

**OPTIMASI FITUR SUARA PADA KLASIFIKASI SUARA
BATUK BASAH/KERING ANAK-ANAK DENGAN
ALGORITME GENETIKA**

TUGAS AKHIR

Untuk memenuhi salah satu persyaratan

Mencapai derajat sarjana S1



Disusun oleh:

Ibnu Hafid Prihastomo

13524109

Jurusan Teknik Elektro

Fakultas Teknologi Industri

Universitas Islam Indonesia

Yogyakarta

2018

LEMBAR PENGESAHAN

**OPTIMASI FITUR SUARA PADA KLASIFIKASI SUARA BATUK
BASA/KERING ANAK-ANAK DENGAN ALGORITME GENETIKA**



**Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Teknik
pada Program Studi Teknik Elektro
Fakultas Teknologi Industri
Universitas Islam Indonesia**

Disusun oleh:

Ibnu Hafid Prihastomo

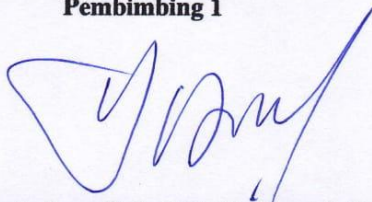
13524109

الجامعة الإسلامية
الاستاذة
الاستاذة

Yogyakarta, 8 Februari 2018

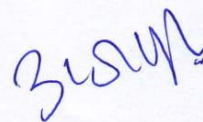
Menyetujui,

Pembimbing 1



Yusuf Aziz Amrulloh, S.T., M.Eng., Ph.D.
045240101

Pembimbing 2



Elvira Sukma Wahyuni, S.Pd., M.Eng.
155231301

LEMBAR PERNYATAAN

Dengan ini Saya menyatakan bahwa:

1. Tugas Akhir ini tidak mengandung karya yang diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan Saya juga tidak mengandung karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.
2. Informasi dan materi Tugas Akhir yang terkait hak milik, hak intelektual, dan paten merupakan milik bersama antara tiga pihak yaitu penulis, dosen pembimbing, dan Universitas Islam Indonesia. Dalam hal penggunaan informasi dan materi Skripsi terkait paten maka akan diskusikan lebih lanjut untuk mendapatkan persetujuan dari ketiga pihak tersebut di atas.

Yogyakarta, 8 Februari 2018



Ibnu Hafid Prihastomo

LEMBAR PENGESAHAN DOSEN PENGUJI

**OPTIMASI FITUR SUARA PADA KLASIFIKASI SUARA BATUK
BASAH/KERING ANAK-ANAK DENGAN ALGORITME GENETIKA**

TUGAS AKHIR

Disusun oleh :

Nama : Ibnu Hafid Prohastomo

No. Mahasiswa : 13524109

**Telah Dipertimbangkan di Depan Sidang Penguji Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Elektro Fakultas Teknologi Industri
Universitas Islam Indonesia**

Yogyakarta, 23 Februari 2018

Tim Penguji,

Yusuf Aziz Amrulloh, S.T., M.Eng., Ph.D.

Ketua

Ida Nurcahyani, S.T., M.Eng.

Anggota I

Dzata Farahiyah, S.T., M.Sc.

Anggota II

Mengetahui,

Ketua Program Studi Teknik Elektro

Universitas Islam Indonesia



Dr. Eng. Hendra Setiawan, S.T., M.T.

KATA PENGANTAR



Assalamualaikum Warahmatullahi Wabarakatuh

Segala puji bagi Allah Subhanahu Wa Ta'ala, Tuhan semesta alam yang telah melimpahkan rahmat dan hidayah kepada hamba-Nya, sehingga Tugas Akhir ini dapat diselesaikan. Sholawat serta salam semoga selalu tercurah kepada Rasulullah Muhammad Shallallahu Allaihi Wasallam beserta para keluarga, sahabat dan para pengikutnya hingga akhir zaman. Tugas Akhir yang berjudul “Optimasi Fitur Suara Pada Klasifikasi Suara Batuk Basah/Kering Anak-Anak Dengan Algoritme Genetika” ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Program Studi Teknik Elektro, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Penulis menyadari sepenuhnya bahwa isi dari Tugas Akhir ini masih jauh dari sempurna karena keterbatasan ilmu dan pengetahuan penulis sebagai mahasiswa, oleh karena itu dengan segala kerendahan hati, penulis menerima saran dan kritikan yang sifatnya membangun dari semua pihak.

Pada kesempatan ini, ungkapan rasa terima kasih yang sebesar-besarnya diucapkan kepada berbagai pihak yang telah memberikan bantuan, bimbingan, dukungan, kerja sama, fasilitas dan kemudahan lainnya. Untuk itu, dengan ketulusan hati saya mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Kedua Orang Tua yang telah memberikan semangat baik secara materil maupun moril, hingga Tugas Akhir ini dapat diselesaikan.
2. Bapak Hendra Setiawan, S.T., M.T., Ph.D. selaku Ketua Program Studi Teknik Elektro Universitas Islam Indonesia.
3. Bapak Yusuf Aziz Amrulloh S.T., M.Eng., Ph.D. selaku Dosen Pembimbing 1 Tugas Akhir yang telah meluangkan waktunya sampai terselesaikannya Tugas Akhir ini.
4. Ibu Elvira S. Wahyuni, S.Kom., M.Eng. selaku Dosen Pembimbing 2 Tugas Akhir yang telah meluangkan waktunya sampai terselesaikannya Tugas Akhir ini.
5. Adik-Adik tercinta Hanifa dan Luthfi serta seluruh keluarga besar yang selalu mendoakan dan memotivasi penulis dalam menulis Tugas Akhir ini.
6. Bapak dr. Soeroyo Mahfudz, MPH, SP.A(K) atas informasi dan pengetahuan yang dibutuhkan untuk penulisan laporan Tugas Akhir ini.

7. Pak Agus dan Pak Heri yang selalu menyediakan tempat, dan alat-alat di Lab. Dasar Teknik Elektro dan Lab. Komputer dan Simulasi.
8. Teman seperjuangan Viko Rona Sajri yang telah bersama-sama berjuang, membantu dan memberikan motivasi serta persahabatan selama ini.
9. Teman – teman Program Studi Teknik Elektro yang telah selalu mendukung dan membantu dalam senang maupun susah.
10. Dan banyak pihak lain yang tidak dapat kami sebutkan seluruhnya yang telah membantu dalam penyelesaian Tugas Akhir ini.

Akhir kata penulis sampaikan pula harapan semoga tugas akhir ini dapat memberikan manfaat yang cukup berarti khususnya bagi penulis dan bagi pembaca pada umumnya. Semoga Allah Subhanahu Wa Ta'ala senantiasa selalu memberikan rahmat dan hidayah – Nya kepada kita semua. Amiin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 8 Februari 2018

Penulis

ABSTRAK

Batuk adalah respon alamiah tubuh manusia pada saluran pernapasan. Di dalam dunia medis dikenal dengan dua jenis batuk yaitu batuk kering dan batuk basah. Pada anak-anak suara batuk dapat digunakan sebagai indikator suatu penyakit. Untuk itu perlu dilakukan metode klasifikasi suara batuk secara otomatis. Penelitian untuk membedakan suara batuk kering dan basah telah dilakukan. Namun penelitian tersebut masih memiliki kinerja kurang dari 90%. Hal ini bisa disebabkan oleh beberapa fitur tidak relevan. Pada penelitian ini, kami akan melakukan klasifikasi suara batuk kering dan batuk basah pada anak-anak dengan metode seleksi fitur. Pada penelitian ini melibatkan 30 subjek pasien anak-anak penderita penyakit pernapasan. Subjek tersebut terdiri dari 15 subjek penderita batuk kering dan 15 subjek penderita batuk basah. Metode seleksi fitur yang digunakan adalah algoritme genetika. Sedangkan pada proses klasifikasi menggunakan jaringan syaraf tiruan. Pada penelitian ini sebelum melakukan seleksi fitur kami melakukan klasifikasi suara batuk dengan jaringan syaraf tiruan. Hasil klasifikasi tersebut memiliki kinerja akurasi 96,42%, sensitivitas 96,68% dan spesifisitas 96,78%. Selanjutnya dilakukan seleksi fitur untuk menghilangkan fitur yang tidak relevan. Dari hasil seleksi fitur kinerja yang dihasilkan memiliki nilai akurasi 96,41%, sensitivitas 96,55% dan spesifisitas 96,65%. Dengan menghilangkan 25% fitur masih memiliki nilai kinerja yang relatif sama dibuktikan dengan nilai signifikansi, yaitu pada akurasi 0,00011, sensitivitas 0,00002 dan spesifisitas 0,00001.

Kata Kunci—Batuk kering, Batuk basah, Algoritme genetika, Jaringan syaraf tiruan.

