) = (1/2)(1/2) = 0,25$

Maka, perhitungan *goodness of split* untuk *threshold* $X_2=11$ adalah sebagai berikut.

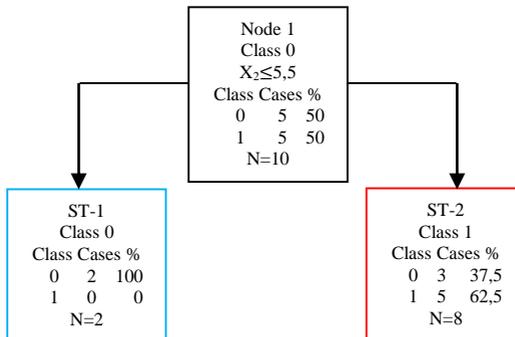
$$\phi(s,t) = 0,25 - 0,8(0,25) - 0,2(0,25) = 0$$

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

Tabel 2.10 Perbandingan *Goodness of Split* dari Pemilah 1

Variabel	Threshold	<i>Goodness of Split</i> ($\phi(s,t)$)
X ₁	2,5	0
	3,5	0,01
	5	0
X ₂	5,5	0,0625*
	8	0,01
	11	0

Berdasarkan Tabel 2.10, dapat diketahui pemilah terbaik dipilih dari *threshold* dengan *goodness of split* terbesar. Nilai *goodness of split* terbesar terdapat pada variabel pemilah X₂ (Jarak antara *superior border alveolar crest* dengan *superior border foramen mental*) dengan *threshold* 5,5. Selanjutnya, dilakukan pemilahan pada data sampel berdasarkan pemilah terbaik atau pemilah utama tersebut. Sehingga data sampel pada variabel X₂ yang kurang dari 5,5 masuk ke dalam simpul kiri sedangkan data yang lebih dari 5,5 masuk dalam simpul kanan. Berikut ini merupakan hasil pembentukan pohon berdasarkan pemilahan utama yang ditunjukkan pada Gambar 2.6 sebagai berikut.



Gambar 2.2 Pohon Klasifikasi Ilustrasi Hasil Pemilahan I

Berdasarkan Gambar 2.6, dapat diketahui simpul kanan yang dihasilkan terdiri dari 8 data dan belum homogen, untuk itu dilakukan pemilahan kembali secara *recursive* hingga didapatkan pohon klasifikasi maksimal dengan simpul yang homogen. Data sampel yang berasal dari simpul kanan digunakan untuk pemilahan kedua yang ditunjukkan pada Tabel 2.11 sebagai berikut.

Tabel 2.11 Data Sampel Ilustrasi CART Pemilahan 2

Data Ke-	Y	X ₁	X ₂
2	1	6	10
3	0	2	6
4	0	4	10
5	0	6	12
6	1	2	6
8	1	3	10
9	1	3	12
10	1	4	6

Hasil perbandingan *goodness of split* setiap kemungkinan *threshold* dari sampel pada simpul kanan yang digunakan sebagai pemilah kedua ditunjukkan dalam Tabel 2.12 sebagai berikut.

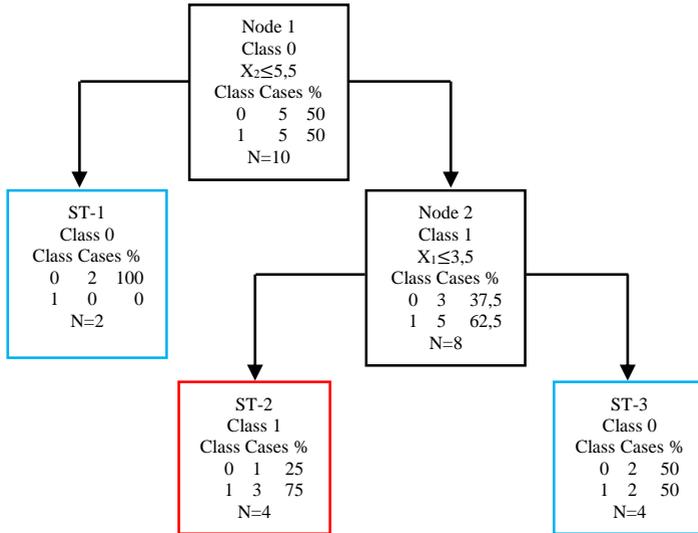
Tabel 2.12 Perbandingan *Goodness of Split* dari Pemilah 2

Variabel	Threshold	<i>Goodness of Split</i> ($\phi(s,t)$)
X ₁	2,5	0,05068
	3,5	0,05937*
	5	0,05068
X ₂	5,5	0,04775
	8	0,05117

**goodness of split* terbesar

Berdasarkan Tabel 2.12 dapat diketahui nilai *goodness of split* terbesar yakni 0,05937 diperoleh pada variabel X₁ dengan *threshold* sebesar 3,5, sehingga pemilah terbaik yang dihasilkan adalah variabel X₁ dan berperan sebagai pemilahan II. Oleh karena itu, data sampel berjumlah 8 data akan dipilah berdasarkan pemilahan 2 menjadi dua anak simpul, dimana data variabel X₁ yang bernilai kurang dari 3,5 akan masuk ke dalam simpul kiri dan sampel dengan nilai lebih dari 3,5 akan masuk ke simpul kanan.

Sehingga, hasil pemilahan untuk simpul kanan pada kedalaman 2 dapat digambarkan pada pohon klasifikasi berikut ini.



Gambar 2.3 Pohon Klasifikasi Hasil Pemilahan II

Berdasarkan Breiman, Friedman, Olshen, dan Stone (1984) pembentukan pohon akan berhenti ketika jumlah anggota pada simpul terminal berjumlah kurang dari sama dengan 5 pengamatan atau dengan jumlah *nodes* minimum yang telah ditentukan. Maka pohon yang terbentuk pada Gambar 2.7 dapat dikatakan sebagai pohon maksimal dan simpul terminal yang terbentuk adalah sebanyak 3 ditunjukkan oleh kotak yang berwarna merah (jenis kelamin perempuan) dan biru (jenis kelamin laki-laki).

Selanjutnya dilakukan pelabelan kelas pada tiap *terminal nodes* untuk mengetahui karakteristik masing-masing kelas pada variabel respon. Sehingga dapat memudahkan dalam hal pengklasifikasian dengan cara melihat karakteristik yang telah dihasilkan. Pemberian label kelas yaitu label kelas 0 untuk jenis kelamin laki-laki dan label kelas 1 untuk jenis kelamin perempuan yang ditentukan berdasarkan persamaan (2.7) dengan mencari nilai

peluang kelas yang tertinggi. Penandaan label pada simpul terminal dapat dilihat pada Tabel 2.14 berikut.

Tabel 2.13 Pemberian Label pada Simpul Terminal 1

Label Kelas	Simpul Terminal	Presentase
Laki-Laki	1	100%
(2 Simpul Terminal)	3	50%
Perempuan	2	75%
(1 simpul terminal)		

Karakteristik jenis kelamin dapat dilihat pada simpul terminal yang memiliki persentase kelas terbesar. Karakteristik kelas 0 (Laki-Laki) dapat diketahui dengan menelusuri simpul terminal 1 dan karakteristik kelas 1 (perempuan) dapat diketahui dengan menelusuri simpul terminal 2. Sehingga dengan menelusuri aturan pemilahan pada simpul terminal 1 dan 2 akan didapatkan karakteristik jenis kelamin

2.6. Boosting

Salah satu variasi dari metode *Ensemble* yang paling banyak digunakan adalah *Boosting*, dimana *Boosting* merupakan metode umum yang digunakan untuk meningkatkan performansi dari suatu algoritma pembelajaran. Fungsi dari metode ini adalah untuk menemukan klasifikasi yang sangat akurat pada data pelatihan, dengan menggabungkan “*weak hypothesis*”, yang masing-masing hanya perlu menjadi cukup akurat pada data pelatihan (Schapire, 1990). Salah satu ide utama dari algoritma *boosting* adalah untuk mempertahankan distribusi atau bobot dari data *training*. Bobot dari distribusi data *training* dinotasikan dengan $D_t(i)$. *Boosting* dimulai dengan memberikan bobot yang sama pada semua data *training*, kemudian menentukan *base learner* atau *weak learner*. Salah satu tugas dari *weak learner* adalah untuk mencari *weak hypothesis* $h_t: x \rightarrow \{-1, +1\}$ yang sesuai dengan distribusi w_t (Freund & Schapire, 1999).

Diberikan sekumpulan data sampel pelatihan $\{y_i, x_i\}_1^N$, kemudian akan dicari fungsi $F^*(x)$ yang memetakan x ke y , dengan nilai yang diharapkan dari beberapa fungsi kerugian $\Psi(y, F(x))$ adalah minimum. Pada umumnya fungsi $F(x)$ adalah anggota dari

kelas parameter fungsi $F(x;P)$. Fungsi $F(x;P)$ dirumuskan sebagai berikut.

$$F(x;P) = \sum_{m=0}^M \beta_m h(x; a_m) \quad (2.22)$$

Dimana $P = \{\beta_m, a_m\}_0^M$ adalah kumpulan dari parameter. Kemudian $F(x)$ dievaluasi pada setiap x sebagai parameter dan meminimumkan:

$$\Phi(F(x)) = E_y[\Psi(y, F(x) | x)] \quad (2.23)$$

pada masing-masing x dengan memperhatikan $F(x)$. Solusinya adalah dari jenis:

$$F^*(x) = \sum_{m=0}^M f_m(x) \quad (2.24)$$

Dimana $f_0(x)$ adalah dugaan awal dan $\{f_m\}_1^M$ adalah fungsi tambahan, dikenal sebagai “steps” atau “boosts”. Dengan menggunakan *steepest-descent* didapatkan:

$$f_m(x) = -\rho_m g_m(x) \quad (2.25)$$

dimana, disini diasumsikan bahwa diferensiasi dan integrasi dapat ditukar,

$$g_m(x) = E_y \left[\frac{\partial \Psi(y, F(x))}{\partial F(x)} \Big| x \right]_{F(x)=F_{m-1}(x)} \quad (2.26)$$

dan

$$F_{m-1}(x) = \sum_{i=0}^{m-1} f_i(x) \quad (2.27)$$

ketika distribusi gabungan dari (y,x) di representasikan dengan data sampel yang terbatas, maka $E_y[\cdot|x]$ tidak dapat dievaluasi secara akurat pada setiap x_i dan jika akan dilakukan optimasi parameter maka solusi akan sulit untuk didapatkan. Pada penelitian ini, diberikan pendekatan $F_{m-1}(x)$ pada iterasi m -th, fungsi $\beta_m h(x;a)$ adalah tahap dari metode *greedy* yang terbaik terhadap solusi untuk meminimumkan $F^*(x)$, berdasarkan batasan bahwa pada tahap $h(x,a_m)$ menjadi anggota dari kelas parameter fungsi $h(x,a)$. Salah

satu kemungkinan adalah dengan memilih anggota dari kelas paramater $h(x;a)$ yang paling paralel di ruang data berdimensi N dengan gradien negatif $\{-g_m(x_i)\}_1^N$. Hal ini dimungkinkan untuk menggunakan $h(x,a_m)$ daripada gradien negatif $-g_m(x_i)$. Bobot dari ρ_m dirumuskan sebagai berikut.

$$\rho_m = \arg \min_{\rho} \sum_{i=1}^N \Psi(y_i, F_{m-1}(x_i) + \rho h(x_i; a_m)) \quad (2.28)$$

Ketika $y \in \{-1, 1\}$ dan fungsi kerugian $\Psi(y, F)$ bergantung pada y dan F hanya melalui $\Psi(y, F) = \Psi(yF)$, algoritma dapat mengurangi *Boosting*. Jika fungsi kerugian adalah $\Psi(y, F) = (y - F)^2 / 2$ maka gradien *boosting* menghasilkan pendekatan *stagewise* dari iteratif residual yang sesuai (Okun, Valentini, & Re, *Ensembles in Machine Learning Applications*, 2011).

Berikut merupakan algoritma *adaboost* yang digunakan dalam penelitian ini.

Input:

Dataset $D = (x_1, y_1), \dots, (x_m, y_m)$; $x_i \in X$, $y_i \in Y = \{-1, +1\}$

Proses:

- (i) Menginisialisasi bobot sampel data $D_1(i) = 1/m$
- (ii) Untuk $t = 1, \dots, T$:
 1. Melakukan prediksi h_t untuk dataset D yang telah diboboti D_t : $H_t = L(D, D_t)$
 2. Menghitung bobot D_t dari data yang salah diklasifikasikan
 3. Menghitung bobot voting $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right)$
 4. Memperbaharui bobot sampel data:

$$\begin{aligned} D_{t+1(i)} &= \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{jika } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{jika } h_t(x_i) \neq y_i \end{cases} \quad (2.29) \\ &= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \end{aligned}$$

dimana Z_i adalah faktor normalisasi agar $\sum_{i=1}^m D_{t+1}(i) = 1$

(iii) Selesai.