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN.....	i
LEMBAR PERNYATAAN	ii
LEMBAR PENGESAHAN DOSEN PENGUJI.....	iii
KATA PENGANTAR.....	iv
ABSTRAK	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	2
1.5 Batasan Penelitian	2
BAB 2 TINJAUAN PUSTAKA	4
2.1 Tinjauan Pustaka	4
2.2 Landasan Teori	5
2.2.1 Suara Batuk.....	5
2.2.2 Seleksi Fitur Menggunakan Algoritme Genetika	5
2.2.3 Jaringan Syaraf Tiruan.....	7
BAB 3 METODOLOGI PENELITIAN.....	9
3.1 Akuisisi Data	9
3.2 Perancangan Dataset.....	10
3.3 Pengurangan <i>Noise</i>	11

3.4 Segmentasi.....	12
3.5 Ekstraksi Fitur	12
3.6 Normalisasi.....	15
3.7 Pengujian	15
3.7.1 Pengujian LRM.....	15
3.7.2 Seleksi Fitur LRM	16
3.7.3 Pengujian LRM Tanpa Fitur yang Telah Direduksi	17
3.7.4 Pengujian Jaringan Syaraf Tiruan.....	17
3.7.5 Seleksi Fitur Dengan Algoritme Genetika.....	17
3.7.6 Pengujian Jaringan Syaraf Tiruan Tanpa Fitur yang Telah Direduksi	18
BAB 4 HASIL DAN PEMBAHASAN.....	20
4.1 Dataset Suara Batuk	20
4.2 Pengujian Data.....	20
4.2.1 Pengujian Dengan <i>Logistic Regression Model</i> (LRM)	20
4.2.2 Seleksi Fitur Dengan LRM	21
4.2.3 Pengujian LRM Tanpa Fitur yang Telah Direduksi.....	22
4.2.4 Pengujian Dengan Jaringan Syaraf Tiruan.....	22
4.2.5 Seleksi Fitur Dengan Algoritme Genetika	23
4.2.6 Pengujian Jaringan Syaraf Tiruan Tanpa Fitur yang Telah Direduksi.....	24
BAB 5 KESIMPULAN DAN SARAN.....	28
5.1 Kesimpulan.....	28
5.2 Saran.....	28
DAFTAR PUSTAKA	29
LAMPIRAN	31
Lampiran 1 Fitur-fitur yang dihilangkan pada proses seleksi fitur	31

DAFTAR GAMBAR

Gambar 2.1 Siklus algoritme genetika	6
Gambar 2.2 Pindah silang satu titik pada kromosom induk dan menghasilkan kromosom anak ...	7
Gambar 2.3 Model neuron; $X_1 - X_n$ merupakan input layer, $W_1 - W_n$ merupakan bobot	8
Gambar 3.1 Diagram blok penelitian seleksi fitur dan klasifikasi suara batuk anak-anak	9
Gambar 3.2 Satu episode batuk dari mulai hingga akhir batuk pada anak-anak	10
Gambar 3.3 Sinyal suara batuk kering dan batuk basah pada anak-anak	11
Gambar 3.4 Urutan fitur ekstraksi yang digunakan pada penelitian	14
Gambar 3.5 Pembagian data <i>training</i> dan <i>testing</i> pada <i>cross validation</i>	16
Gambar 3.6 Pembagian data <i>training</i> , <i>validation</i> dan <i>testing</i> pada <i>cross validation</i>	17
Gambar 3.7 Prosedur penghilangan fitur algoritme genetika mulai 1 fitur hingga K fitur	18

DAFTAR TABEL

Tabel 4.1 Pengujian data dengan menggunakan LRM dengan semua metode ekstraksi fitur.....	20
Tabel 4.2 Nilai P masing-masing fitur terhadap <i>baseline</i>	21
Tabel 4.3 Hasil kinerja pengujian klasifikasi dengan tidak mengikutsertakan fitur yang direduksi	22
Tabel 4.4 Pengujian data dengan menggunakan jaringan syaraf tiruan dengan semua metode ekstraksi fitur.....	23
Tabel 4.5 Kinerja hasil seleksi fitur	23
Tabel 4.6 Tabel fitur yang dominan direduksi saat seleksi fitur	24
Tabel 4.7 Kinerja pengujian klasifikasi dengan tidak mengikutsertakan fitur yang direduksi lebih dari 1 kali.....	25
Tabel 4.8 Kinerja pengujian klasifikasi dengan tidak mengikutsertakan fitur yang direduksi sebanyak jumlah yang sama.....	26
Tabel 4.9 Kinerja pengujian klasifikasi dengan tidak mengikutsertakan semua fitur yang pernah direduksi	26

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Batuk merupakan pertahanan alami tubuh manusia yang terjadi pada saluran pernapasan. Batuk akan mencegah kotoran yang berbahaya masuk kedalam tubuh manusia melalui saluran pernapasan. Selain itu juga batuk akan membantu membersihkan saluran pernapasan manusia dari kotoran [1]. Di dalam dunia medis batuk dikategorikan menjadi dua yaitu, batuk basah (berdahak) dan batuk kering [2]. Kedua jenis batuk ini juga memiliki suara khas yang berbeda.

Pada awalnya batuk bukanlah dikategorikan kedalam sebuah penyakit, jika batuk tersebut tidak disertai dengan gejala-gejala lain. Batuk yang terjadi pada anak-anak dan orang dewasa juga memiliki perbedaan [3]. Pada anak-anak batuk dapat digunakan sebagai indikator sebuah penyakit yang diderita pada anak tersebut [3][4]. Batuk basah yang terjadi pada anak-anak mengindikasikan terjadinya infeksi pada saluran pernapasan bagian bawah [4][5]. Hal ini terjadi karena batuk basah mengindikasikan lendir yang berlebihan di saluran pernapasan anak-anak [6].

Pada saat ini klasifikasi suara batuk masih menggunakan keputusan subjektif yang dilakukan oleh dokter. Metode klasifikasi yang digunakan juga masih manual. Metode tersebut dilakukan dengan cara mendengarkan suara batuk yang terjadi pada pasien. Metode klasifikasi ini memiliki kekurangan yang diakibatkan oleh beberapa faktor. Salah satu faktornya adalah tidak semua dokter mampu untuk mengklasifikasikan suara batuk kering dan basah terutama pada anak-anak. Hal ini dipengaruhi oleh pengalaman yang dimiliki oleh dokter tersebut. Selain itu juga dapat dipengaruhi oleh kepekaan pendengaran dari dokter tersebut [7].

Untuk itu perlu dibuat suatu sistem yang mampu mengklasifikasikan suara batuk secara otomatis dan akurat. Beberapa penelitian telah dilakukan untuk dapat membuat sistem tersebut. Namun penelitian tersebut memiliki hasil yang kurang baik karena pada proses klasifikasi menggunakan metode linier [1][2]. Suara batuk merupakan sinyal yang memiliki sifat non-stasioner. Maka perlu digunakan sistem yang dapat menyelesaikan masalah non-linier. Selain itu proses klasifikasi suara batuk biasanya akan banyak sekali menggunakan metode ekstraksi fitur dan juga jumlah data yang cukup besar. Ditambah bisa jadi tidak semua fitur relevan dalam proses klasifikasi.

Pada penelitian ini peneliti mengusulkan sebuah metode klasifikasi suara batuk basah dan kering pada anak-anak. Metode yang diusulkan yaitu metode klasifikasi dan optimasi atau seleksi fitur. Metode klasifikasi yang akan digunakan adalah jaringan syaraf tiruan dan algoritma genetika sebagai algoritma optimasi atau seleksi fitur. Algoritma genetika akan digunakan untuk

mereduksi fitur-fitur yang tidak relevan. Algoritme genetika dipilih karena dapat memberikan hasil terbaik dengan cukup cepat pada masalah yang memiliki dimensi tinggi [8]. Dengan melakukan penelitian ini diharapkan proses diagnosa terhadap pasien menjadi lebih baik lagi.

1.2 Rumusan Masalah

Masalah yang dirumuskan pada tugas akhir ini adalah bagaimana mengimplementasikan algoritma genetika dalam melakukan seleksi fitur untuk mengoptimalkan proses klasifikasi suara batuk basah dan batuk kering pada anak-anak.

1.3 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah:

1. Mengimplementasikan algoritme genetika dan metode LRM dalam proses seleksi fitur pada klasifikasi suara batuk basah dan batuk kering pada anak-anak.
2. Mendapatkan fitur yang relevan untuk digunakan pada proses klasifikasi suara batuk basah dan batuk kering pada anak-anak.

1.4 Manfaat Penelitian

Manfaat yang didapatkan dari melakukan penelitian ini antara lain:

1. Kita dapat mengetahui fitur yang relevan dan fitur yang tidak relevan yang digunakan pada klasifikasi suara batuk basah dan batuk kering pada anak-anak.
2. Dapat menghilangkan fitur yang tidak relevan dengan mengetahui fitur yang relevan sehingga beban komputasi menjadi lebih ringan dan tanpa menurunkan nilai kinerja.
3. Dapat menghilangkan fitur-fitur *redundant*, sehingga beban komputasi akan menjadi berkurang.
4. Dapat melakukan klasifikasi suara batuk basah dan batuk kering pada anak-anak dengan mudah dan memiliki hasil yang baik.

1.5 Batasan Penelitian

Pada penelitian ini peneliti memberikan beberapa batasan masalah sebagai berikut:

1. Pada penelitian ini peneliti menggunakan ekstraksi fitur antara lain : *Non-Gaussian Score (NGS)*, *Formants Frequencies (FF)*, *Kurtosis*, *Log Energy (LogE)*, *Pitch*, *mel frequency cepstral coefficient, (MFCC)*, *Zero Crossing Rate (ZCR)* dan *Linear Predictive coding (LPC)*
2. Algoritme genetika yang digunakan pada penelitian ini yaitu metode seleksi roda rolet (*Roulette Wheel Selection*). Pada tahap pindah silang menggunakan metode *1-Point Crossover*.

3. Algoritme klasifikasi yang digunakan pada penelitian ini adalah algoritme jaringan syaraf tiruan dengan tiga lapisan tersembunyi.
4. *Software* yang digunakan MATLAB 2013, Audacity 2.1.1, Adobe Audition CC 2017
5. Perekaman data diperoleh dari pasien anak-anak dengan rentang usia 0 – 72 bulan di Rumah Sakit dr. Sardjito, Yogyakarta.

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Suara batuk cukup penting karena batuk dapat menjadi ciri-ciri awal dari beberapa penyakit pernapasan. Klasifikasi suara batuk secara manual memberikan hasil yang sangat subjektif [9]. Mengingat bahwa tindakan medis pada kedua jenis batuk ini sangat berbeda maka diperlukan klasifikasi suara batuk yang akurat. *Digital Signal Processing (DSP)* memungkinkan melakukan penelitian suara batuk melalui rekaman suara batuk yang diambil dari pasien. Untuk membedakan suara batuk dilakukan penelitian yang mempelajari berbagai fase batuk kering dan fase batuk basah oleh Chatzarrin. Mereka membagi batuk menjadi 3 fase, fase pertama pembukaan awal glotis, fase kedua aliran udara yang berisik dan fase ketiga penutupan glotis. Mereka mendapatkan bahwa batuk kering pada fase kedua memiliki energi yang lebih rendah dibandingkan batuk basah. Penelitian ini mencatat bahwa selama fase ini sebagian besar kekuatan sinyal terkandung pada batuk basah antara 0-750 Hz dan pada batuk kering antara 1500-2250 Hz. Penelitian ini berhasil mengidentifikasi 14 batuk basah dan batuk kering dengan akurasi 100%, menggunakan metode *thresholding* sederhana [10]. Namun penelitian yang dilakukahn oleh chatzarrin ini hanya menggunakan subjek yang sedikit yaitu 14 subjek. Sehingga penelitian yang dilakukan menjadi kurang baik. Mengingat bahwa suara batuk merupakan suara yang bersifat non-stasioner sehingga perlu melibatkan banyak subjek untuk melakukan penelitian.

Penelitian klasifikasi suara batuk basah dan batuk kering lainnya dilakukan dengan menggunakan metode baru yaitu *intra-cough shimmer jump (ICSJ)* dan *intra-cough teager energy operator jump (ICTeJ)* untuk membedakan batuk basah dan batuk kering. Hasil dari penelitian tersebut didapatkan nilai akurasi sebesar 76% dan nilai sensitivitas 77%, sedangkan nilai spesifisitas yang dihasilkan 75%. Penelitian tersebut menggunakan ekstraksi fitur *Temporal amplitude perturbation (shimmer)* dan *Teager energy operator perturbation* [9]. Pada penelitian lainnya, dilakukan untuk membedakan batuk basah dan batuk kering secara otomatis. Pada penelitian tersebut juga menggunakan metode klasifikasi *logistic regression model (LRM)*. Metode ekstraksi fitur yang digunakan *bispectrum score (BGS)*, *non-gaussianity score (NGS)*, *formants frequencies (FF)*, *log energy (LogE)*, *kurtosis (kurt)*, *pitch* dan *mel frequency cepstral coefficient (MFCC)*. Penelitian ini dilakukan terhadap subjek pasien anak-anak dengan penyakit pernapasan. Hasil yang didapatkan pada penelitian tersebut yaitu memiliki nilai sensitivitas sebesar 87% dan nilai spesifisitas sebesar 88% [1]. Klasifikasi suara batuk kering dan batuk

basah menggunakan metode *logistic regression model* (LRM) memiliki nilai sensitivitas, akurasi dan spesifisitasnya masih relatif rendah yaitu masih dibawah 90%.

Untuk menghasilkan kinerja klasifikasi yang lebih baik peneliti menggunakan konsep baru yang dapat digunakan untuk melakukan klasifikasi suara batuk basah dan kering pada anak-anak. Konsep yang gunakan tersebut adalah dengan melakukan seleksi fitur untuk memilih fitur mana saja yang berkontribusi terhadap target kelas. Selain itu juga dengan menggunakan metode seleksi fitur ini dapat mengurangi fitur-fitur yang *redundant*. Metode seleksi fitur yang digunakan adalah algoritme genetika. Sedangkan pada proses klasifikasi menggunakan jaringan syaraf tiruan. Penelitian yang dilakukan ini dengan melibatkan 30 subjek pasien. Diharapkan dengan digunakannya sebuah konsep baru ini dapat membantu proses klasifikasi suara batuk basah dan batuk kering yang lebih akurat.

2.2 Landasan Teori

2.2.1 Suara Batuk

Batuk dialami oleh setiap manusia dan merupakan mekanisme pertahanan penting dimana batuk akan mengeluarkan lendir, zat berbahaya dan infeksi laring, trakea dan bronkus. Batuk merupakan mekanisme yang efisien dalam membersihkan saluran pernafasan bagian atas [11]. Batuk juga dapat berfungsi untuk melindungi organ pernapasan bagian bawah dari benda asing yang masuk melalui saluran pernapasan. Selain menjadi mekanisme pertahanan, batuk juga merupakan gejala yang umum diamati pada beberapa penyakit pernapasan [12].

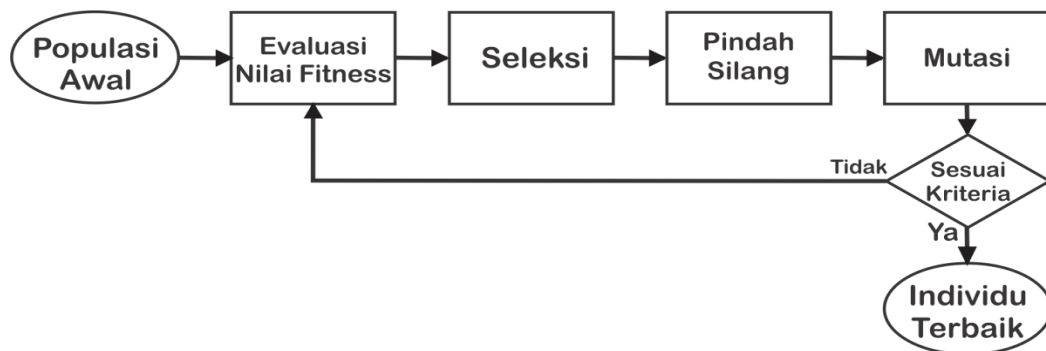
Mekanisme batuk dimulai dengan inspirasi, diikuti oleh penutupan glotis. Hal ini memungkinkan tekanan meningkat di dalam rongga *thoracic* melalui kompresi isometrik otot ekspirasi. Saat glotis terbuka, maka tekanan udara akan keluar dalam jumlah yang besar dari paru-paru secara singkat dan akan mengeluarkan kotoran atau sekresi [13].

2.2.2 Seleksi Fitur Menggunakan Algoritme Genetika

Seleksi fitur merupakan suatu tahapan yang terdapat pada proses klasifikasi suara. Seleksi fitur dilakukan pada tahap praproses data. Tujuan dari dilakukannya seleksi fitur adalah untuk mengurangi dimensi data, menghilangkan data yang tidak relevan, dan dapat meningkatkan akurasi. Selain dapat mengurangi fitur yang tidak relevan seleksi fitur juga dilakukan untuk menghilangkan fitur-fitur *redundant* [14]. Pada teknik seleksi fitur ini mempelajari bagaimana dapat mengidentifikasi dan memilih informasi fitur untuk membuat sebuah model [15].

Algoritme Genetika atau *Genetic Algorithm* (GA) dikenalkan oleh John Holland dalam menyelesaikan masalah optimasi [16]. Algoritme genetika merupakan salah satu algoritme optimasi yang dapat digunakan untuk seleksi fitur. Algoritme genetika bekerja seperti pada

proses evolusi alamiah atau seleksi alam, dimana individu terkuatlah yang akan bertahan hidup. Individu akan dilakukan seleksi berdasarkan nilai kebugarannya (*Fitness Value*). Gambar 2.1 merupakan siklus pada algoritme genetika.



Gambar 2.1 Siklus algoritme genetika

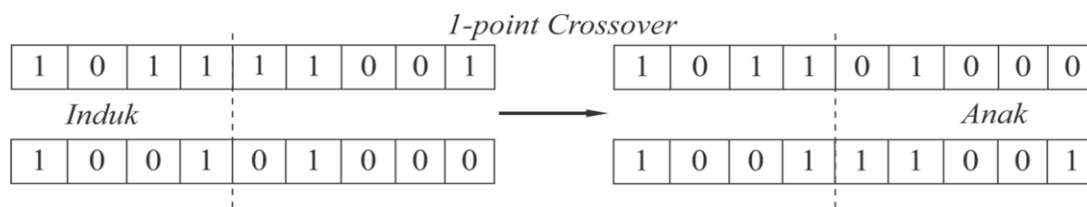
Pada populasi terdiri dari beberapa individu yang merupakan kandidat solusi. Individu dikatakan unggul jika individu tersebut memiliki nilai kebugaran yang tinggi. Individu unggul akan mampu menyesuaikan diri dan akan bertahan hidup, sedangkan yang tidak mampu bertahan akan mati [17]. Pada setiap iterasi pada algoritme ini (generasi), kemungkinan solusi (kromosom) dihasilkan menggunakan operator genetik. Operator genetik yang biasa digunakan pada algoritme genetika antara lain, mutasi (*mutation*), pindah silang (*crossover*), rekombinan (*recombinant*) [18].

Pada tahap seleksi individu terdapat beberapa metode yang dapat digunakan yaitu, *roulette wheel selection*, *tournament selection* dan *rank-based selection*. Namun metode seleksi yang sering digunakan adalah seleksi menggunakan roda rolet (*roulette wheel selection*) [17]. Pada seleksi roda rolet setiap kromosom yang ada akan menempati potongan lingkaran roda rolet dengan besar potongan proporsional sesuai dengan nilai kebugarannya. Probabilitas terpilihnya individu h_i ditentukan oleh Persamaan 2.1.

$$P(h_i) = \frac{F(h_i)}{\sum_{i=1}^p (h_i)} \quad (2.1)$$

dimana nilai $F(h_i)$ merupakan nilai kebugaran dari individu (h_i) [15].

Pindah silang memiliki tujuan untuk mewariskan sifat yang dimiliki oleh individu induk kepada anak yang dihasilkan. Operator pindah silang ini akan melibatkan paling sedikit 2 individu, dimana kromosom kromosom individu tersebut akan saling bertukar [19]. Terdapat beberapa metode pindah silang dalam algoritme genetika yaitu, pindah silang 1 titik, pindah silang n-titik dan pindah silang *shuffle*. Metode pindah silang yang paling sederhana ialah metode pindah silang 1 titik. Pindah silang 1 titik dapat dilihat pada Gambar 2.2.