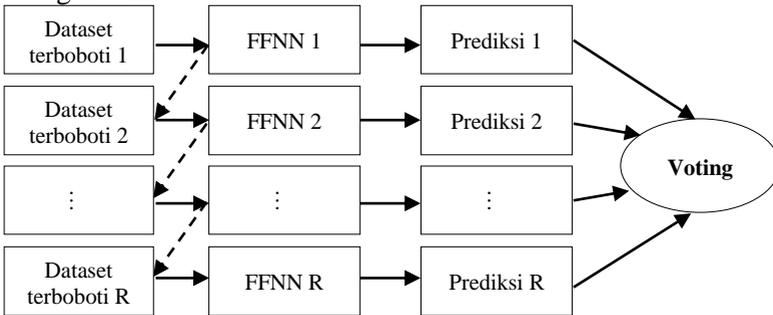
Output:

Voting prediksi klasifikasi akhir

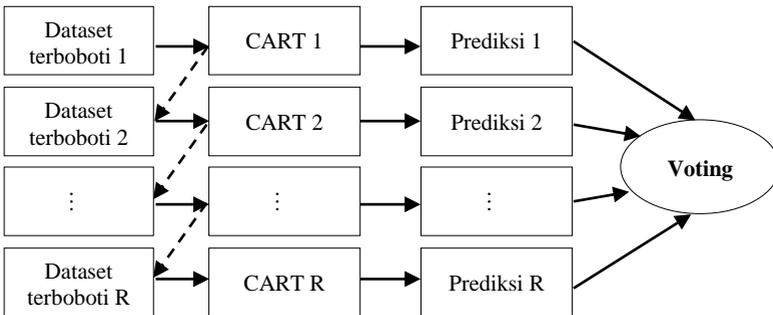
$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad (2.30)$$

2.7. Ilustrasi Boosting

Ilustrasi penerapan algoritma Boosting pada suatu *Dataset* dengan metode FFNN dan metode CART dapat digambarkan sebagai berikut.



Gambar 2.4 Ilustrasi Boosting CART



Gambar 2.5 Ilustrasi Boosting CART

Langkah pertama adalah pemberian bobot D_1 pada masing-masing sampel data dengan nilai yang sama. Selanjutnya saat dilakukan pembentukan pohon dengan algoritma CART, bobot D_1 digunakan dalam perhitungan indeks gini sebagai pengganti $p(j/t)$ dan saat dilakukan pembentukan model (struktur jaringan) FFNN, bobot D_1 digunakan dalam inisialisasi bobot awal agar jaringan FFNN dapat mencapai global minimum terhadap nilai *error*. Setelah pohon klasifikasi dan struktur jaringan FFNN selesai dibentuk, dilakukan prediksi kelas pada data sampel. Berdasarkan hasil prediksi tersebut, dihitung jumlah bobot dari sampel yang salah diklasifikasikan untuk menghitung bobot voting.

Hasil prediksi awal diamati untuk menentukan bobot pada replikasi selanjutnya D_2 . Bobot sampel yang salah diklasifikasikan selanjutnya bernilai lebih besar, sedangkan sampel yang benar di klasifikasikan bernilai lebih kecil dengan perhitungan sesuai persamaan (2.29). Bobot baru inilah yang akan digunakan untuk pembentukan pohon klasifikasi pada replikasi selanjutnya. Langkah-langkah tersebut diulang kembali sampai sejumlah replikasi yang diinginkan. Setelah replikasi selesai, dilakukan proses voting sesuai persamaan (2.30) dengan melihat tanda nilai tersebut.

2.8. Ukuran Ketepatan Klasifikasi

Metode *K-folds Cross Validation* (KCV) merupakan suatu metode yang dapat diandalkan (*reliable*) untuk memprediksi kesalahan dalam suatu klasifikasi. Metode ini banyak digunakan oleh peneliti untuk mengurangi bias yang terjadi karena pengambilan sampel data yang akan digunakan. KCV secara berulang-ulang membagi data menjadi data *training* dan data *testing*, dimana setiap data berkesempatan menjadi data *testing*. Akurasi atau ketepatan klasifikasi yang dihasilkan dengan menggunakan metode KCV adalah sebagai berikut.

$$CVA = \frac{1}{K} \sum_{k=1}^K A_k \quad (2.31)$$

dimana K adalah banyaknya *fold* yang digunakan, dan A_k adalah hasil akurasi yang dihasilkan pada setiap *fold*. Tahapan yang dilakukan dalam menggunakan KCV adalah sebagai berikut.

1. Tentukan nilai K atau banyak *fold*.
2. Melakukan pengambilan data untuk data *training*.
3. Membuat model dengan menggunakan data *training*.
4. Menghitung performa model.
5. Ulangi langkah 2 hingga langkah 4 sebanyak K .
6. Hitung rata-rata performa metode.

Pada kasus klasifikasi dua kelas, model klasifikasi menghasilkan prediksi yang merupakan peluang keanggotaan suatu kelas yang merupakan prediksi dari model tersebut. Perbandingan kelas aktual dengan kelas prediksi dari model akan menghasilkan empat kemungkinan yang terjadi, yaitu.

1. Jika aktual *positive* dan prediksi *positive* maka *true positive*
2. Jika aktual *positive* dan prediksi *negative* maka *false positive*
3. Jika aktual *negative* dan prediksi *negative* maka *true negative*
4. Jika aktual *negative* dan prediksi *positive* maka *false negative*.

Kemungkinan-kemungkinan yang terjadi diatas dapat diperlihatkan dalam suatu *confusion matrix* berikut ini.

Tabel 2.14 *Confusion Matrix*

Kelas Prediksi	Kelas Aktual	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	TP	FP
<i>Negative</i>	FN	TN
Total	P	N

Performa dari suatu metode klasifikasi dapat diukur dengan menggunakan *accuracy*, *sensitivity*, dan *specificity*. *Accuracy* adalah banyaknya pengamatan yang tepat diklasifikasikan sesuai kategorinya, *sensitivity* adalah banyaknya pengamatan yang berkategori positif yang tepat diklasifikasikan, dan *specificity* adalah banyaknya pengamatan yang berkategori negatif yang tepat diklasifikasikan. Perhitungan dilakukan dgn persamaan berikut.

$$Sensitivity = \frac{TP}{P} \quad (2.32)$$

$$\text{Specificity} = \frac{TN}{N} \quad (2.33)$$

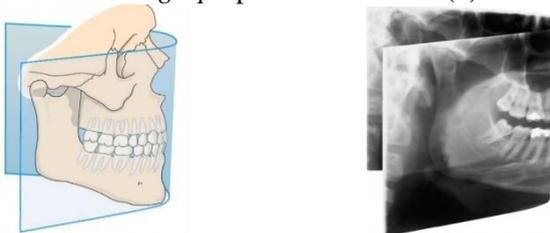
$$\text{Accuracy} = \frac{TP + TN}{P + N} \quad (2.34)$$

dimana TP adalah pengamatan berkategori positif tepat diklasifikasikan kategori positif, FP adalah pengamatan berkategori positif diklasifikasikan kategori negatif, TN adalah pengamatan berkategori negatif tepat diklasifikasikan kategori negatif, dan FN adalah pengamatan berkategori negatif diklasifikasikan kategori positif.

2.9. Panoramic radiograph

Radiograph adalah alat yang sangat diperlukan dalam antropologi forensik. Radiografi menyediakan kemampuan untuk melihat seluruh bagian mandibula dan memungkinkan untuk mendapatkan lokasi yang lebih akurat dari foramen mental. Pada bidang kesehatan, radiografi dibuat menggunakan radiasi *electromacnetic X-rays* untuk melihat isi dalam tubuh manusia, seperti tulang, otak, *dental*, dan *fetus*. (Chandra et. al, 2013)

Panoramic radiograph adalah teknik yang digunakan secara luas karena memiliki keuntungan, diantaranya adalah dengan dosis radiasi yang relatif rendah pada pasien dan biaya lebih rendah jika dibandingkan dengan teknik yang lebih canggih. Keakuratan pengukuran pada radiografi didasarkan pada kualitas radiografi (Singal & Sharma, 2016). Ilustrasi posisi dalam pengambilan *panoramic radiograph* dapat dilihat pada Gambar 2.1 (a) dan hasil dari *panoramic radiograph* pada Gambar 2.1 (b).

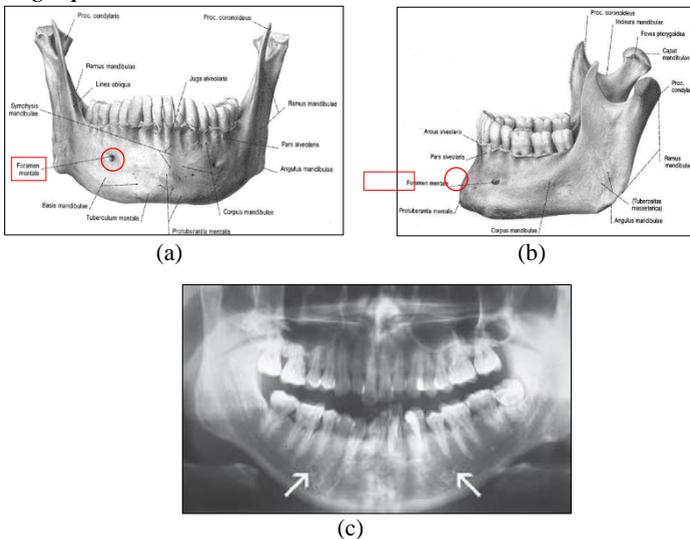


Gambar 2.6 (a) Posisi pengambilan radiografi panoramik, (b) Hasil radiografi panoramik

2.10. Foramen Mental

Foramen mental adalah suatu saluran terbuka pada korpus mandibular tempat keluarnya arteri, vena, dan nervus mental yang merupakan cabang nervus alveolaris inferior. Mengetahui lokasi foramen mental sangat penting ketika mempertimbangkan prosedur tindakan bedah mulut dan prosedur invasif lainnya di sekitar foramen mental (Kholilullah, 2016). Foramen mental merupakan salah satu karakteristik mandibula yang dapat digunakan untuk estimasi usia dan penentuan jenis kelamin. Perkembangan foramen mental dipengaruhi berbagai faktor, antara lain faktor usia, jenis kelamin dan ras, oleh karena itu foramen mental merupakan salah satu karakteristik mandibula yang dapat digunakan untuk penentuan jenis kelamin (Gupta & Soni, 2012).

Gambar 2.2 (a) merupakan tulang mandibula tampak dari depan, Gambar 2.2 (b) merupakan tulang mandibula tampak dari samping kiri, dan Gambar 2.2 (c) merupakan hasil *panoramic radiograph* dari foramen mental.



Gambar 2.7 (a) Tulang mandibula tampak depan, (b) Tulang mandibula tampak samping kiri, (c) Panoramic radiograph pada foramen mental

BAB III METODOLOGI PENELITIAN

3.1. Sumber Data

Penelitian ini menggunakan data sekunder, dimana data yang digunakan didapatkan dari penelitian yang dilakukan oleh Afrista pada tahun 2016 yang berjudul “Determinasi Jenis Kelamin dengan Metode Pengukuran Linear pada Foramen Mental menggunakan Radiografi Panoramik”. Data yang digunakan merupakan *panoramic radiograph* yang diambil dari 62 orang yang terdiri dari 31 laki-laki dan 31 perempuan yang berumur lebih dari 21 tahun.

3.2. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini merupakan variabel yang berisi informasi mengenai hasil pengamatan radiograf yang terdiri dari 3 variabel prediktor dan 1 variabel respon. Berikut adalah variabel penelitian yang digunakan.

Tabel 3.1 Variabel yang digunakan dalam penelitian

	Variabel	Skala	Keterangan
X ₁	Jarak antara <i>superior border</i> alveolar <i>crest</i> terhadap <i>inferior border</i> mandibula	Rasio	Satuan <i>centimeter</i>
X ₂	Jarak antara <i>superior border</i> alveolar <i>crest</i> dengan <i>superior border</i> foramen mental	Rasio	Satuan <i>centimeter</i>
X ₃	Jarak antara <i>inferior border</i> foramen mental dengan <i>inferior border</i> mandibula	Rasio	Satuan <i>centimeter</i>
Y	Jenis kelamin	Nominal	0 : laki-laki 1 : perempuan

3.3. Definisi Operasional

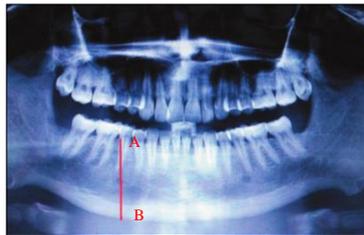
Cara pengukuran untuk variabel penelitian adalah hasil *panoramic radiograph* di klip dengan kertas mika kemudian diletakkan diatas *viewer*, setelah itu, ditarik garis vertikal antara

superior border alveolar crest hingga *inferior border* mandibula melalui titik tengah foramen mental (jarak A-B), garis vertikal antara *alveolar crest* dengan *superior border* foramen mental (jarak A-C), dan garis vertikal antara *inferior border* foramen mental dengan *inferior border* mandibula (jarak D-B), kemudian diukur menggunakan kaliper atau jangka sorong.