Gambar 2.2 Pindah silang satu titik pada kromosom induk dan menghasilkan kromosom anak

Individu yang terpilih dari proses seleksi akan dilakukan pindah silang. Pada pindah silang 1 titik dengan memilih secara acak titik yang akan dilakukan pindah silang. Kemudian pada titik tersebut akan dipotong dan digabungkan kromosom dari kedua induk, sehingga akan menghasilkan 2 kromosom anak. Operator genetika lainnya adalah mutasi.

Mutasi merupakan operator genetika yang bersifat individual, yaitu operasi ini hanya dibutuhkan satu individu. Setiap individu memiliki kemungkinan untuk bermutasi. Besarnya kemungkinan mutasi pada setiap individu dipengaruhi dengan besarnya nilai probabilitas mutasi (P_m). Pada algoritme genetika ini disarankan untuk memberikan nilai yang kecil pada nilai probabilitas mutasi [19].

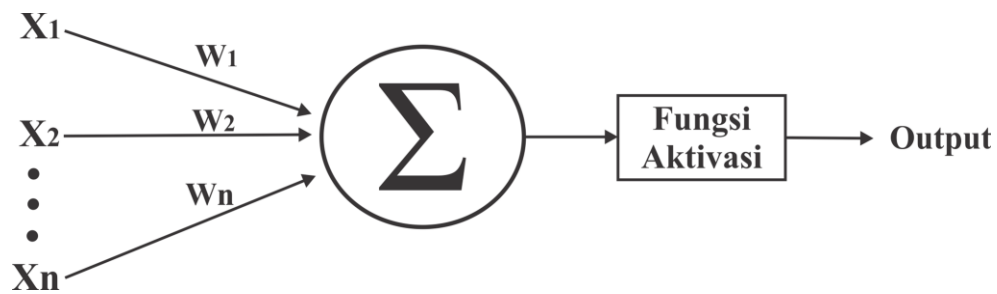
2.2.3 Jaringan Syaraf Tiruan

Jaringan syaraf tiruan merupakan sistem komputasi yang arsitektur dan operasinya terinspirasi oleh jaringan syaraf biologi pada otak [20]. Terdapat jutaan neuron pada otak manusia, neuron-neuron ini saling terhubung sehingga membentuk suatu jaringan [21]. Jaringan syaraf tiruan merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*). Jaringan syaraf tiruan ini pertama kali diperkenalkan oleh McCulloch dan Pitts pada tahun 1943 [21]. Jaringan syaraf tiruan pertama kali digunakan pada bidang kendali pada tahun 1964 oleh Widrow dan Smith yaitu *adaptive linear network (ADALINE)* [21]. Karakteristik yang dimiliki oleh jaringan syaraf tiruan adalah:

1. Pola hubungan antar neuron (arsitekturanya)
2. Metode penentuan bobot yang digunakan
3. Fungsi aktivasi

Jaringan syaraf tiruan memiliki beberapa elemen antara lain neuron, unit, sel atau *node* [21]. Dalam jaringan syaraf tiruan (JST) terdapat banyak neuron yang akan saling terhubung. Setiap neuron tersebut akan saling berkorelasi bersama membentuk jaringan sesuai dengan arsitektur eksplisitnya [22]. Bobot digunakan sebagai informasi yang akan digunakan untuk menyelesaikan permasalahan yang dihadapi. Sedangkan fungsi aktivasi berperan untuk

mengaktifkan atau menonaktifkan neuron. Gambar 2.3 merupakan gambar model neuron pada jaringan syaraf tiruan.



Gambar 2.3 Model neuron; $X_1 - X_n$ merupakan input layer, $W_1 - W_n$ merupakan bobot [21]

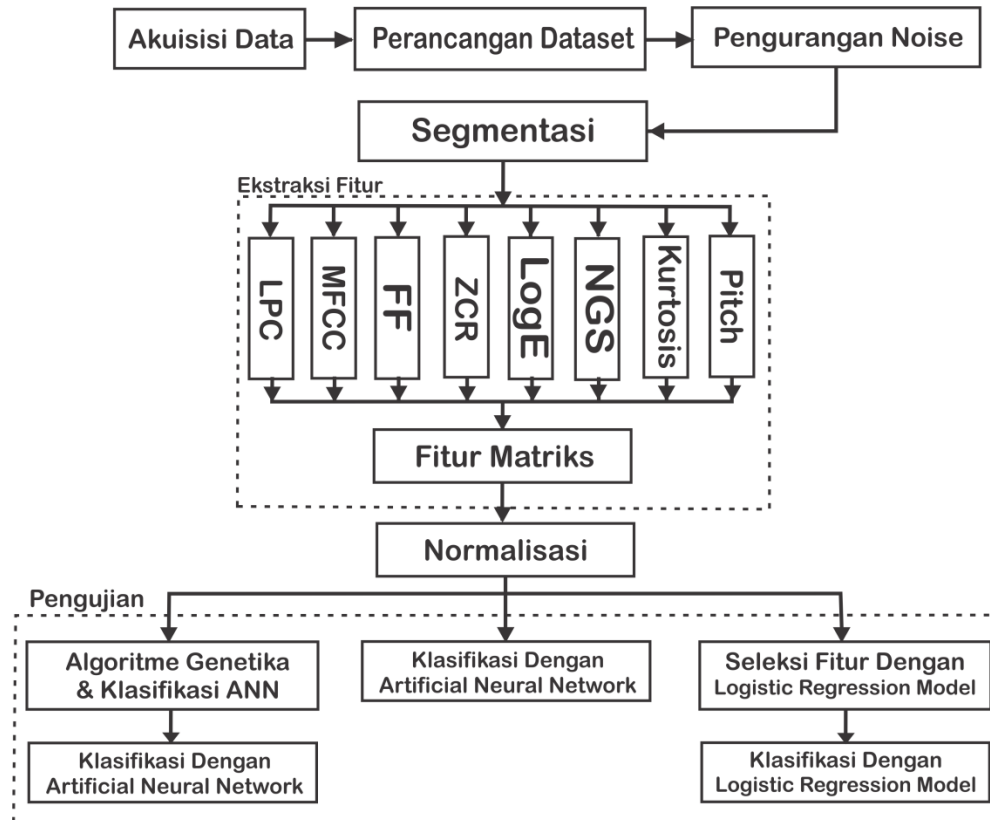
Jaringan syaraf tiruan memiliki tiga jenis arsitektur yaitu, jaringan lapisan tunggal (*single-layer*), jaringan lapisan multilapis (*multi-layer*) dan jaringan lapisan kompetitif (*competitive-layer*) [21]. Pada jaringan syaraf tiruan terdapat dua metode pembelajaran, yaitu metode pembelajaran terbimbing dan tidak terbimbing.

Jaringan syaraf tiruan telah banyak digunakan untuk melakukan klasifikasi dan pengenalan masalah non-linier [23]. Jaringan syaraf tiruan memiliki latar belakang teori yang baik dan pemetaan fitur dengan akurasi yang tinggi [24]. Oleh karena itulah mengapa pada penelitian ini jaringan syaraf digunakan sebagai metode klasifikasi. Jaringan syaraf tiruan yang digunakan adalah jaringan lapisan multilapis (*multi-layer*) yang memiliki tiga lapisan tersembunyi.

BAB 3

METODOLOGI PENELITIAN

Pada penelitian yang dilakukan dalam 7 tahap alur penelitian. Tahapan pada penelitian yang dilakukan ini digambarkan pada Gambar 3.1. Rincian tahapan yang dilakukan akan dijelaskan pada sub-bab 3.1-3.7.



Gambar 3.1 Diagram blok penelitian seleksi fitur dan klasifikasi suara batuk anak-anak

3.1 Akuisisi Data

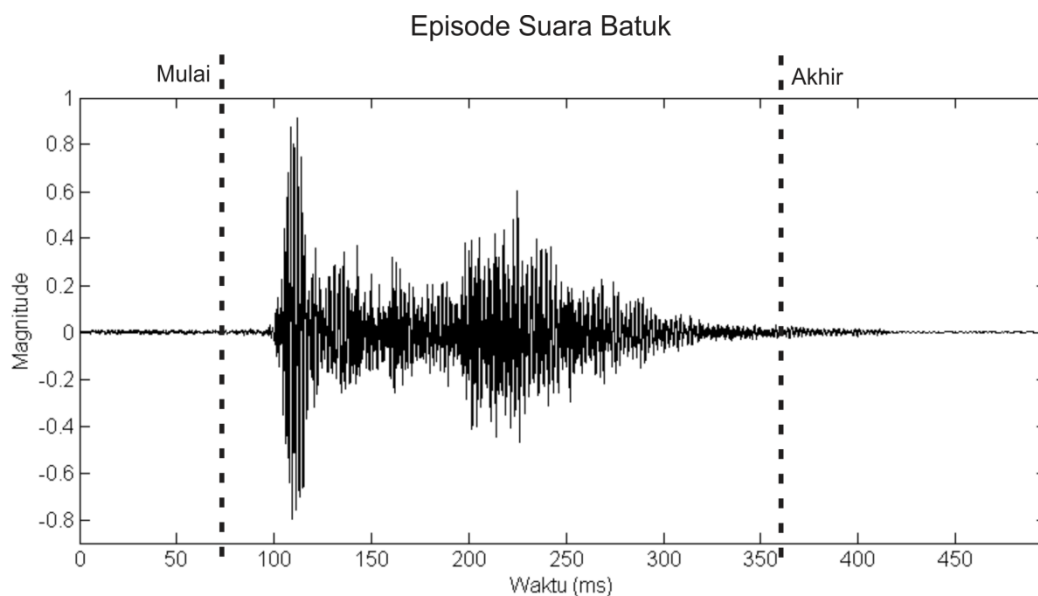
Pada penelitian ini pengambilan data dilakukan di rumah sakit Dr. Sardjito, Yogyakarta. Subjek yang diambil datanya merupakan pasien yang mengalami keluhan pada pernapasan. Pada penelitian ini subjek yang digunakan merupakan usia anak-anak yaitu antara 0 – 72 bulan.