Pada penelitian ini definisi yang digunakan dalam menjelaskan variabel bersumber dari *Indian Journal of Forensic Medicine & Toxicology* oleh Kusum Singal dan Sapna Sharma. Berikut merupakan definisi mengenai variabel penelitian yang digunakan.

1. Jarak *superior border alveolar crest* terhadap *inferior border* mandibula (jarak A-B)

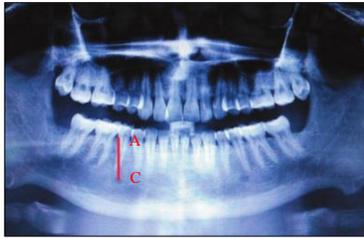
Jarak *superior border alveolar crest* terhadap *inferior border* mandibula merupakan jarak yang diukur dari *superior border alveolar crest* yang ditunjukkan oleh huruf A hingga *inferior border* mandibula yang ditunjukkan oleh huruf B. Berikut adalah *panoramic radiograph* dari foramen mental untuk menunjukkan jarak yang telah dijelaskan.



Gambar 3.1 Jarak A-B

2. Jarak antara *superior border alveolar crest* dengan *superior border* foramen mental (jarak A-C)

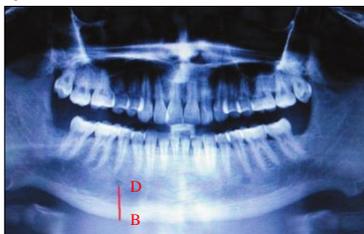
Jarak *superior border alveolar crest* dengan *superior border* foramen mental merupakan jarak yang diukur dari *superior border alveolar crest* yang ditunjukkan oleh huruf A hingga *superior border* foramen mental yang ditunjukkan oleh huruf C. Berikut adalah *panoramic radiograph* dari foramen mental untuk menunjukkan jarak yang telah dijelaskan.



Gambar 3.2 Jarak A-C

3. Jarak antara *inferior border* foramen mental dengan *inferior border* mandibula (jarak D-B)

Jarak *inferior border* foramen mental dengan *inferior border* mandibula merupakan jarak yang diukur dari *inferior border* foramen mental yang ditunjukkan oleh huruf D hingga *inferior border* mandibula yang ditunjukkan oleh huruf B. Berikut adalah *panoramic radiograph* dari foramen mental untuk menunjukkan jarak yang telah dijelaskan.



Gambar 3.3 Jarak D-B

3.4. Langkah Analisis

Tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut.

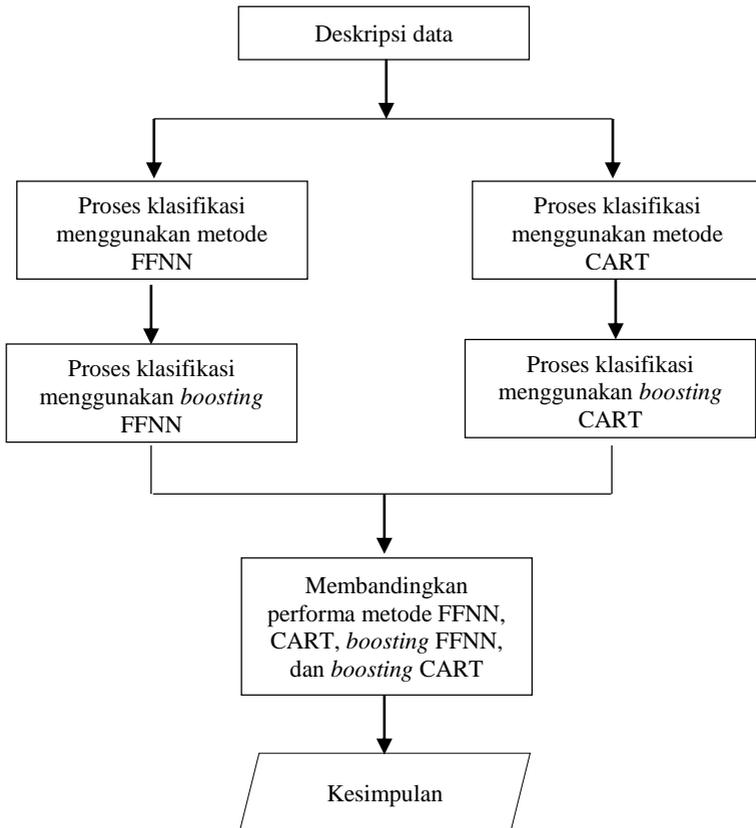
1. Mendeskripsikan data secara visual menggunakan *boxplot* dan *scatterplot*.
2. Melakukan klasifikasi dengan menggunakan metode *Feedforward Neural network* (FFNN).
3. Melakukan klasifikasi dengan menggunakan metode *Classification and Regression Tree* (CART).

4. Melakukan proses *boosting* pada metode *Feedforward Neural network* (FFNN), dengan tahapan sebagai berikut.
 - 1) Melakukan klasifikasi data dengan metode FFNN dengan banyak *neuron* adalah 1 sampai 10.
 - 2) Menentukan hasil klasifikasi FFNN yang terbaik berdasarkan banyak *neuron* pada *hidden layer*.
 - 3) Melakukan pemberian bobot pada variabel respon. Variabel prediktor yang digunakan dalam proses *boosting* adalah variabel prediktor yang juga digunakan dalam FFNN.
 - 4) Mengklasifikasikan variabel respon yang telah diboboti dengan variabel prediktor dengan menggunakan metode FFNN dengan banyak *neuron* yang telah didapatkan pada tahap kedua.
 - 5) Menghitung pengamatan yang misklasifikasi, kemudian memperbaharui bobot pada variabel respon dari pengamatan yang misklasifikasi.
 - 6) Mendapatkan prediksi akhir (voting) dan menghitung tingkat akurasi dari *boosting* FFNN.
5. Melakukan proses *boosting* pada metode *Classification and Regression Tree* (CART), dengan tahapan sebagai berikut.
 - 1) Memilah *learning* sampel yang masih heterogen menjadi dalam himpunan bagian yang lebih homogen dengan menggunakan Indeks Gini.
 - 2) Menentukan kriteria *Goodness of Split*.
 - 3) Penghentian pohon (*Stopping Rule*).
 - 4) Melakukan pemangkasan (*Prunning*) pada pohon yang telah terbentuk berdasarkan *cost complexity minimum*.
 - 5) Mengklasifikasikan variabel respon yang telah diboboti dengan variabel prediktor dengan metode CART.
 - 6) Menghitung pengamatan yang misklasifikasi, kemudian memperbaharui bobot pada variabel respon dari pengamatan yang misklasifikasi.
 - 7) Mendapatkan prediksi akhir (voting) dan menghitung tingkat akurasi dari *boosting* CART.

6. Membandingkan performa dari metode FFNN dan CART yang telah didapat. Ukuran performa yang digunakan sebagai pembanding adalah *accuracy*, *sensitivity*, dan *specitivity*.

3.5. Diagram Alir

Diagram alir unruk langkah-langkah yang akan dilakukan dalam penelitian ini adalah sebagai berikut.



Gambar 3.4 Langkah-langkah yang dilakukan dalam penelitian

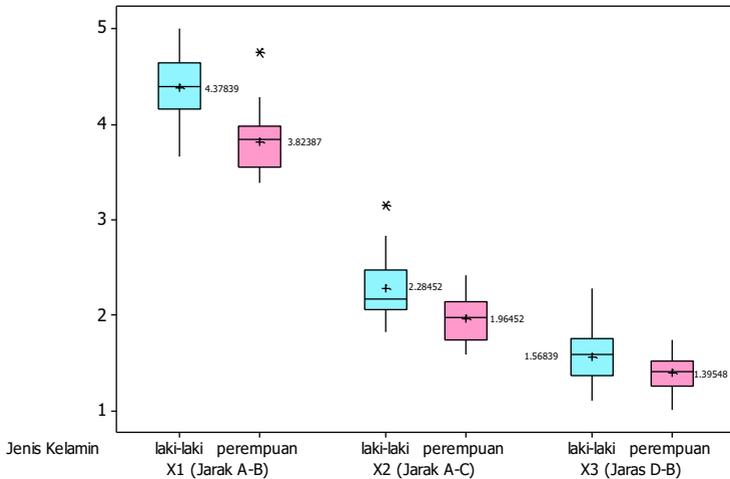
(Halaman ini sengaja dikosongkan)

BAB IV ANALISIS DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai karakteristik data hasil pengukuran linier pada foramen mental. Kemudian akan dilanjutkan dengan melakukan klasifikasi menggunakan metode *Feedforward* Neural Network (FFNN), klasifikasi dengan Boosting pada metode *Feedforward* Neural Network (FFNN), klasifikasi menggunakan metode Classification and Regression Tree (CART), dan klasifikasi dengan Boosting pada metode Classification and Regression Tree (CART)

4.1. Deskripsi Data Hasil Pengukuran Linier pada Foramen Mental

Deskripsi data akan dilakukan secara visual menggunakan *boxplot* dan didapatkan hasil sebagai berikut.

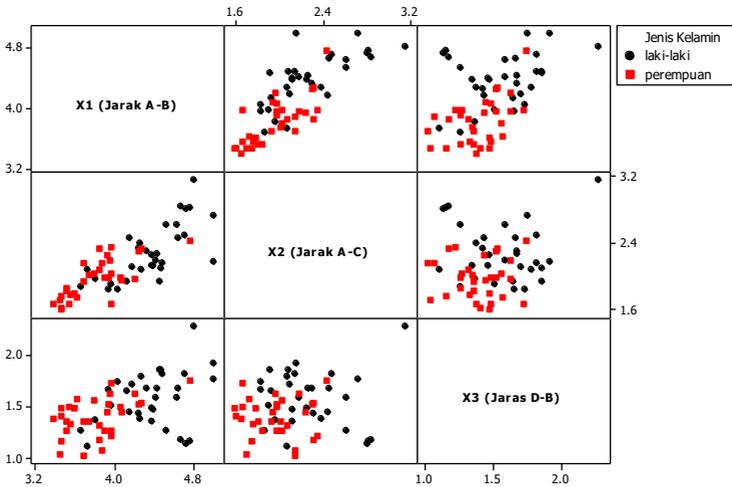


Gambar 4.1 Perbedaan sebaran data antara jenis kelamin laki-laki dan jenis kelamin perempuan

Melalui Gambar 4.1 dapat disimpulkan bahwa terdapat perbedaan sebaran data hasil pengukuran linier pada foramen mental antara manusia dengan jenis kelamin laki-laki dan jenis

kelamin perempuan, dimana jenis kelamin laki-laki cenderung memiliki jarak yang lebih lebar daripada jenis kelamin perempuan, hal tersebut terjadi pada semua variabel prediktor.

Kemudian akan dibuat *scatterplot* antar setiap variabel prediktor, yaitu jarak *superior border alveolar crest* terhadap *inferior border mandibula* (jarak A-B), jarak *superior border alveolar crest* dengan *superior border foramen mental* (jarak A-C), jarak *inferior border foramen mental* dengan *inferior border mandibula* (jarak D-B), untuk melihat apakah sebaran data antar kategori dapat terpisah secara linier (*linearly separable*) atau tidak. Berikut adalah *scatter plot* yang didapatkan.



Gambar 4.2 Sebaran data antar kategori jenis kelamin laki-laki dan perempuan pada setiap variabel prediktor

Dapat diketahui bahwa secara umum sebaran data antara jenis kelamin laki-laki dan perempuan tidak *linearly separable*, dimana terdapat beberapa data manusia dengan jenis kelamin laki-laki berada pada daerah manusia dengan jenis kelamin perempuan dan juga terjadi sebaliknya, dimana kasus ini disebut sebagai *linearly non-separable data*. Hal ini mengindikasikan bahwa pada kasus ini sebaiknya menggunakan metode klasifikasi non linier.

4.2. Klasifikasi Jenis Kelamin Berdasarkan Foramen Mental

Setelah melakukan deskripsi data hasil pengukuran linier pada foramen mental, selanjutnya akan dilakukan pembuatan model untuk mengklasifikasikan manusia dengan jenis kelamin laki-laki atau perempuan. Metode klasifikasi yang digunakan adalah metode *Feedforward* Neural Network (FFNN) dan Classification and Regression Tree (CART) yang kemudian akan dilanjutkan dengan metode *ensemble* yaitu boosting pada masing-masing metode klasifikasi. Berikut pembahasan lebih lanjut untuk masing-masing metode.

4.2.1. Klasifikasi dengan Metode *Feedforward* Neural Network (FFNN)

Metode pertama yang akan digunakan dalam klasifikasi adalah metode *Feedforward* Neural Network. Langkah pertama yang dilakukan dalam pembentukan struktur jaringan FFNN adalah menentukan algoritma yang digunakan. Dalam penelitian ini algoritma yang digunakan yaitu *backpropagation* dikarenakan algoritma ini cocok digunakan dalam kasus klasifikasi. Langkah kedua yaitu menentukan fungsi aktivasi, fungsi aktivasi merupakan fungsi matematis yang berguna untuk membatasi dan menentukan jangkauan *output* suatu neuron. Penelitian ini menggunakan fungsi aktivasi sigmoid dikarenakan nilai yang dihasilkan dari fungsi ini memiliki nilai pada interval 0 sampai 1. Langkah ketiga yaitu menentukan banyaknya input dan banyaknya neuron pada hidden layer. Pada kasus klasifikasi jumlah input yang digunakan berdasarkan pada banyaknya variabel prediktor yang digunakan, sehingga jumlah input adalah 3. Dalam mendapatkan model FFNN yang optimal, banyaknya neuron pada hidden layer akan dicoba mulai dari 1 neuron sampai dengan 20 neuron, dimana masing-masing jumlah neuron akan memberikan hasil yang berbeda.

Untuk memprediksi kesalahan dalam suatu klasifikasi maka digunakan metode *K-folds Cross Validation* (KCV) dengan nilai $K=5$. Berikut adalah rincian hasil *accuracy*, *sensitivity*, dan *spcitivity* yang didapat pada setiap *fold*.

Tabel 4.1 Performa metode FFNN pada setiap *fold*s

Jumlah Neuron	Fold	Accuracy		Sensitivity		Specificity	
		Training	Testing	Training	Testing	Training	Testing
1	1	85.42	85.71	87.50	71.43	83.33	100.00
1	2	88.00	66.67	80.00	66.67	96.00	66.67
1	3	88.00	75.00	88.00	66.67	88.00	83.33
1	4	82.00	66.67	84.00	66.67	80.00	66.67
1	5	86.00	83.33	84.00	83.33	88.00	83.33
2	1	89.58	85.71	91.67	71.43	87.50	100.00
2	2	86.00	66.67	80.00	66.67	92.00	66.67
2	3	86.00	75.00	80.00	66.67	92.00	83.33
2	4	88.00	75.00	88.00	100.00	88.00	50.00
2	5	84.00	83.33	72.00	83.33	96.00	83.33
3	1	93.75	71.43	95.83	57.14	91.67	85.71
3	2	92.00	83.33	84.00	83.33	100.00	83.33
3	3	86.00	75.00	80.00	66.67	92.00	83.33
3	4	82.00	75.00	72.00	66.67	92.00	83.33
3	5	94.00	75.00	100.00	83.33	88.00	66.67
4	1	97.92	78.57	100.00	57.14	95.83	100.00
4	2	100.00	83.33	100.00	83.33	100.00	83.33
4	3	90.00	66.67	92.00	66.67	88.00	66.67
4	4	92.00	83.33	92.00	100.00	92.00	66.67
4	5	96.00	66.67	96.00	83.33	96.00	50.00
5	1	93.75	78.57	91.67	71.43	95.83	85.71
5	2	100.00	83.33	100.00	83.33	100.00	83.33
5	3	100.00	75.00	100.00	66.67	100.00	83.33
5	4	100.00	75.00	100.00	66.67	100.00	83.33
5	5	100.00	83.33	100.00	100.00	100.00	66.67
6	1	100.00	71.43	100.00	57.14	100.00	85.71
6	2	100.00	91.67	100.00	100.00	100.00	83.33
6	3	100.00	66.67	100.00	66.67	100.00	66.67
6	4	100.00	91.67	100.00	83.33	100.00	100.00
6	5	100.00	75.00	100.00	83.33	100.00	66.67
7	1	100.00	57.14	100.00	28.57	100.00	85.71
7	2	100.00	75.00	100.00	66.67	100.00	83.33
7	3	100.00	66.67	100.00	50.00	100.00	83.33
7	4	100.00	66.67	100.00	50.00	100.00	83.33
7	5	100.00	83.33	100.00	100.00	100.00	66.67
8	1	100.00	64.29	100.00	57.14	100.00	71.43
8	2	100.00	75.00	100.00	66.67	100.00	83.33
8	3	100.00	91.67	100.00	83.33	100.00	100.00
8	4	100.00	75.00	100.00	66.67	100.00	83.33
8	5	100.00	83.33	100.00	83.33	100.00	83.33

Tabel 4.1 Performa metode FFNN pada setiap *fold*s (Lanjutan)

Jumlah Neuron	Fold	Accuracy		Sensitivity		Specificity	
		Training	Testing	Training	Testing	Training	Testing
9	1	100.00	78.57	100.00	71.43	100.00	85.71
9	2	100.00	75.00	100.00	66.67	100.00	83.33
9	3	100.00	58.33	100.00	66.67	100.00	50.00
9	4	100.00	75.00	100.00	83.33	100.00	66.67
9	5	100.00	83.33	100.00	83.33	100.00	83.33
10	1	100.00	78.57	100.00	57.14	100.00	100.00
10	2	100.00	75.00	100.00	66.67	100.00	83.33
10	3	100.00	75.00	100.00	66.67	100.00	83.33
10	4	100.00	66.67	100.00	50.00	100.00	83.33
10	5	100.00	83.33	100.00	100.00	100.00	66.67

Setelah didapatkan rincian hasil *accuracy*, *sensitivity*, dan *specitivity* pada setiap *fold*, selanjutnya akan dihitung rata-rata pada setiap *fold* agar mendapatkan hasil *accuracy*, *sensitivity*, dan *specitivity* pada setiap neuron pada *hidden layer*. Berikut adalah hasilnya.

Tabel 4.2 Performa metode FFNN pada setiap neuron pada *hidden layer*

Jumlah Neuron	Accuracy		Sensitivity		Specificity	
	Training	Testing	Training	Testing	Training	Testing
1	85.88	75.48	84.70	70.95	87.07	80.00
2	86.72	77.14	82.33	77.62	91.10	76.67
3	89.55	75.95	86.37	71.43	92.73	80.48
4	95.18	75.71	96.00	78.10	94.37	73.33
5	98.75	79.05	98.33	77.62	99.17	80.48
6	100.00	79.29	100.00	78.10	100.00	80.48
7	100.00	69.76	100.00	59.05	100.00	80.48
8	100.00	77.86	100.00	71.43	100.00	84.29
9	100.00	74.05	100.00	74.29	100.00	73.81
10	100.00	75.71	100.00	68.10	100.00	83.33

Dapat diketahui bahwa model FFNN terbaik yang didapat adalah model dengan jumlah neuron pada *hidden layer* sebanyak 6, dimana *accuracy*, *sensitivity*, dan *specitivity* yang didapat untuk data *testing* masing-masing adalah 79.29%, 78.10%, dan 84.29%, dimana nilai ini lebih besar daripada nilai yang didapat pada model

FFNN lainnya. Model terbaik FFNN yang didapatkan yaitu dengan menggunakan 3 input, 1 *hidden layer* dengan jumlah neuron 6, dan 1 output.

4.2.2. Klasifikasi dengan Boosting pada Metode *Feedforward Neural Network* (FFNN)

Metode FFNN menghasilkan model optimal dengan menggunakan 3 input dan 6 neuron pada *hidden layer*, dimana model optimal ini akan digunakan sebagai *base learner* untuk proses boosting. Berikut adalah *confussion matrix* Boosting FFNN untuk data *training* yang didapatkan.

Tabel 4.3 *Confussion Matrix Boosting FFNN untuk Data Training*

Kelas Prediksi	Kelas Aktual	
	Laki-Laki	Perempuan
Laki-Laki	25	0
Perempuan	0	25
Total	25	25

Berdasarkan tabel 4.3 dapat diketahui bahwa terdapat 25 pengamatan yang tepat diklasifikasikan ke kategori laki-laki atau *True Positive* (TP) dan 25 pengamatan yang tepat diklasifikasikan ke kategori perempuan atau *True Negative* (TN). Performa dari Boosting FFNN untuk data *training* dapat diukur dengan menggunakan *accuracy*, *sensitivity*, dan *specificity*. Perhitungan *accuracy*, *sensitivity*, dan *specificity* disajikan dalam perhitungan berikut ini.

$$Sensitivity = \frac{TP}{P} = \frac{25}{25} = 1$$

$$Specificity = \frac{TN}{N} = \frac{25}{25} = 1$$

$$Accuracy = \frac{TP + TN}{P + N} = \frac{25 + 25}{25 + 25} = 1$$

Berdasarkan hasil perhitungan yang telah dilakukan, diperoleh tingkat ketepatan klasifikasi sebesar 100%, dimana ketepatan klasifikasi kategori laki-laki sebesar 100% dan ketepatan

klasifikasi kategori perempuan sebesar 100%. Setelah menerapkan Boosting FFNN pada data *training*, selanjutnya Boosting FFNN diterapkan untuk data *testing*. Berikut adalah *confussion matrix* Boosting FFNN untuk data *testing* yang didapatkan.

Tabel 4.4 *Confoussion Matrix Boosting FFNN untuk Data Testing*

Kelas Prediksi	Kelas Aktual	
	Laki-Laki	Perempuan
Laki-Laki	5	1
Perempuan	1	5
Total	6	6

Berdasarkan tabel 4.3 dapat diketahui bahwa terdapat 5 pengamatan yang tepat diklasifikasikan ke kategori laki-laki atau *True Positive* (TP) dan 5 pengamatan yang tepat diklasifikasikan ke kategori perempuan atau *True Negative* (TN). Performa dari Boosting FFNN untuk data *testing* dapat diukur dengan menggunakan *accuracy*, *sensitivity*, dan *specificity*. Perhitungan *accuracy*, *sensitivity*, dan *specificity* disajikan dalam perhitungan berikut ini.

$$Sensitivity = \frac{TP}{P} = \frac{5}{6} = 0.8333$$

$$Specificity = \frac{TN}{N} = \frac{5}{6} = 0.8333$$

$$Accuracy = \frac{TP + TN}{P + N} = \frac{5 + 5}{6 + 6} = 0.8333$$

Berdasarkan hasil perhitungan yang telah dilakukan, diperoleh tingkat ketepatan klasifikasi sebesar 83.33%, dimana ketepatan klasifikasi kategori laki-laki sebesar 83.33% dan ketepatan klasifikasi kategori perempuan sebesar 83.33%.

4.2.3. Klasifikasi dengan Metode Classification and Regression Tree (CART)

Metode kedua yang digunakan adalah metode *Classification and Regression Tree* (CART) yang merupakan salah satu jenis dari *Decision Tree* dengan kelas biner. Analisis CART diawali dengan

pembentukan pohon klasifikasi maksimal yang menghasilkan jumlah simpul terminal terbanyak dengan kedalaman yang tinggi, setelah itu dilakukan pemangkasan pohon dengan menggunakan pendekatan *cost complexity minimum* sampai didapatkan pohon klasifikasi optimal dengan metode *5-fold cross validation estimate*. Berikut ini merupakan tahapan dalam analisis CART untuk klasifikasi jenis kelamin pada foramen mental.

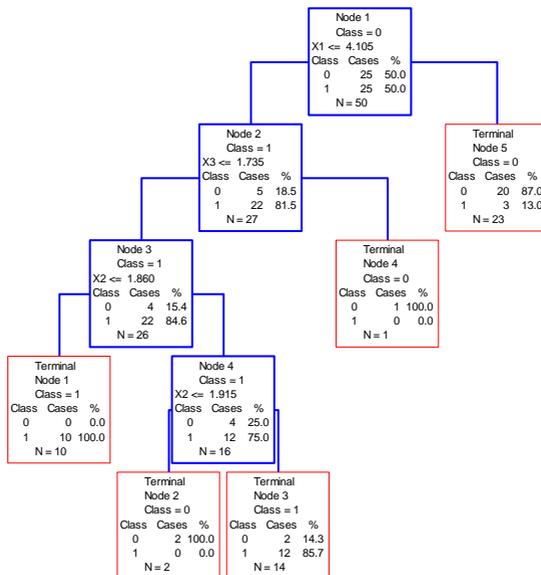
a. Pembentukan Pohon Klasifikasi Maksimal

Pembentukan pohon klasifikasi maksimal diawali dengan pemilihan *classifier* yakni variabel prediktor terpenting yang digunakan sebagai pemilah utama. Berdasarkan variabel prediktor yang digunakan, dilakukan pemilihan pemilah berdasarkan skor kontribusi terbesar yang dihasilkan oleh masing-masing variabel prediktor. Besarnya kontribusi masing-masing variabel prediktor ditampilkan dalam Tabel 4.3 berikut.