Perekaman data suara batuk dilakukan dengan menggunakan dua buah mikrofon. Mikrofon yang digunakan adalah *low noise microphone* (Model NT3, RODE ®, Sydney, Australia) disertai *pre-amplifier*. Dan sebuah *Analog to Digital converter* (Model Mobile Pre-USB, M-Audio ®, CA, USA). *Sampling rate* yang digunakan 44100 Hz sampel/detik dengan resolusi 16-bit agar mendapatkan suara yang terbaik. Mikrofon dipasang didekat mulut pasien dengan jarak 50 cm, dimana satu mikrofon menghadap pasien sedangkan yang lainnya menghadap arah yang berlawanan. Penempatan mikrofon secara berlawanan arah ini bertujuan untuk memfasilitasi diferensiasi antara suara batuk dari subjek dengan suara latar belakang. Sumber latar belakang

itu berupa suara percakapan orang dewasa, langkah kaki, mainan anak-anak, suara yang dihasilkan oleh troli dan gerakan pintu. Panjang rekaman data yang dilakukan bervariasi antara 30 menit hingga 4 jam. Pada tahap perekaman data ini diambil data rekaman suara batuk (R) dari 30 subjek (ID).

3.2 Perancangan Dataset

Sebelum merancang dataset yang akan digunakan pada penelitian. Peneliti terlebih dahulu mendengarkan dataset rekaman suara batuk basah dan batuk kering yang telah dibuat oleh dokter yang ahli di bidang tersebut. Hal ini bertujuan agar peneliti mendapatkan pengetahuan dan dapat membedakan suara batuk basah dan batuk kering. Data rekaman R itu terdiri dari 15 subjek (ID1-ID15) pasien penderita batuk kering dan 15 subjek (ID16-ID30) pasien penderita batuk basah. Hasil dari data rekaman masing-masing subjek tersebut kemudian dilakukan pemilihan episode suara batuk. Satu episode batuk dihitung dari mulai terjadinya hingga akhir dari satu kali batuk seperti pada Gambar 3.2.



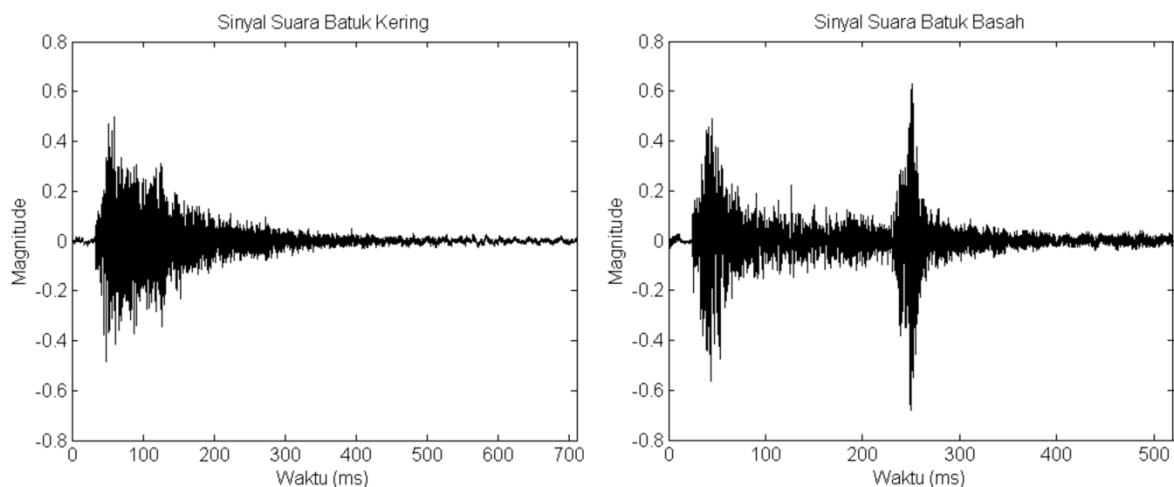
Gambar 3.2 Satu episode batuk dari mulai hingga akhir batuk pada anak-anak

Pemilihan episode suara batuk dilakukan dengan mendengarkan rekaman suara batuk dan melihat spektrum serta gelombangnya secara manual satu persatu. Perangkat lunak yang digunakan untuk mendengarkan rekaman suara batuk yaitu Adobe Audition. Contoh sinyal suara batuk kering dan batuk basah dapat dilihat pada Gambar 3.3.

Pada data rekaman tersebut peneliti mengambil 10 episode suara batuk terbaik dari masing-masing subjek. Kriteria episode suara batuk terbaik yang diambil adalah episode suara batuk yang tidak memiliki latar belakang *noise* dan data rekaman tidak terpotong (*clipping*).

Episode suara batuk yang terpilih kemudian dicatat waktu mulai batuk dan waktu berakhir batuknya. Dari data waktu yang telah dicatat ini maka data rekaman suara yang panjang tadi akan dipotong sesuai dengan waktu yang tercatat. Data rekaman tersebut dipotong dengan menggunakan *software* MATLAB. Pemotongan ini dilakukan pada semua data subjek yang akan digunakan pada penelitian.

Proses pemotongan data ini maka akan dihasilkan data rekaman suara batuk yang hanya berisi episode suara batuk saja. Data episode suara batuk yang didapatkan berjumlah 150 episode suara batuk kering dan 150 episode suara batuk basah, sehingga total episode suara batuk yang digunakan berjumlah 300 episode suara batuk. Kemudian episode suara batuk tersebut digabungkan menjadi sebuah dataset (D_s) suara batuk basah dan kering. Pada dataset (D_s) tersebut memiliki urutan yaitu pada bagian awal semua episode suara batuk kering dan dilanjutkan dengan episode suara batuk basah. Setelah dataset (D_s) digabungkan maka dicatat juga waktu mulai episode batuk kering dan basah serta waktu berakhirnya, pencatatan ini bertujuan untuk membuat target matriks. Target matriks (T_c) ini berguna untuk membedakan target kelas. Pada data batuk kering peneliti memberikan target bilangan biner 0 dan pada data batuk basah diberikan target bilangan biner 1.



Gambar 3.3 Sinyal suara batuk kering dan batuk basah pada anak-anak

3.3 Pengurangan *Noise*

Pada tahap ini dataset (D_s) yang telah dibentuk selanjutnya dilakukan pengurangan *noise*. Proses pengurangan *noise* ini dilakukan dengan menggunakan metode *High Pass Filter* (HPF) *Butterworth* orde 4. HPF dirancang untuk mengurangi *noise* yang kurang dari (f_c) 10 Hz. Penggunaan nilai (f_c) ini berdasarkan *noise* frekuensi rendah yang dihasilkan oleh getaran kabel dan tiang penyangga mikrofon. Pada tahap ini keluaran yang didapatkan dari proses pengurangan *noise* dinotasikan sebagai \widehat{D}_s .

3.4 Segmentasi

Pada dataset (\widehat{Ds}) akan dilakukan tahap berikutnya yaitu tahap segmentasi sinyal suara batuk. Pada tahap ini bertujuan untuk membagi sinyal suara batuk menjadi beberapa subblok sepanjang t . Pembagian subblok tersebut menggunakan Persamaan 3.1.

$$S = \frac{T}{t} \quad (3.1)$$

Dimana S merupakan jumlah subblok yang dihasilkan, dan T adalah waktu total dari dataset \widehat{Ds} dalam *second*. Sedangkan t panjang durasi tiap sub-blok yang digunakan untuk membagi sinyal suara batuk dalam *second*. Waktu pada tiap subblok yang digunakan pada penelitian ini adalah 20 ms dengan pergeseran antar subblok menggunakan *non-overlapping subblock*. Pemilihan waktu subblok 20 ms karena pada rentang waktu tersebut sinyal suara dianggap stasioner [25].

3.5 Ekstraksi Fitur

Setelah dilakukan segmentasi sinyal suara yang telah terbagi menjadi subblok tersebut diturunkan fitur matematisnya. Ekstraksi fitur dilakukan dengan menggunakan beberapa metode ekstraksi $f(n)$, dimana f adalah notasi dari setiap metode ekstraksi dan n merupakan urutan dari metode ekstraksi tersebut. Ekstraksi fitur ini dilakukan ke masing-masing subblok segmentasi (S) yang telah dibuat. Metode ekstraksi fitur yang digunakan yaitu :

1. Linear Predictive coding (LPC)

Metode LPC bekerja dengan beranggapan bahwa sinyal ucapan dihasilkan oleh *buzzer*. Pada kondisi batuk, saat fase ketiga glotis akan terbuka dan menghasilkan *buzz*, yang ditandai dengan intensitas (kenyaringan) dan frekuensi. Perhitungan LPC menggunakan Persamaan 3.2.

$$f_{(1)}(S) = \sum_{i=1}^h a_i S \quad (3.2)$$

Dimana $f_{(1)}(S)$ adalah nilai prediksi data hasil segmentasi (S). Sedangkan h adalah orde LPC dan a_i adalah koefisien prediktor LPC.

2. Mel frequency cepstral coefficient (MFCC)

MFCC sering digunakan dalam analisis suara musik dan suara ucapan [1]. MFCC meniru beberapa bagian dari manusia yang memproduksi ucapan dan bagian yang mempersepsi ucapan. Pada MFCC meniru persepsi logaritmik dari kenyaringan (*loudness*) dan suatu nada (*pitch*) dari sistem pendengaran manusia. Metode ini juga sangat berguna untuk menganalisis suara mendengkur dan suara batuk. MFCC memberikan representasi yang baik dari sifat saluran pernapasan bagian atas [26]. Pada penelitian ini digunakan 14 koefisien MFCC. Nilai MFCC dihitung pada data hasil segmentasi Y dengan Persamaan 3.3.

$$f_{(2)}(S) = \sum_c^C L_k(c) \cos \left\{ \frac{r(2c-1)\pi}{2C} \right\} \quad (3.3)$$

L_k merupakan keluaran *log energy* pada c mel filter banks dari data tiap subblok data segmentasi (S) dan r merupakan jumlah dari koefisien *cepstral*.