Tabel 4.5 Variabel terpenting dari pohon klasifikasi maksimal

Variabel	Nama Variabel	Skor
X ₁	Jarak A-B	100.00
X ₂	Jarak A-C	54.48
X ₃	Jarak D-B	52.49

Berdasarkan Tabel 4.3, dapat diketahui variabel terpenting yang digunakan sebagai pemilah utama atau dapat dikatakan sebagai simpul induk (*parent node*) pada pohon klasifikasi maksimal adalah X₁ (jarak A-B) dibuktikan dari skor kontribusi tertinggi yang dihasilkan yakni sebesar 100. Selanjutnya proses pemilahan terjadi sampai terbentuk pohon klasifikasi maksimal, dimana sifat dari pohon klasifikasi maksimal adalah menghasilkan simpul terminal terbanyak dengan kedalaman tertinggi. Jumlah simpul terminal yang dihasilkan dari pohon klasifikasi maksimal adalah sebanyak 5 simpul dengan tingkat kedalaman yang dihasilkan sebesar 5. *Topology* pohon maksimal dengan menggunakan pemilahan pemilah indeks gini ditunjukkan pada Gambar 4.10 berikut ini.

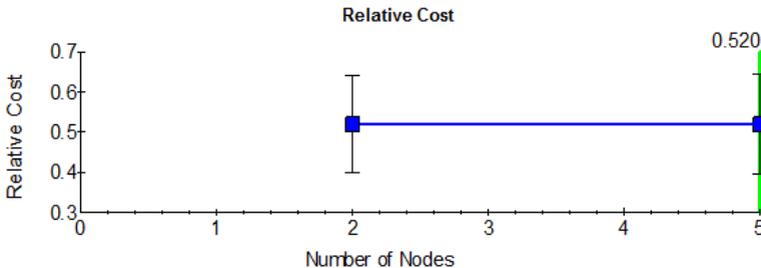


Gambar 4.3 Topology pohon klasifikasi maksimal

b. Pemangkasan (*Pruning*) Pohon Klasifikasi Maksimal

Setelah dihasilkan pohon klasifikasi maksimal, langkah selanjutnya adalah melihat apakah pohon klasifikasi maksimal tersebut perlu dilakukan pemangkasan (*pruning*) atau tidak. *Pruning* bertujuan untuk menghindari adanya kasus *overfitting* (nilai akurasi yang dihasilkan melebihi kenyataan yang ada) atau *underfitting* (nilai akurasi yang dihasilkan sangat kurang dibandingkan dengan nilai *actual*). Untuk melihat perlu tidaknya dilakukan pemangkasan pohon, dapat dilihat pada nilai *relative cost* yang dihasilkan. Pemangkasan pohon dilakukan dengan menggunakan pendekatan *cost complexity minimum* dan dalam menentukan pohon klasifikasi optimal digunakan metode *5-fold cross validation estimate*, yakni membagi data menjadi 5 bagian dimana 4 bagian akan digunakan untuk membentuk pohon klasifikasi dan 1 bagian lainnya digunakan untuk pendugaan kesalahan klasifikasi. Setiap hasil pemangkasan memiliki nilai

relative cost tertentu, kemudian dipilih hasil pemangkasan dengan nilai *relative cost* yang paling minimum. Berikut ini akan ditunjukkan *plot relative cost* pada klasifikasi jenis kelamin pada foramen mental.



Gambar 4.4 Plot *relative cost*

Berdasarkan Gambar 4.14, dapat diketahui bahwa jumlah pohon klasifikasi maksimal dengan jumlah simpul terminal sebanyak 5 yang diperoleh sudah dapat dikatakan optimal ditunjukkan dengan nilai *relative cost* yang dihasilkan sebesar 0.520, ditandai dengan garis berwarna hijau. Sedangkan, untuk biaya kesalahan (*cross-validation relative cost*) menghasilkan nilai yang paling minimum yakni sebesar $0,520 \pm 0,124$ atau bekisar antara 0,644 sampai 0,396 dengan nilai kompleksitas sebesar 0. Berikut ini akan ditunjukkan urutan dalam pembentukan pohon klasifikasi sampai didapatkan pohon klasifikasi optimal yang ditampilkan pada Tabel 4.5 berikut.

Tabel 4.6 Urutan Pembentukan Pohon Klasifikasi (*Tree Sequence*)

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Cross-Validated Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity</i>
1**	5	0.520 ± 0.124	0.200	0.000
2	2	0.520 ± 0.121	0.320	0.020
3	1	1.000 ± 0.000	1.000	0.340

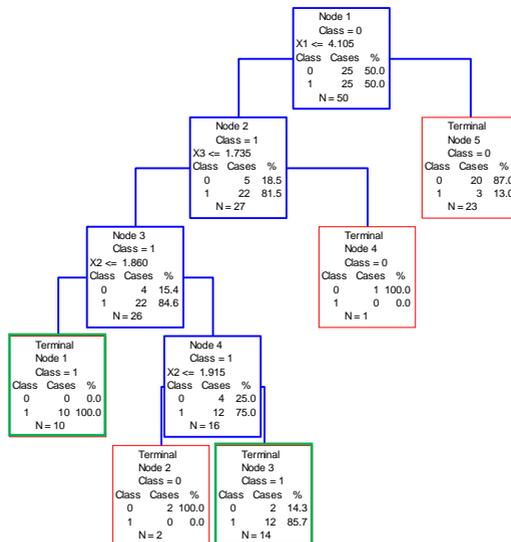
** pohon klasifikasi optimal

Berdasarkan Tabel 4.5, jika dilihat pada biaya kesalahan (*cross-validation relative cost*) yang dihasilkan, pohon klasifikasi 1 memiliki biaya kesalahan yang paling minimum, sehingga pohon

klasifikasi dengan simpul terminal sebanyak 5 simpul sudah layak atau dapat dikatakan optimal tanpa harus dilakukan pemangkasan terlebih dahulu.

c. Pemilihan Pohon Klasifikasi Optimal

Pohon klasifikasi maksimal yang dihasilkan juga dapat dikatakan sebagai pohon klasifikasi optimal karena tidak perlu dilakukan *prunning*. Langkah selanjutnya dalam analisis CART adalah menentukan pohon klasifikasi optimal yang digunakan untuk mengetahui akurasi yang dihasilkan dari klasifikasi jenis kelamin pada foramen mental. Variabel-variabel apa saja yang berperan penting dalam pemilahan dapat dilihat pada besarnya nilai skor atau kontribusi yang diberikan oleh masing-masing variabel. Berdasarkan Tabel 4.5, dapat diketahui variabel terpenting dengan skor tertinggi yang berperan sebagai simpul utama adalah variabel X_1 . Proses pemilahan yang dilakukan oleh variabel terpenting dalam pembentukan pohon optimal secara visual akan ditampilkan oleh Gambar 4.12 berikut ini.



Gambar 4.5 Topology pohon klasifikasi optimal

Setiap simpul terminal merupakan titik akhir dari suatu pemilahan berstruktur pohon, simpul ini tidak bisa dipilih kembali, simpul terminal merupakan simpul yang mengandung amatan-amatan homogen dan akhirnya dimasukkan sebagai sesuatu kelas tertentu. Pohon klasifikasi yang dihasilkan oleh pohon klasifikasi optimal sama dengan pohon klasifikasi maksimal yakni sebanyak 5 simpul terminal, sebanyak 2 simpul terminal diprediksi sebagai kategori perempuan dan 3 simpul terminal diprediksi sebagai kategori laki-laki, dengan tingkat kedalaman (*depth*) 5. Berikut adalah simpul terminal yang terbentuk pada penelitian ini.

a. Simpul Terminal 1

Simpul terminal 1 terdiri dari 10 pengamatan yang diprediksi sebagai kelompok kelas perempuan. Karakteristik jenazah berjenis kelamin perempuan dari simpul ini adalah memiliki jarak antara *superior border alveolar crest* dengan *superior border foramen mental* kurang dari 1.860, jarak antara *inferior border foramen mental* dengan *inferior border mandibula* kurang dari 1.735, dan jarak antara *superior border alveolar crest* terhadap *inferior border mandibula* kurang dari 4.105.

b. Simpul Terminal 2

Simpul terminal 2 terdiri dari 2 pengamatan yang diprediksi sebagai kelompok kelas laki-laki. Karakteristik jenazah berjenis kelamin laki-laki dari simpul ini adalah memiliki jarak antara *superior border alveolar crest* dengan *superior border foramen mental* kurang dari 1.915, jarak antara *superior border alveolar crest* dengan *superior border foramen mental* lebih dari 1.860, jarak antara *inferior border foramen mental* dengan *inferior border mandibula* kurang dari 1.735, dan jarak antara *superior border alveolar crest* terhadap *inferior border mandibula* kurang dari 4.105.

c. Simpul Terminal 3

Simpul terminal 3 terdiri dari 14 pengamatan yang diprediksi sebagai kelompok kelas perempuan. Karakteristik jenazah berjenis kelamin perempuan dari simpul ini adalah

memiliki jarak antara *superior border alveolar crest* dengan *superior border* foramen mental lebih dari 1.915, jarak antara *superior border alveolar crest* dengan *superior border* foramen mental lebih dari 1.860, jarak antara *inferior border* foramen mental dengan *inferior border* mandibula kurang dari 1.735, dan jarak antara *superior border alveolar crest* terhadap *inferior border* mandibula kurang dari 4.105.

d. Simpul Terminal 4

Simpul terminal 4 terdiri dari 1 pengamatan yang diprediksi sebagai kelompok kelas laki-laki. Karakteristik jenazah berjenis kelamin laki-laki dari simpul ini adalah memiliki jarak antara *superior border alveolar crest* dengan *superior border* foramen mental kurang dari 1.860, jarak antara *inferior border* foramen mental dengan *inferior border* mandibula kurang dari 1.735, dan jarak antara *superior border alveolar crest* terhadap *inferior border* mandibula kurang dari 4.105.

e. Simpul Terminal 5

Simpul terminal 5 terdiri dari 23 pengamatan yang diprediksi sebagai kelompok kelas laki-laki. Karakteristik jenazah berjenis kelamin laki-laki dari simpul ini adalah memiliki jarak antara *superior border alveolar crest* terhadap *inferior border* mandibula lebih dari 4.105.

Simpul terminal yang dihasilkan oleh pohon klasifikasi optimal akan diberi label kelas untuk mengetahui karakteristik dari masing-masing tipe (kelas) pada variabel respon. Berdasarkan Gambar 4.15 dapat dilihat bahwa simpul terminal dengan warna merah merupakan label kelas untuk jenis kelamin laki-laki (Kelas 0) dan simpul terminal dengan berwarna hijau merupakan label kelas untuk jenis kelamin perempuan (Kelas 1). Pemberian label pada simpul terminal berguna untuk memudahkan dalam mengetahui karakteristik atau sifat yang dihasilkan oleh klasifikasi jenis kelamin pada foramen mental. Pemberian label untuk masing-masing simpul terminal dilakukan dengan melihat persentase kelas terbesar yang terdapat pada simpul terminal tersebut.

Jumlah simpul terminal untuk kelas 0 (laki-laki) adalah sebanyak 3 simpul dan untuk kelas 1 (perempuan) adalah sebanyak 2 simpul. Kemudian setelah dilakukan pelabelan kelas pada masing-masing simpul terminal, langkah selanjutnya adalah menelusuri pohon klasifikasi optimal berdasarkan simpul terminal pada masing-masing kelas untuk mengetahui karakteristik atau sifat yang terdapat pada masing-masing kategori atau jenis kelamin. Pelabelan kelas untuk masing-masing simpul terminal yang dihasilkan oleh pohon klasifikasi optimal dapat dilihat pada Tabel 4.17.

Tabel 4.7 Pelabelan Kelas Simpul Terminal

Kelas	Simpul Terminal	Persentase (%)
Laki-Laki (3 simpul terminal)	2	100.0
	4	100.0
	5	87.0
Perempuan (2 simpul terminal)	1	100.0
	3	85.7

Karakteristik jenis kelamin dapat dilihat pada simpul terminal yang memiliki persentase kelas terbesar. Karakteristik kelas 0 (Laki-Laki) dapat diketahui dengan menelusuri simpul terminal 2 dan karakteristik kelas 1 (perempuan) dapat diketahui dengan menelusuri simpul terminal 1. Sehingga dengan menelusuri aturan pemilahan pada simpul terminal 1 dan 2 akan didapatkan karakteristik jenis kelamin pada Tabel 4.18 sebagai berikut.

Tabel 4.8 Karakteristik Jenis Kelamin Berdasarkan Simpul Terminal dengan Persentase Terbesar

Kelas	Simpul Terminal	Karakteristik
Laki-Laki	ST-2	- Manusia yang memiliki jarak <i>superior border alveolar crest</i> terhadap <i>inferior border mandibula</i> kurang dari 4.105 cm, jarak antara <i>inferior border foramen mental</i> dengan <i>inferior border mandibula</i> kurang dari 1.735, jarak antara <i>superior border alveolar crest</i> dengan <i>superior border foramen mental</i> lebih dari 1.860, dan jarak antara <i>superior border alveolar crest</i> dengan <i>superior border foramen mental</i> kurang dari 1.915.

Tabel 4.8 Karakteristik Jenis Kelamin Berdasarkan Simpul Terminal dengan Persentase Terbesar (Lanjutan)

Kelas	Simpul Terminal	Karakteristik
Perempuan	ST-1	- Manusia yang memiliki jarak <i>superior border alveolar crest</i> terhadap <i>inferior border</i> mandibula kurang dari 4.105 cm, jarak antara <i>inferior border</i> foramen mental dengan <i>inferior border</i> mandibula kurang dari 1.735, jarak antara <i>superior border alveolar crest</i> dengan <i>superior border</i> foramen mental kurang dari 1.860.

Untuk mengklasifikasikan jenazah yang tidak diketahui jenis kelaminnya dapat digunakan pohon klasifikasi optimal yakni dengan cara menelusuri pohon klasifikasi dari atas ke bawah. Berikut ini ilustrasi dalam mengklasifikasikan jenazah untuk mengetahui jenis kelamin berdasarkan pohon klasifikasi optimal pada Gambar 4.15. Misalnya, terdapat 10 data baru berdasarkan pengukuran jarak pada foramen mental yang ditunjukkan pada Tabel 4.8 berikut ini.

Tabel 4.9 Ilustrasi Data Baru

No	Jarak A-B	Jarak A-C	Jaras A-D
1	4.64	2.62	1.59
2	3.56	1.76	1.33
3	3.75	2.01	1.35
4	4.28	2.07	1.79
5	4.38	2.12	1.35
6	4.38	2.25	1.48
7	4.53	2.62	1.26
8	3.52	1.81	1.36
9	3.86	2.32	1.18
10	4.07	1.98	1.49

Dengan melakukan penelusuran pada pohon klasifikasi optimal yang dihasilkan, maka hasil pengklasifikasian data baru adalah sebagai berikut.

1. Untuk jenazah pertama, apakah jenazah tersebut memiliki jarak *superior border alveolar crest* terhadap *inferior border*

- mandibula (Jarak A-B) lebih dari 4.105 cm? Jika iya, maka jenazah tersebut diklasifikasikan berjenis kelamin laki-laki.
2. Untuk jenazah kedua, apakah jenazah tersebut memiliki jarak *superior border alveolar crest* terhadap *inferior border* mandibula (Jarak A-B) kurang dari 4.105 cm? Jika iya, maka lihat jarak antara *inferior border* foramen mental dengan *inferior border* mandibula (Jarak A-D), apakah kurang dari 1.735? Jika iya, maka lihat jarak antara *superior border alveolar crest* dengan *superior border* foramen mental (Jarak A-C), apakah kurang dari 1.860? Jika iya, maka jenazah tersebut diklasifikasikan berjenis kelamin perempuan.
 3. Untuk jenazah kedua, apakah jenazah tersebut memiliki jarak *superior border alveolar crest* terhadap *inferior border* mandibula (Jarak A-B) kurang dari 4.105 cm? Jika iya, maka lihat jarak antara *inferior border* foramen mental dengan *inferior border* mandibula (Jarak A-D), apakah kurang dari 1.735? Jika iya, maka lihat jarak antara *superior border alveolar crest* dengan *superior border* foramen mental (Jarak A-C), apakah lebih dari 1.860? Jika iya, maka lihat jarak antara *superior border alveolar crest* dengan *superior border* foramen mental (Jarak A-C), apakah lebih dari 1.915? Jika iya, maka jenazah tersebut diklasifikasikan berjenis kelamin perempuan.

Begitu seterusnya untuk jenazah keempat sampai dengan jenazah terakhir. Berikut ini merupakan hasil klasifikasi data baru yang menunjukkan jenis kelamin yang nantinya akan menjadi variabel respon (kelas) yang dihasilkan pada Tabel 4.9.

Tabel 4.10 Hasil Klasifikasi Data Baru

No	Jarak A-B	Jarak A-C	Jarak A-D	Jenis Kelamin
1	4.64	2.62	1.59	0
2	3.56	1.76	1.33	1
3	3.75	2.01	1.35	1
4	4.28	2.07	1.79	0
5	4.38	2.12	1.35	0

Tabel 4.10 Hasil Klasifikasi Data Baru (Lanjutan)

No	Jarak A-B	Jarak A-C	Jarak A-D	Jenis Kelamin
6	4.38	2.25	1.48	0
7	4.53	2.62	1.26	0
8	3.52	1.81	1.36	1
9	3.86	2.32	1.18	1
10	4.07	1.98	1.49	1

d. Ketepatan Klasifikasi Pohon CART

Setelah terbentuk pohon klasifikasi optimal, langkah selanjutnya adalah menghitung ketepatan klasifikasi atau akurasi dari hasil klasifikasi jenis kelamin pada foramen mental menggunakan *5-fold cross-validation* untuk data *training* dan *testing*. Berikut adalah rincian hasil *accuracy*, *sensitivity*, dan *specitivity* yang didapat pada setiap *fold*.

Tabel 15 4.11 *Confusion Matrix* CART

Fold	<i>Accuracy</i>		<i>Sensitivity</i>		<i>Specificity</i>	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
1	83.30	85.70	83.30	71.40	83.30	100.00
2	86.00	75.00	80.00	83.30	92.00	66.70
3	86.00	75.00	84.00	66.70	88.00	83.30
4	80.00	75.00	84.00	100.00	76.00	50.00
5	90.00	83.30	92.00	83.30	88.00	83.30
Mean	85.06	78.8	84.66	80.94	85.46	76.66

Dapat diketahui bahwa pohon optimal analisis CART untuk klasifikasi jenis kelamin pada foramen mental dengan jumlah simpul terminal sebanyak 5 simpul dan tingkat kedalaman 5 menghasilkan *accuracy*, *sensitivity*, dan *specitivity* untuk data *testing* masing-masing sebesar 78.80%, 80.94%, dan 76.66%.

4.2.4. Klasifikasi dengan Boosting pada Metode *Classification and Regression Tree* (CART)

Sebelumnya telah dilakukan klasifikasi dengan metode CART yang menghasilkan pohon terbaik pada *fold* ke-5. Proses selanjutnya adalah boosting CART dimana data pada *fold* ke-5

digunakan sebagai *base learner*. Berikut adalah *confussion matrix* Boosting CART untuk data *training* yang didapatkan.

Tabel 4.12 *Confoussion Matrix Boosting CART untuk Data Training*

Kelas Prediksi	Kelas Aktual	
	Laki-Laki	Perempuan
Laki-Laki	25	0
Perempuan	0	25
Total	25	25

Berdasarkan tabel 4.3 dapat diketahui bahwa terdapat 25 pengamatan yang tepat diklasifikasikan ke kategori laki-laki atau *True Positive* (TP) dan 25 pengamatan yang tepat diklasifikasikan ke kategori perempuan atau *True Negative* (TN). Performa dari Boosting CART untuk data *training* dapat diukur dengan menggunakan *accuracy*, *sensitivity*, dan *specificity*. Perhitungan *accuracy*, *sensitivity*, dan *specificity* sebagai berikut.

$$Sensitivity = \frac{TP}{P} = \frac{25}{25} = 1$$

$$Specificity = \frac{TN}{N} = \frac{25}{25} = 1$$

$$Accuracy = \frac{TP + TN}{P + N} = \frac{25 + 25}{25 + 25} = 1$$

Berdasarkan hasil perhitungan yang telah dilakukan, diperoleh tingkat ketepatan klasifikasi sebesar 100%, dimana ketepatan klasifikasi kategori laki-laki sebesar 100% dan ketepatan klasifikasi kategori perempuan sebesar 100%. Setelah menerapkan Boosting CART pada data *training*, selanjutnya Boosting CART diterapkan untuk data *testing*.

Tabel 4.13 *Confoussion Matrix Boosting CART untuk Data Testing*

Kelas Prediksi	Kelas Aktual	
	Laki-Laki	Perempuan
Laki-Laki	6	1
Perempuan	0	5
Total	6	6

Berdasarkan tabel 4.3 dapat diketahui bahwa terdapat 6 pengamatan yang tepat diklasifikasikan ke kategori laki-laki atau *True Positive* (TP) dan 5 pengamatan yang tepat diklasifikasikan ke kategori perempuan atau *True Negative* (TN). Performa dari Boosting FFNN untuk data *testing* dapat diukur dengan menggunakan *accuracy*, *sensitivity*, dan *specificity*. Perhitungan *accuracy*, *sensitivity*, dan *specificity* disajikan dalam perhitungan berikut ini.

$$Sensitivity = \frac{TP}{P} = \frac{6}{6} = 1$$

$$Specificity = \frac{TN}{N} = \frac{5}{6} = 0.8333$$

$$Accuracy = \frac{TP + TN}{P + N} = \frac{6 + 5}{6 + 6} = 0.9167$$

Berdasarkan hasil perhitungan yang telah dilakukan, diperoleh tingkat ketepatan klasifikasi sebesar 91.67%, dimana ketepatan klasifikasi kategori laki-laki sebesar 100% dan ketepatan klasifikasi kategori perempuan sebesar 83.33%.

4.3. Perbandingan Performa Antar Metode

Setelah didapatkan nilai-nilai *accuracy*, *sensitivity*, dan *specificity* dari metode *Feedforward* Neural Network, Boosting FFNN, *Classification and Regression Tree*, dan Boosting CART, maka selanjutnya akan dilakukan perbandingan metode berdasarkan nilai-nilai tersebut untuk memilih metode yang terbaik. Rangkuman hasil performa untuk setiap metode yang digunakan dapat dilihat pada Tabel 4.7 berikut.

Tabel 4.14 Rangkuman performa dari metode FFNN, boosting FFNN, CART, dan boosting CART

Metode	Accuracy		Sensitivity		Specificity	
	Training	Testing	Training	Testing	Training	Testing
FFNN	100.00	79.29	100.00	78.10	100.00	80.48
Boosting FFNN	100.00	83.33	100.00	83.33	100.00	83.33
CART	85.06	78.8	84.66	80.94	85.46	76.66
Boosting CART	100.00	91.67	100.00	100.00	100.00	83.33

Pada Tabel 4.14 dapat dilihat bahwa metode yang terbaik adalah metode Boosting CART karena memiliki *accuracy*, *sensitivity*, dan *specitivity* pada data *testing* yang lebih besar dibandingkan dengan metode yang lain.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan yang didapat berdasarkan hasil analisis yang telah dilakukan adalah terdapat perbedaan sebaran data hasil pengukuran linier pada foramen mental antara manusia dengan jenis kelamin laki-laki dan jenis kelamin perempuan, dimana jenis kelamin laki-laki cenderung memiliki jarak yang lebih lebar daripada jenis kelamin perempuan, hal tersebut terjadi pada semua variabel prediktor yang digunakan, yaitu jarak *superior border alveolar crest* terhadap *inferior border mandibula* (jarak A-B), jarak *superior border alveolar crest* dengan *superior border foramen mental* (jarak A-C), jarak *inferior border foramen mental* dengan *inferior border mandibula* (jarak D-B).

Model terbaik FFNN untuk klasifikasi jenis kelamin yang didapatkan yaitu dengan menggunakan 3 input, 1 *hidden layer* dengan jumlah neuron 6 unit, dan 1 output. Pohon optimal analisis CART yaitu dengan jumlah simpul terminal sebanyak 5 simpul dan tingkat kedalaman 5. Kesimpulan lain yang didapat adalah metode *Boosting Classification and Regression Tree* (CART) merupakan metode yang terbaik dalam mengklasifikasikan jenis kelamin.

5.2. Saran

Berdasarkan analisis yang telah dilakukan pada penelitian ini, maka saran yang dapat diberikan adalah diharapkan pada penelitian selanjutnya menggunakan metode klasifikasi yang berbeda atau pengembangan dari metode yang telah digunakan pada tugas akhir ini, sehingga dapat dibandingkan dengan metode yang telah digunakan pada penelitian ini dan dipilih metode yang lebih baik.