3. Formants Frequencies (FF)

Dalam analisis suara manusia, *formant* adalah resonansi dari saluran suara manusia. Pada analisis suara batuk peneliti beranggapan bahwa resonansi yang terjadi pada saluran suara manusia berkontribusi terhadap suara batuk yang direpresentasikan dalam *formant structure*. Lendir yang terdapat pada saluran suara dapat mengubah akustik dari saluran suara [1]. Pada penelitian ini digunakan 5 frekuensi *formant* pertama. Peneliti menghitung 5 frekuensi *formant* dengan menggunakan LPC. Orde LPC yang digunakan adalah orde 11.

4. Zero-Crossing Rate (ZCR)

ZCR merupakan suatu metode yang digunakan untuk mengetahui perubahan dari suatu sinyal. Perubahan sinyal yang diamati adalah tingkat perubahan dari positif ke negatif. Metode ini sangat sederhana namun dapat digunakan untuk mendeteksi sifat periodik dari sinyal. Metode ini dapat merepresentasikan getaran *glottis* dan juga dapat memisahkan sinyal suara dan tidak bersuara [26]. Nilai *zero-crossing rate* dihitung pada setiap data hasil segmentasi Y menggunakan Persamaan (3.4).

$$f_{(4)}(S) = \frac{1}{Tl-1} \sum_{t=1}^{T-1} H(S_t \cdot S_{t-1}) \quad (3.4)$$

Dimana S_t adalah sinyal dari data hasil segmentasi (S) saat waktu ke t . Tl adalah panjang dari data segmentasi (S) sedangkan H merupakan fungsi indikator (akan bernilai 0 saat argumen negatif dan bernilai 1 saat argumen positif).

5. Log Energy (LogE)

Fitur ini merupakan fitur pembantu yang digunakan pada metode MFCC, dimana fitur ini mampu meningkatkan akurasi dalam pengealan suara otomatis. *Log Energy* dihitung pada data hasil segmentasi Y menggunakan Persamaan 3.5.

$$f_{(5)}(S) = 10 \log_{10} \left(\varepsilon + \frac{1}{F} \sum_{k=1}^F (S^i(k)^2) \right) \quad (3.5)$$

Dengan nilai ε dalam (%) adalah konstanta positif kecil yang ditambahkan untuk menghindari perhitungan logaritma 0.

6. Non-Gaussian Score (NGS)

NGS menyediakan metode yang mudah untuk mengukur penyimpangan sinyal yang diberikan dari model *Gaussian*. Fitur ini menunjukkan kemampuan untuk menyaring apnea saat

tidur pada analisis suara mendengkur. *Non-Gaussianity* mungkin timbul saat glottis tiba-tiba terbuka saat fase ekspulsif [26]. Perhitungan fitur NGS dari data hasil segmentasi Y menggunakan Persamaan 3.6.

$$f_{(6)}(S) = 1 - \left(\frac{\sum_{j=1}^F (q[j] - p)^2}{\sum_{j=1}^F (q[j] - \bar{q})^2} \right) \quad (3.6)$$

Dimana nilai p dan q merepresentasikan *normal probability plot* dari data normal referensi dan data yang dianalisis. Sedangkan nilai j memiliki nilai antara 1 sampai dengan F.

7. Kurtosis

Kurtosis adalah ukuran dari pengukuran *peakedness* yang terkait dengan probabilitas distribusi dari data hasil segmentasi (S) [1]. Fitur *kurtosis* dihitung dengan menggunakan Persamaan 3.7.

$$f_{(7)}(S) = \frac{E(Y - \mu)^4}{\sigma^4} \quad (3.7)$$

Dimana μ adalah rata-rata hitung (mean) sedangkan σ adalah standar deviasi dari masing-masing data hasil segmentasi (S).

8. Pitch

Dalam analisis *pitch* (nada) didefinisikan sebagai frekuensi dasar yang ada pada pita suara [1]. Setiap orang mempunyai *pitch* yang khas yang dipengaruhi oleh aspek fisiologis manusia (laring). Pada saat kondisi batuk saraf dan otot di laring mengendalikan pembukaan dan penutupan pita suara yang diperlukan untuk membantu membersihkan saluran pernapasan.

Pada tahap ini akan didapatkan sebuah matriks yang memiliki dimensi $B \times F$, dimana F adalah banyaknya fitur yang digunakan. Matriks yang didapatkan tersebut dinamakan fitur matriks dan dinotasikan sebagai x . Persamaan fitur matriks dapat dilihat pada Persamaan 3.8.

$$x = [f_{(1)}(S) \ f_{(2)}(S) \ f_{(3)}(S) \ f_{(4)}(S) \ f_{(5)}(S) \ f_{(6)}(S) \ f_{(7)}(S) \ f_{(8)}(S)] \quad (3.8)$$

dimana :

$f_{(1)}(S)$ = dimensi matriks B x 8 fitur LPC; $f_{(5)}(S)$ = dimensi matriks B x 1 fitur LogE

$f_{(2)}(S)$ = dimensi matriks B x 14 fitur MFCC; $f_{(6)}(S)$ = dimensi matriks B x 1 fitur NGS

$f_{(3)}(S)$ = dimensi matriks B x 5 fitur FF; $f_{(7)}(S)$ = dimensi matriks B x 1 fitur Kurt

$f_{(4)}(S)$ = dimensi matriks B x 1 fitur ZCR; $f_{(8)}(S)$ = dimensi matriks B x 1 fitur Pitch

Urutan dari fitur yang terdapat pada fitur matriks x dapat dilihat pada Gambar 3.5.

LPC								MFCC														FF					ZCR	LogE	NGS	Kurt	Pitch
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32

Gambar 3.4 Urutan fitur ekstraksi yang digunakan pada penelitian

3.6 Normalisasi

Setelah data selesai diekstraksi maka data perlu dilakukan normalisasi agar data tersebut tidak memiliki jangkauan data yang terlalu jauh antara data tertinggi dan data terendah. Normalisasi yang digunakan yaitu antara 1 sampai dengan -1. Pada penelitian ini normalisasi menggunakan Persamaan 3.9.

$$x' = \frac{((2 \times x) - (x \max + x \min))}{(x \max - x \min)} \quad (3.9)$$

Dimana x adalah data yang akan dinormalisasi, sedangkan $x \min$ adalah data terkecil dari seluruh data, dan $x \max$ adalah data terbesar dari seluruh data [27].

3.7 Pengujian

Pada penelitian ini yang akan digunakan sebagai baseline adalah nilai kinerja yang dihasilkan dari proses klasifikasi. Parameter-parameter kinerja yang digunakan yaitu akurasi. Sensitivitas dan spesifisitas. Adapun masing-masing parameter tersebut dapat dihitung menggunakan Persamaan 3.10, Persaman 3.11 dan Persaman 3.12.

$$Akurasi = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (3.10)$$

$$Sensitivitas = \frac{TP}{(TP + FN)} \quad (3.11)$$

$$Spesifisitas = \frac{TN}{(FP + TN)} \quad (3.12)$$

Dimana TP adalah *True Positif*, FP adalah *False Positif*, TN adalah *True Negatif*, FN adalah *False Negatif*. TP, FP, TN dan FN merupakan jumlah keanggotan dari data input target yang disesuaikan dengan hasil klasifikasi.

3.7.1 Pengujian LRM

Logistic regression model (LRM) adalah pendekatan linier yang digunakan untuk memodelkan hubungan dengan menggunakan fungsi prediktor linier dimana parameter model yang pada awalnya tidak diketahui dapat diperkirakan atau diestimasi [28].

Pada tahap ini data matriks x' akan dilakukan pengujian klasifikasi menggunakan metode *logistic regression model* (LRM). Sebelum melakukan pengujian data matriks x' akan dibagi menjadi dua bagian yaitu, *training* dan *testing*. Dimana pembagian data matriks x' tersebut berdasarkan *sample* tiap subjek (ID). Setiap subjek akan secara bergantian digunakan sebagai *training* maupu *testing* (*N-Cross validation*). Penggunaan *Cross validation* dapat dilihat pada Gambar 3.5.

Training																										Testing					
ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID16	ID17	ID18	ID19	ID20	ID21	ID22	ID23	ID24	ID25	ID26	ID12	ID13	ID14	ID15	ID27	ID28	ID29	ID30		
ID15	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID30	ID16	ID17	ID18	ID19	ID20	ID21	ID22	ID23	ID24	ID25	ID11	ID12	ID13	ID14	ID26	ID27	ID28	ID29		
ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12	ID17	ID18	ID19	ID20	ID21	ID22	ID23	ID24	ID25	ID26	ID27	ID13	ID14	ID15	ID1	ID28	ID29	ID30	ID16		

Gambar 3.5 Pembagian data *training* dan *testing* pada *cross validation*

Pengujian dengan LRM ini akan menghasilkan keluaran yang dinotasikan dengan L' . Keluaran dari LRM L' ini akan dilakukan perhitungan kinerja menggunakan Persamaan 3.10, Persaman 3.11 dan Persaman 3.12.

Sehingga akan didapatkan nilai akurasi, sensitivitas dan spesifisitas dari pengujian LRM. Nilai akurasi dinotasikan dengan $accL'$ dan nilai sensitivitas dinotasikan dengan $SensL'$ sedangkan nilai spesifisitas yang dinotasikan dengan $SpecL'$. Nilai kinerja yang didapatkan ini akan dijadikan *baseline*.