(Halaman ini sengaja dikosongkan)

DAFTAR PUSTAKA

- Blau, S. (2006). The Role of Forensik Anthropology in Disaster Victim Identification: A Brief Overview. *DVI Workshop*, (pp.25-27). Bandung.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. United States of America: Chapman & Hall/CRA.
- Breiman, L., Schwenk, H., & Bengio, Y. (2000). Boosting Neural Networks. *Neural Competition 12*, 1869-1807.
- Chandra, A., Singh, A., Badni, M., Jaiswal, R., & Agnihotri, A. (2013). Determination of sex by radiographic analysis of mental foramen in North Indian population. *Journal of Forensic Dental Sciences*, 52. doi: 10.4103/0975-1475.114556
- Fausett, L. (1994). *Fundamental of Neural Networks: Architectures Algorithm, and Applications*. New Jersey: Prentice Hall, Englewood Cliffs.
- Fernanda, J. W., & Otok, B. W. (2012). Boosting Neural Network dan Boosting Cart Pada Klasifikasi Diabetes Militus Tipe II. *Jurnal Matematika Vol. 2 No. 2, Desember 2012. ISSN : 1693-1394*, 33-49.
- Freund, Y., & Schapire, R. E. (1999). A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5):771-780.
- Ghouse, N., Nagaraj, T., James, L., Swamy, N. N., & C. D. Jagdish, T. B. (2016). Digital analysis of linear measurements related to the mental and mandibular foramina in sex determination. *Journal Med Radiol Pathol Surg*, 2:5-9. doi: 10.15713/ins.jmrps.55
- Gupta, S., & Soni, J. S. (2012). Study of anatomical variations and incidence of mental foramen and accessory mental foramen in dry human mandibles. *National Journal of MedRes*, 2: 28-30. doi: 10.4103/2249-5975.119803

- Haykin, S. (1999). *Neural Network: A Comprehensive Foundation (2nd ed.)*. New Jersey: Prentice Hall.
- Hu, K.-S., Koh, K.-S., Han, S.-H., Shin, K.-J., & Kim, H.-J. (2006). Sex Determination Using Nonmetric Characteristics of the Mandible in Koreans. *Journal Forensic Sciences*, Vol. 51, No. 6. doi: 10.1111/j.1556-4029.2006.00270.x
- Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning*. New York: Springer Science+Business Media, LLC.
- Khasanah, P. M. (2015). *Klasifikasi Penderita Retardasi Mental Di Rumah Sakit Jiwa Menur Provinsi Jawa Timur Menggunakan Synthetic Minority Oversampling Technique (SMOTE) - Classification And Regression Trees (CART)*. Institut Teknologi Sepuluh Nopember Surabaya, Statistika, Surabaya.
- Kholilullah, R. (2016). *Prevalensi Posisi dan Tipe Foramen Mental melalui Pengamatan Radiograf Panoramik Pasien RSGM FKG Universitas Airlangga*. Surabaya: Repository Universitas Airlangga.
- Langland, O. E., Langlais, R. P., & Preece, J. (2002). *Principles of dental imaging*. Philadelphia: Lippincott Williams & Wilkins.
- Malik, M., Laller, S., Saini, R. S., Mishra, R. K., Hora, I., & Dahiya, N. (2016). Mental foramen: An Indicator for Gender Determination - A Radiographic Study. *Santosh University Journal of Health Sciences*, 2(1):12-14. doi: 10.4103/0975-1475.114556
- Naroor, N., Shenai, P., Chatra, L., Veena, K. M., Rao, P. K., & Shetty, P. (2015). Gender determination using the mental foramen. *Journal Cranio Max Dis*, 4:144-7. doi: 10.4103/2278-9588.163260
- Okun, O. (2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations*. United States of America: IGI Global.

- Okun, O., Valentini, G., & Re, M. (2011). *Ensembles in Machine Learning Applications*. German: Springer-Verlag Berlin Heidelberg.
- Prawestiningtyas, E., & Algozi, A. M. (2009). Identifikasi Forensik Berdasarkan Pemeriksaan Primer dan Sekunder Sebagai Penentu Identitas Korban pada Dua Kasus Bencana Massal. *Jurnal Kedokteran Brawijaya*, Vol XXV, No. 2, 87-94.
- Rashid, S., & Ali, J. (2011). Sex Determination Using Linear Measurements Related to The Mental and Mandibular Foramina Vertical Positions on Digital Panoramic Images. *J Bagh Coll Dentistry*, 23, 59-64.
- Rokach, L., & Maimon, O. (2015). *Data Mining With Decision Trees: Theory and Applications*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Saparwoko, E. (2006). DVI in Indonesia: An Overview. *DVI Workshop*, (pp. 25-27). Bandung.
- Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, 5(2):197-227. doi: 10.1109/sfcs.1989.63451
- Setiawan, A. D., Savitri, Y., & Mulyani, S. W. (2016). *Determinasi Jenis Kelamin dengan Metode Pengukuran Linear pada Foramen Mental menggunakan Radiografi Panoramik*. Surabaya: Universitas Airlangga.
- Siang, J. (2009). *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan MATLAB*. Yogyakarta: ANDI.
- Singal, K., & Sharma, S. (2016). Gender Determination by Mental Foramen Using Linear Measurements on Radiographs: A Study in Haryana Population. *Indian Journal of Forensic Medicine & Toxicology*, 44-49. doi: 10.5958/0973-9130.2016.00011.6
- Singh, Y., & Chauhan, A. S. (2009). Neural Network in Data Mining. *Journal of Theoretical and Applied Information Technology*. doi: 10.1002/widm.1052
- Thakur, M., Reddy, K. V., Y.Sivaranjani, & Khaja, S. (2014). Gender Determination by Mental Foramen and Height of

the Body of the Mandible in Dentulous Patients A Radiographic Study. *Journal of Indian Academy of Forensic Medicine*, 13-18. doi: 10.5958/0973-9130.2016.00011.6

White, S., & Pharoah, M. (2009). *Oral Radiology Principles and Interpretation. 6th ed.* United States: Mosby, Inc.

Zhang, Y. (2010). *New Advances in Machine Learning.* Croatia: InTech. doi: 10.5772/225

LAMPIRAN

Lampiran 1. Data Hasil Pengukuran Jarak pada Foramen Mental

X1	X2	X3	Y
4.71	2.48	1.82	0
4.64	2.62	1.59	0
4.68	2.84	1.18	0
4.25	2.32	1.43	0
4.46	1.92	1.86	0
3.97	1.9	1.51	0
3.73	2.07	1.11	0
4.4	2.12	1.47	0
3.95	1.83	1.66	0
4.73	2.81	1.14	0
4.28	2.07	1.79	0
4.27	2.39	1.38	0
4.48	2.09	1.86	0
4.16	2.45	1.44	0
4.38	2.12	1.35	0
4.19	2.1	1.71	0
4.38	2.25	1.48	0
4.25	2.3	1.52	1
3.61	1.79	1.48	1
⋮	⋮	⋮	⋮
3.52	1.81	1.36	1
4.21	1.96	1.63	1
3.86	2.32	1.18	1
3.56	1.76	1.33	1
3.85	2.07	1.32	1
4.28	2.32	1.53	1
3.91	1.98	1.26	1
3.96	2.18	1.63	1
3.7	1.93	1.36	1
3.94	2.24	1.44	1
3.8	2.02	1.56	1
3.39	1.65	1.38	1
4.07	1.98	1.49	1

Lampiran 2. Syntax Feedforward Neural Network (FFNN)

```

library(neuralnet)
library(caret)
xdata <- read.csv("D:/skripsi01.csv")
maxs <- apply(xdata, 2, max)
mins <- apply(xdata, 2, min)
data <- as.data.frame(scale(xdata, center = mins, scale = maxs - mins))
ind
c(4,5,2,1,1,3,3,4,5,3,5,4,4,4,5,3,5,2,2,1,3,1,1,2,2,3,1,4,1,2,5,4,4,3,1,
1,2,4,1,3,4,2,2,5,5,5,5,1,3,4,2,2,3,4,5,1,1,3,2,1,3,5)
datax <- cbind(data,ind)
colnames(datax) <- c("X1", "X2", "X3", "Y", "id")
nn=ncol(data)
n=nrow(data)
jmlnode=10
jmlk=5
subset=seq(1,n,1)
result<-array(vector(mode = 'numeric'), c(jmlk*jmlnode, 8))
dimnames(result)[[2]]<-c("Fold", "Hidden", "aktrain", "aktest", "sentrain",
"sentest", "spetrain", "spetest")

z <- 0
for(i in 1:jmlk) {
  datatraining=datax[subset[datax[, (nn+1)]!=i],]
  datatesting=datax[subset[datax[, (nn+1)]==i],]
  datatrain <- datatraining[, (1:nn)]
  datatest <- datatesting[, (1:nn)]
  for(j in (1:jmlnode)) {
    z<- z+1;
    set.seed(123)
    nn1 <- neuralnet(Y~X1+X2+X3, data=datatrain, hidden=j, err.fct="ce",
                    linear.output=FALSE)

    set.seed(123)
    probtrain <- compute(nn1, datatrain[, 1:3])
    probtrain_ <- probtrain$net.result*(maxs[4]-mins[4])+mins[4]
    predtrain <- round(probtrain_)
    CMtrain <- table(predtrain, datatrain[, 4])
    ptr <- as.factor(predtrain)
    atr <- as.factor(datatrain[, 4])
    sentrain <- sensitivity(ptr, atr)
    spetrain <- specificity(ptr, atr)
    aktrain <- sum(diag(CMtrain)/nrow(datatrain))
    set.seed(123)
    nn2 <- compute(nn1, datatest[, 1:3])
    probtest <- nn2$net.result*(maxs[4]-mins[4])+mins[4]
    predtest <- round(probtest)
    CMtest <- table(predtest, datatest[, 4])
    pts <- as.factor(predtest)
    ats <- as.factor(datatest[, 4])
    sentest <- sensitivity(pts, ats)
    spetest <- specificity(pts, ats)
    aktest <- sum(diag(CMtest))/nrow(datatest)
    result[z,1] <- i
    result[z,2] <- j
    result[z,3] <- aktrain
  }
}

```

```

result[z,4] <- aktest
result[z,5] <- sentrain
result[z,6] <- sentest
result[z,7] <- spetrain
result[z,8] <- spetest
}
}
xresult <- data.frame(result)
#View(xresult)
akhirresult<-array(vector(mode = 'numeric'), c(jmlnode, 7))
dimnames(akhirresult)[[2]]<-c("Hidden","aktrain","aktest","sentrain",
                               "sentest","spetrain","spetest")
zz<-0
for (l in 1:jmlnode) {
  zz<- zz+1
  akhirresult[zz,1] <- unique(xresult[xresult$Hidden==l,]$Hidden)
  akhirresult[zz,2] <- mean(xresult[xresult$Hidden==l,]$aktrain)
  akhirresult[zz,3] <- mean(xresult[xresult$Hidden==l,]$aktest)
  akhirresult[zz,4] <- mean(xresult[xresult$Hidden==l,]$sentrain)
  akhirresult[zz,5] <- mean(xresult[xresult$Hidden==l,]$sentest)
  akhirresult[zz,6] <- mean(xresult[xresult$Hidden==l,]$spetrain)
  akhirresult[zz,7] <- mean(xresult[xresult$Hidden==l,]$spetest)
}
}
FFNN <- data.frame(akhirresult)
#View(FFNN)

```

Lampiran 3 Output Syntax Feedforward Neural Network (FFNN)

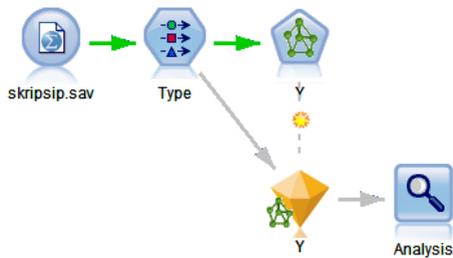
```

> xresult
  Fold Hidden   aktrain   aktest sentrain   sentest   spetrain   spetest
1     1      1  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
2     1      2  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
3     1      3  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
4     1      4  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
5     1      5  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
6     1      6  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
7     1      7  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
8     1      8  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
9     1      9  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
10    1     10  0.8541666667 0.8571428571   0.875 0.7142857143 0.8333333333 1.0000000000
11    2      1  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
12    2      2  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
13    2      3  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
14    2      4  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
15    2      5  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
16    2      6  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
17    2      7  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
18    2      8  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
19    2      9  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
20    2     10  0.9000000000 0.7500000000   0.800 0.6666666667 1.0000000000 0.8333333333
21    3      1  0.9800000000 0.5833333333   1.000 0.6666666667 0.9600000000 0.5000000000
22    3      2  0.9800000000 0.5833333333   1.000 0.6666666667 0.9600000000 0.5000000000
23    3      3  0.9800000000 0.5833333333   1.000 0.6666666667 0.9600000000 0.5000000000
24    3      4  0.9800000000 0.5833333333   1.000 0.6666666667 0.9600000000 0.5000000000
25    3      5  0.9800000000 0.5833333333   1.000 0.6666666667 0.9600000000 0.5000000000
26    3      6  0.9800000000 0.5833333333   1.000 0.6666666667 0.9600000000 0.5000000000
27    3      7  0.9800000000 0.5833333333   1.000 0.6666666667 0.9600000000 0.5000000000

```

28	3	8	0.9800000000	0.5833333333	1.000	0.6666666667	0.9600000000	0.5000000000
29	3	9	0.9800000000	0.5833333333	1.000	0.6666666667	0.9600000000	0.5000000000
30	3	10	0.9800000000	0.5833333333	1.000	0.6666666667	0.9600000000	0.5000000000
31	4	1	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
32	4	2	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
33	4	3	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
34	4	4	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
35	4	5	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
36	4	6	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
37	4	7	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
38	4	8	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
39	4	9	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
40	4	10	1.0000000000	0.7500000000	1.000	0.6666666667	1.0000000000	0.8333333333
41	5	1	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
42	5	2	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
43	5	3	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
44	5	4	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
45	5	5	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
46	5	6	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
47	5	7	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
48	5	8	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
49	5	9	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
50	5	10	1.0000000000	0.8333333333	1.000	0.8333333333	1.0000000000	0.8333333333
> FFNN								
Hidden aktrain aktest sentrain sentest spetrain spetest								
1	1	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
2	2	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
3	3	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
4	4	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
5	5	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
6	6	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
7	7	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
8	8	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
9	9	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	
10	10	0.9468333333	0.7547619048	0.935	0.7095238095	0.9586666667	0.8	

Lampiran 4 Boosting FFNN



Lampiran 5 Output Boosting FFNN

Results for output field Y

Comparing \$N-Y with Y

'P'	1.0	2.0
Correct	50 100%	10 83.33%
Wrong	0 0%	2 16.67%
Total	50	12

Coincidence Matrix for \$N-Y (rows show actuals)

'P' = 1.0	0.000000	1.000000
0.000000	25	0
1.000000	0	25
'P' = 2.0	0.000000	1.000000
0.000000	5	1
1.000000	1	5

Performance Evaluation

'P' = 1.0	
0.000000	0.693
1.000000	0.693
'P' = 2.0	
0.000000	0.511
1.000000	0.511

Lampiran 6 Syntax untuk memilih rpart.control yang terbaik

```
# membagi training testing 80:20 #
library(caret)
skripsi <- read.csv("D:/skripsi01.csv")
set.seed(123)
partisi <- createDataPartition(skripsi$Y, p = 0.8, list=F)
train <- skripsi[partisi, 1:4]
test <- skripsi[-partisi, 1:4]
=====
# memilih rpart.control yg terbaik #
library(e1071)
rpart.ranges <- list(minsplit = seq(5, 50, by = 5), cp = c(0,
  0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2,0.5), maxdepth =
  1:10)
set.seed(123)
regtree.tune <- tune(rpart,Y ~ ., data = train, ranges = rpart.ranges)
```

Lampiran 7 Output untuk memilih rpart.control yang terbaik

```
> regtree.tune

Parameter tuning of 'rpart':

- sampling method: 10-fold cross validation

- best parameters:
  minsplit cp maxdepth
    10    0         5

- best performance: 0.156978773
```

Lampiran 8 *Syntax Boosting Classification and Regression Tree (CART)*

```

library(adabag)
data <- read.csv("E:/skripsilp.csv")
ind <- c(4,5,2,1,1,3,3,4,5,3,5,4,4,4,5,3,5,2,2,1,3,1,1,2,2,3,1,4,
1,2,5,4,4,3,1,1,2,4,1,3,4,2,2,5,5,5,5,1,3,4,2,2,3,4,5,1,1,3,2,1,3,5)
datax <- cbind(data,ind)
colnames(datax) <- c("X1", "X2", "X3", "Y", "id")
jmlk <- 5
result <- array(vector(mode = 'numeric'), c(jmlk, 7))
dimnames(result)[[2]] <- c("Fold", "aktrain", "aktest", "sentrain",
"sentest", "spetrain", "spetest")

z <- 0
for(i in 1:jmlk) {
  z<- z+1;
  datatrain <- datax[subset[datax[, (nn+1)]!=i],]
  datatest <- datax[subset[datax[, (nn+1)]==i],]
  datatraining <- datatrain[,1:4]
  datatesting <- datatest[,1:4]
  ntrain <- nrow(datatraining)
  ntest <- nrow(datatesting)
  # BOOSTING CART
  set.seed(1234)
  bcart <- boosting(Y ~ ., data = datatraining, boos = TRUE, mfinal = 50,
  coeflearn = "Breiman")

  # AKURASI
  set.seed(1234)
  predtrain <- predict.boosting(bcart, newdata = datatraining)
  tabeltrain <- table(predtrain$class, datatraining[,4])
  ptr <- as.factor(predtrain$class)
  atr <- as.factor(datatraining[,4])
  sentrain <- sensitivity(ptr, atr)
  spetrain <- specificity(ptr, atr)
  aktrain <- sum(diag(tabeltrain))/ntrain
  set.seed(1234)
  predtest <- predict.boosting(bcart, newdata = datatesting)
  tabeltest <- table(predtest$class, datatesting[,4])
  pts <- as.factor(predtest$class)
  ats <- as.factor(datatesting[,4])
  sentest <- sensitivity(pts, ats)
  spetest <- specificity(pts, ats)
  aktest <- sum(diag(tabeltest))/ntest
  result[z,1] <- i
  result[z,2] <- aktrain
  result[z,3] <- aktest
  result[z,4] <- sentrain
  result[z,5] <- sentest
  result[z,6] <- spetrain
  result[z,7] <- spetest
}
boostcart400 <- data.frame(result)
View(boostcart400)

```

Lampiran 9 Output CART Fold 1 (data *training*)

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

Actual Class	Predicted Class		Actual Total
	0	1	
0	20.00	4.00	24.00
1	4.00	20.00	24.00
PRED. TOT.	24.00	24.00	48.00
CORRECT	0.833	0.833	
SUCCESS IND.	0.333	0.333	
TOT. CORRECT	0.833		

SENSITIVITY:	0.833	SPECIFICITY:	0.833
FALSE REFERENCE:	0.167	FALSE RESPONSE:	0.167
REFERENCE = "0", RESPONSE = "1"			

Lampiran 10 Output CART Fold 1 (data *testing*)

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

Actual Class	Predicted Class		Actual Total
	0	1	
0	5.00	2.00	7.00
1	0.00	7.00	7.00
PRED. TOT.	5.00	9.00	14.00
CORRECT	0.714	1.000	
SUCCESS IND.	0.214	0.500	
TOT. CORRECT	0.857		

SENSITIVITY:	0.714	SPECIFICITY:	1.000
FALSE REFERENCE:	0.000	FALSE RESPONSE:	0.222
REFERENCE = "0", RESPONSE = "1"			

Lampiran 11 Output CART Fold 2 (data *training*)

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

Actual Class	Predicted Class		Actual Total
	0	1	
0	20.00	5.00	25.00
1	2.00	23.00	25.00
PRED. TOT.	22.00	28.00	50.00
CORRECT	0.800	0.920	
SUCCESS IND.	0.300	0.420	
TOT. CORRECT	0.860		

SENSITIVITY:	0.800	SPECIFICITY:	0.920
FALSE REFERENCE:	0.091	FALSE RESPONSE:	0.179
REFERENCE = "0", RESPONSE = "1"			

Lampiran 12 Output CART Fold 2 (data *testing*)

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

Actual Class	Predicted Class		Actual Total
	0	1	
0	5.00	1.00	6.00
1	2.00	4.00	6.00
PRED. TOT.	7.00	5.00	12.00
CORRECT	0.833	0.667	
SUCCESS IND.	0.333	0.167	
TOT. CORRECT	0.750		
SENSITIVITY: 0.833 SPECIFICITY: 0.667			
FALSE REFERENCE: 0.286 FALSE RESPONSE: 0.200			
REFERENCE = "0", RESPONSE = "1"			

Lampiran 13 Output CART Fold 3 (data *training*)

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

Actual Class	Predicted Class		Actual Total
	0	1	
0	21.00	4.00	25.00
1	3.00	22.00	25.00
PRED. TOT.	24.00	26.00	50.00
CORRECT	0.840	0.880	
SUCCESS IND.	0.340	0.380	
TOT. CORRECT	0.860		
SENSITIVITY: 0.840 SPECIFICITY: 0.880			
FALSE REFERENCE: 0.125 FALSE RESPONSE: 0.154			
REFERENCE = "0", RESPONSE = "1"			

Lampiran 14 Output CART Fold 3 (data *testing*)

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

Actual Class	Predicted Class		Actual Total
	0	1	
0	4.00	2.00	6.00
1	1.00	5.00	6.00
PRED. TOT.	5.00	7.00	12.00
CORRECT	0.667	0.833	
SUCCESS IND.	0.167	0.333	
TOT. CORRECT	0.750		
SENSITIVITY: 0.667 SPECIFICITY: 0.833			
FALSE REFERENCE: 0.200 FALSE RESPONSE: 0.286			
REFERENCE = "0", RESPONSE = "1"			

Lampiran 15 iOutput CART Fold 4 (data *training*)

LEARNING SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class		Actual Total
	0	1	
0	21.00	4.00	25.00
1	6.00	19.00	25.00
PRED. TOT.	27.00	23.00	50.00
CORRECT	0.840	0.760	
SUCCESS IND.	0.340	0.260	
TOT. CORRECT	0.800		
SENSITIVITY: 0.840 SPECIFICITY: 0.760			
FALSE REFERENCE: 0.222 FALSE RESPONSE: 0.174			
REFERENCE = "0", RESPONSE = "1"			

Lampiran 16 Output CART Fold 4 (data *testing*)

CASE-BY-CASE SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class		Actual Total
	0	1	
0	6.00	0.00	6.00
1	3.00	3.00	6.00
PRED. TOT.	9.00	3.00	12.00
CORRECT	1.000	0.500	
SUCCESS IND.	0.500	0.000	
TOT. CORRECT	0.750		
SENSITIVITY: 1.000 SPECIFICITY: 0.500			
FALSE REFERENCE: 0.333 FALSE RESPONSE: 0.000			
REFERENCE = "0", RESPONSE = "1"			

Lampiran 17 Output CART Fold 5 (data *training*)

LEARNING SAMPLE CLASSIFICATION TABLE

Actual Class	Predicted Class		Actual Total
	0	1	
0	23.00	2.00	25.00
1	3.00	22.00	25.00
PRED. TOT.	26.00	24.00	50.00
CORRECT	0.920	0.880	
SUCCESS IND.	0.420	0.380	
TOT. CORRECT	0.900		
SENSITIVITY: 0.920 SPECIFICITY: 0.880			
FALSE REFERENCE: 0.115 FALSE RESPONSE: 0.083			
REFERENCE = "0", RESPONSE = "1"			

Lampiran 18 Surat Pernyataan Legalitas Data**SURAT PERNYATAAN**

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMIPA ITS:

Nama : Nur Lailatul Faizah

NRP : 1313 100 103

Menyatakan bahwa data yang digunakan dalam Tugas Akhir ini merupakan data yang diambil dari:

Sumber : Skripsi FKG Unair oleh Afrista dengan Judul “Determinasi Jenis Kelamin dengan Metode Pengukuran Linier pada Foramen Mental menggunakan Radiografi Panoramik”

Surat ini dibuat dengan sebenarnya, Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Mengetahui,
Pembimbing Tugas Akhir



(Dr. Kartika Fithriasari, M.Si)
NIP. 19691212 199303 2 001

Surabaya, 9 Juni 2017



(Nur Lailatul Faizah)
NRP. 1313 100 103

BIODATA PENULIS



Penulis bernama Nur Lailatul Faizah atau biasa dipanggil Iza lahir di Gresik, 10 Februari 1995. Penulis merupakan anak kedua dari tiga bersaudara yang lahir dari pasangan Suami Istri bernama Pundhi dan Fatkhullah. Penulis telah menyelesaikan pendidikan di TK Aisyiyah Bustanul Athfal 5 Surabaya, SD Negeri Tanah Kali Kedinding II Surabaya, SMP Negeri 15 Surabaya, dan SMA Negeri 9 Surabaya.

Kemudian penulis melanjutkan pendidikan kuliah S1 di Jurusan Statistika ITS melalui jalur Tes Tulis SBMPTN dengan NRP 1313100103. Penulis pernah aktif dalam organisasi kampus sebagai *staff* Departemen Pengembangan Sumber Daya Musik (PSDMus) Paduan Suara Mahasiswa (PSM ITS) periode 2014/2015 dan *staff* Departemen Hubungan Luar Paduan Suara Mahasiswa (PSM ITS) periode 2015/2016. Selain itu Penulis juga pernah mengikuti kompetisi lomba Paduan Suara pada 6th Brawijaya *Choir* Festival pada tahun 2014. Penulis pernah memiliki pengalaman magang selama satu bulan di Kantor Pelayanan Pajak Pratama Surabaya Mulyorejo. Penulis mengucapkan terimakasih kepada Bapak Fatkhullah dan Ibu Pundhi sebagai orang tua. Kakak Ana Fauziatul Khasanah dan Muhammad Jihad Aulia'ul Firdaus sebagai saudara penulis, yang selalu memberikan doa, semangat, serta motivasi. Dan juga kepada Desak, Adheala, Maulana, Aldo, Delinda, Barocca, Kedai27, dan Statistika Angkatan 2013 (*Legendary*) yang telah memberikan pengalaman dan kebersamaan yang sangat berharga. Apabila terdapat kritik, saran, dan diskusi mengenai Tugas Akhir ini dapat disampaikan melalui *email* nurlailatulfaizah@gmail.com.