3.7.2 Seleksi Fitur LRM

Pemilihan fitur ini dilakukan berdasarkan P -value, dimana fitur yang memiliki nilai signifikansi tinggi merupakan fitur yang tidak relevan. Untuk melakukan seleksi fitur pengujian klasifikasi dengan LRM dilakukan pada setiap metode ekstraksi satu persatu ($f_{(1)}(S)$ $f_{(2)}(S)$ $f_{(3)}(S)$ $f_{(4)}(S)$ $f_{(5)}(S)$ $f_{(6)}(S)$ $f_{(7)}(S)$ $f_{(8)}(S)$). Sehingga akan didapatkan keluaran berupa $L'f_{(i)}$, dimana nilai i adalah urutan metode fitur ekstraksi yang digunakan. Pengujian ini juga diterapkan metode *cross validation* seperti pada pengujian pembuatan *baseline*. Setelah didapatkan keluaran pengujian pada masing masing metode fitur ekstraksi, maka dihitung kinerjanya menggunakan Persamaan 3.10, Persaman 3.11 dan Persaman 3.12. Dari proses ini akan didapatkan nilai kinerja dari masing masing metode ekstraksi yang dinotasikan sebagai $SensL'f_{(i)}$ pada sensitifitas dan $SpecL'f_{(i)}$ pada spesifisitasnya.

Nilai $SensL'f_{(i)}$ dan $SpecL'f_{(i)}$ akan digunakan untuk memutuskan apakah fitur tersebut berkontribusi atau tidak. Dari nilai $SensL'f_{(i)}$ dan $SpecL'f_{(i)}$ akan dihitung nilai signifikasinya (P -value) terhadap nilai $SensL'$ dan $SpecL'$ yang telah didapatkan dari pengujian sebelumnya. Setelah mendapatkan nilai signifikansinya maka, tahap selanjutnya melakukan seleksi. Seleksi dilakukan dengan memberikan nilai batas P_{thd} , jika nilai signifikansi lebih kecil dari nilai P_{thd} maka fitur dimasukkan sebaliknya jika lebih besar dari P_{thd} maka fitur tersebut akan dieleminasi. Pada penelitian ini digunakan nilai $P_{thd} = 0,06$.

3.7.3 Pengujian LRM Tanpa Fitur yang Telah Direduksi

Pada tahap ini dilakukan pengujian klasifikasi kembali. Pada tahap ini klasifikasi dilakukan dengan tidak mengikutsertakan fitur yang telah direduksi pada saat seleksi fitur. Tahap ini bertujuan untuk mengetahui kinerja dari seleksi fitur dengan menggunakan LRM.

3.7.4 Pengujian Jaringan Syaraf Tiruan

Pada tahap ini dilakukan pengujian klasifikasi dengan menggunakan jaringan syaraf tiruan. Pada pengujian ini data matriks x' akan digunakan sebagai masukan pada jaringan syaraf tiruan. Pada pengujian ini juga dilakukan *cross validation*, namun *cross validation* yang dilakukan sedikit berbeda. Pada pengujian ini data matriks x' akan dibagi menjadi 3 bagian, yaitu *training*, *validation* dan *testing*. Penggunaan *cross validation* pada pengujian ini dapat dilihat pada Gambar 3.6.

Training															Validation					Testing									
ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID16	ID17	ID18	ID19	ID20	ID21	ID22	ID23	ID24	ID25	ID11	ID12	ID13	ID26	ID27	ID28	ID14	ID15	ID29	ID30
ID15	ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID30	ID16	ID17	ID18	ID19	ID20	ID21	ID22	ID23	ID24	ID10	ID11	ID12	ID25	ID26	ID27	ID13	ID14	ID28	ID29
ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID17	ID18	ID19	ID20	ID21	ID22	ID23	ID24	ID25	ID26	ID12	ID13	ID14	ID27	ID28	ID29	ID15	ID1	ID30	ID16

Gambar 3.6 Pembagian data *training*, *validation* dan *testing* pada *cross validation*

Jaringan syaraf tiruan memiliki struktur *input layer*, *hidden layer* dan *output layer*. Pada pengujian ini peneliti menggunakan *hidden layer* 3 lapis. Pada *hidden layer* memiliki jumlah *neuron* masing-masing secara berurut yaitu 155, 20, 1.

Keluaran dari jaringan syaraf tiruan ini dinotasikan dengan A' . Setelah didapatkan nilai A' maka kembali dilakukan perhitungan performa menggunakan Persamaan 3.10, Persaman 3.11 dan Persaman 3.12. Dari sini akan didapatkan nilai akurasi, sensitivitas, spesifisitas, yang secara berurutan dinotasikan sebagai $accA'$, $SensA'$ dan $SpecA'$.

3.7.5 Seleksi Fitur Dengan Algoritme Genetika

Pada tahap ini dilakukan pengujian dengan menggunakan algoritme optimasi / seleksi fitur. algoritme ini akan memilih fitur yang akan dihilangkan atau tetap dipertahankan. Algoritme optimasi yang digunakan pada penelitian ini adalah algortime genetika.

Pada proses pengujian seleksi fitur dengan algoritme genetika ini matriks x' digunakan sebagai masukan. Proses seleksi fitur dilakukan dengan membuat populasi dengan membangkitkan bilangan acak. Dimana pada populasi tersebut terdiri dari banyak subset fitur. Subset fitur yang terbentuk ini digunakan sebagai kromosom (I) pada algoritme genetika. Kromosom (I) ini akan dilakukan evaluasi nilai kebugaran, untuk memberikan nilai kebugaran

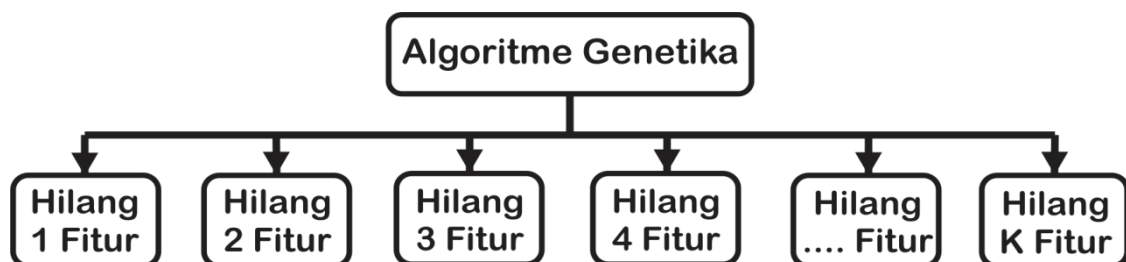
pada tiap-tiap kromosom. Nilai kebugaran diberikan berdasarkan fungsi kebugaran yang digunakan. Pada penelitian ini fungsi kebugaran dihitung berdasarkan nilai *error* dari klasifikasi menggunakan jaringan syaraf tiruan. Persamaan nilai kebugaran pada penelitian ini seperti pada Persamaan 3.13.

$$I = c(x') \quad (3.13)$$

Dimana I adalah merupakan representasi kromosom dari subset fitur yang terpilih. Sedangkan $c(x')$ adalah nilai *error* dari klasifikasi jaringan syaraf tiruan.

Operator seleksi yang digunakan pada penelitian ini adalah seleksi roda rolet (*Roulette Wheel Selection*). Dalam penelitian ini metode pindah silang yang digunakan adalah pindah silang 1 titik. Pindah silang 1 titik ini merupakan pindah silang yang paling sederhana, dimana akan ada sebuah titik yang akan membagi 2 kromosom induk untuk kemudian dilakukan pindah silang. Mutasi dilakukan secara acak, dimana setiap individu memiliki kemungkinan bermutasi sebesar nilai probabilitas mutasinya.

Pada proses seleksi fitur ini algoritme genetika akan menghilangkan fitur-fitur ($f_{(1)}(S) f_{(2)}(S) f_{(3)}(S) f_{(4)}(S) f_{(5)}(S) f_{(6)}(S) f_{(7)}(S) f_{(8)}(S) ; j = 1, 2, 3, \dots, \text{maksimal jumlah fitur}$). Proses seleksi fitur ini memiliki prsedur seperti pada Gambar 3.7.



Gambar 3.7 Prosedur penghilangan fitur algoritme genetika mulai 1 fitur hingga K fitur

Pada awalnya algoritme genetika akan menghilangkan 1 fitur terlebih dahulu dan dilihat apakah kinerjanya mengalami penurunan yang cukup besar. Jika kinerja masih mengalami penurunan yang signifikan maka akan dilakukan seleksi fitur dengan menghilangkan 2 fitur dan dilihat kembali kinerjanya. Proses penghilangan fitur dilakukan hingga menghilangkan fitur sebanyak K-fitur. Dimana K-fitur akan memiliki kinerja klasifikasi yaitu sensitifitas dan spesifisitas 2% lebih rendah dari kinerja *baseline*-nya.

3.7.6 Pengujian Jaringan Syaraf Tiruan Tanpa Fitur yang Telah Direduksi

Pada tahap ini dilakukan kembali pengujian dengan jaringan syaraf tiruan. Pengujian ini dilakukan untuk mengetahui apakah hasil seleksi fitur yang telah dilakukan mendapatkan hasil yang cukup baik. Pengujian ini dilakukan dengan menggunakan jaringan syaraf tiruan dengan

hidden layer 3 lapis masing masing *hidden layer* memiliki *neuron* 155, 20, 1. Pada penelitian ini dilakukan beberapa pengujian antara lain:

1. Pengujian dengan tidak mengikutsertakan fitur yang direduksi lebih dari 1 kali
2. Pengujian dengan tidak mengikutsertakan fitur yang pernah direduksi dalam jumlah yang sama
3. Pengujian dengan tidak mengikutsertakan semua fitur yang menjadi fitur dominan direduksi

BAB 4

HASIL DAN PEMBAHASAN

4.1 Dataset Suara Batuk

Pada penelitian ini digunakan data rekaman suara batuk yang didapatkan dari 30 pasien sebagai subjek. Pasien yang dijadikan subjek merupakan pasien anak-anak penderita penyakit pernapasan dengan rentang usia 0-72 bulan. Data rekaman yang digunakan terdiri dari 15 subjek pasien penderita batuk basah dan 15 subjek pasien penderita batuk kering. Dari masing-masing subjek diambil 10 episode suara batuk.

Pada dataset yang digunakan ini memiliki panjang durasi yaitu 173 detik. Dimana pada dataset tersebut terdiri dari 96 detik suara rekaman batuk basah dan 77 detik suara rekaman batuk kering. Keseluruhan data episode suara batuk yang digunakan yaitu 300 episode yaitu, 150 episode batuk kering dan 150 episode batuk basah.

4.2 Pengujian Data

4.2.1 Pengujian Dengan *Logistic Regression Model* (LRM)

Pengujian ini melakukan klasifikasi suara batuk basah dan kering pada anak-anak dengan LRM. Pada tahap ini klasifikasi dilakukan dengan mengikutsertakan semua fitur. Tahap ini belum dilakukan seleksi fitur. Kinerja yang didapatkan dari pengujian ini terdapat pada Tabel 4.1.

Tabel 4.1 Pengujian data dengan menggunakan LRM
dengan semua metode ekstraksi fitur

	Rata-rata (%)	Standar Deviasi
Akurasi	49,00	6,94
<i>Error</i>	50,00	6,94
Sensitivitas	49,05	6,74
Spesifisitas	49,09	6,74

Dapat dilihat pada tabel 4.1 bahwa kinerja klasifikasi menggunakan metode LRM memiliki hasil yang kurang baik. Nilai akurasi, sensitivitas dan spesifisitasnya masih sangat rendah yaitu akurasi 49,99%, sensitifitas 49,05% dan spesifisitas 49,09%. Selain itu juga pengujian dengan LRM memiliki nilai standar deviasinya cukup tinggi yaitu 6,74% hingga 6,94%. Hasil kinerja yang telah dihasilkan dari pengujian LRM akan digunakan sebagai *baseline* untuk perbandingan antara pengujian sebelum dilakukan seleksi fitur dengan setelah dilakukan seleksi fitur.

4.2.2 Seleksi Fitur Dengan LRM

Pada tahap ini dilakukan seleksi fitur dengan menggunakan LRM. Prosedur seleksi fitur pada tahap ini dengan melakukan klasifikasi pada tiap-tiap fitur. Setelah didapatkan nilai kinerja dari klasifikasi LRM, maka dihitung nilai signifikansi (P -value) dari hasil klasifikasi masing masing fitur terhadap nilai *baseline*. Nilai kinerja yang akan dijadikan acuan untuk dihitung nilai signifikansinya adalah nilai sensitivitas dan spesifisitas. Setelah mendapatkan nilai signifikansi dari kinerja tersebut akan dilakukan pemilihan fitur. Pemilihan ini dilakukan dengan menggunakan nilai *threshold* (P_{thd}). Nilai P_{thd} yang digunakan pada penelitian ini adalah 0,06. Jika nilai signifikansi dari kinerja lebih besar dari nilai P_{thd} maka fitur direduksi, sedangkan jika nilai signifikansi lebih kecil dari P_{thd} maka fitur dipertahankan. Nilai signifikansi dari sensitivitas dan spesifisitas yang dihasilkan pada proses seleksi fitur menggunakan LRM ini dapat dilihat pada Tabel 4.2.

Tabel 4.2 Nilai P masing-masing fitur terhadap *baseline*

Fitur ke	Nilai P		Fitur ke	Nilai P	
	Sensitivitas	Spesifisitas		Sensitivitas	Spesifisitas
1	0,2987	0,2962	17	0,0291	0,0294
2	0,5909	0,5890	18	0,0166	0,0168
3	0,7937	0,7912	19	0,0025	0,0024
4	0,2860	0,2898	20	0,0073	0,0072
5	0,0003	0,0003	21	0,0057	0,0058
6	0,2845	0,2845	22	0,0209	0,0213
7	0,8487	0,8364	23	0,3467	0,3427
8	0,0070	0,0070	24	0,3579	0,3644
9	0,7864	0,7770	25	0,1214	0,1193
10	0,7868	0,7917	26	0,2809	0,2829
11	0,9602	0,9577	27	0,0043	0,0043
12	0,0886	0,0888	28	0,0758	0,0631
13	0,2083	0,2082	29	0,6694	0,6524
14	0,4003	0,3950	30	0,0732	0,0753
15	0,6927	0,6965	31	0,3163	0,3222
16	0,5172	0,5201	32	0,5144	0,4833

Dapat dilihat pada Tabel 4.2 pada beberapa fitur tersebut memiliki nilai signifikansi lebih kecil dari nilai P_{thd} yaitu fitur ke 5, 8, 17, 18, 19, 20, 21, 22, 27. Fitur-fitur yang memiliki nilai signifikansi lebih kecil dari nilai P_{thd} ini akan dipertahankan, sedangkan fitur-fitur yang lainnya

akan direduksi. Pada proses ini terdapat 23 fitur yang direduksi yaitu fitur ke 1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 23, 24, 25, 26, 28, 29, 30, 32.

4.2.3 Pengujian LRM Tanpa Fitur yang Telah Direduksi

Pada Tabel 4.3 dapat dilihat kinerja hasil klasifikasi setelah dilakukan seleksi fitur. Proses ini dengan tidak mengikut sertakan fitur ke 1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 23, 24, 25, 26, 28, 29, 30, 32. Pada tahap ini dilakukan klasifikasi dengan menggunakan metode LRM.

Tabel 4.3 Hasil kinerja pengujian klasifikasi dengan tidak mengikutsertakan fitur yang direduksi

Tanpa Fitur	Kinerja		
	Akurasi (%)	Sensitivitas (%)	Spesifisitas(%)
1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 23, 24, 25, 26, 28, 29, 30, 32	54,3 ± 5,2	49,4 ± 2,9	49,4 ± 2,9

Dapat dilihat pada Tabel 4.3 bahwa nilai kinerja yang didapatkan saat pengujian setelah melakukan seleksi adalah akurasi 54,3%, sensitivitas 49,4% dan spesifisitas 49,4%. Kinerja tersebut memiliki standar deviasi 5,2% pada akurasi, 2,9% pada sensitivitas dan 2,9% pada spesifisitasnya. Nilai kinerja klasifikasi yang dihasilkan setelah dilakukan seleksi fitur meningkat jika dibandingkan dengan kinerja *baseline*-nya, yaitu pada *baseline* memiliki kinerja klasifikasi pada akurasi 49,99%, sensitivitas 49,05%, dan spesifisitas 49,09%. Sedangkan pada nilai standar deviasi memiliki nilai yang lebih rendah jika dibandingkan dengan nilai standar deviasi pada *baseline*. Pada *baseline* memiliki nilai standar deviasi yaitu pada akurasi 6,94%, sensitivitas 6,74% dan pada spesifisitas 6,74%.

4.2.4 Pengujian Dengan Jaringan Syaraf Tiruan

Pada pengujian ini dilakukan klasifikasi suara batuk dengan menggunakan jaringan syaraf tiruan. Proses klasifikasi ini dilakukan dengan menggunakan data fitur hasil ekstraksi fitur semua metode. Hasil dari pengujian ini akan digunakan sebagai *baseline* pada tahap seleksi fitur. Pada pengujian ini dilakukan beberapa kali dan didapatkan nilai rata-rata pengujian seperti pada Tabel 4.4. Nilai kinerja yang didapatkan ini selanjutnya akan digunakan sebagai nilai *baseline* dari penelitian yang dilakukan ini. Nilai tersebut akan menjadi pembanding dengan nilai kinerja setelah dilakukan seleksi fitur.

Tabel 4.4 Pengujian data dengan menggunakan jaringan syaraf tiruan dengan semua metode ekstraksi fitur

	Rata-rata (%)	Standar Deviasi
Akurasi	96,42	0,75
<i>Error</i>	3,57	0,75
Sensitivitas	96,68	0,79
Spesifisitas	96,78	0,79

Pada Tabel 4.4 dapat dilihat bahwa pengujian klasifikasi menggunakan jaringan syaraf tiruan menghasilkan kinerja yang cukup baik karena kinerja yang dihasilkan memiliki nilai diatas 90%. Dapat dilihat bahwa nilai masing-masing kinerja yang didapatkan yaitu akurasi 96,42%, sensitivitas 96,68% dan spesifisitas 96,78%. Nilai standar deviasi dari pengujian ini juga cukup rendah yaitu pada akurasi 0,75%, sensitivitas 0,79% dan spesifisitas 0,79%.

4.2.5 Seleksi Fitur Dengan Algoritme Genetika

Pengujian ini melakukan seleksi fitur untuk mereduksi fitur yang digunakan. Prosedur reduksi fitur yang digunakan adalah pada tahap ini algoritme seleksi fitur akan menghilangkan 1 fitur. Setelah 1 fitur dihilangkan maka seleksi fitur akan menghilangkan 2 fitur, 3 fitur hingga K-fitur. Penentuan jumlah K-fitur yang dihilangkan dilihat dari kinerja sensitivitas dan spesifisitas hasil seleksi fitur. Pada penelitian yang dilakukan ini peneliti menggunakan nilai batas 2%. Jadi proses reduksi fitur akan berhenti setelah kinerja sensitivitas dan spesifisitas yang didapatkan saat menghilangkan fitur sebanyak K dibawah 2% dari kinerja *baseline*-nya.

Tabel 4.5 Kinerja hasil seleksi fitur

Fitur yang Dieleminasi	Akurasi (%)	<i>Error</i> (%)	Sensitivitas (%)	Spesifisitas (%)
1	96,4 ± 0,7	3,57 ± 0,7	96,2 ± 0,8	96,3 ± 0,8
2	96,7 ± 0,7	3,2 ± 0,7	96,5 ± 0,7	96,6 ± 0,7
3	96,6 ± 0,7	3,3 ± 0,7	96,5 ± 0,8	96,6 ± 0,8
4	96,5 ± 0,8	3,4 ± 0,8	96,4 ± 1,0	96,5 ± 1,0
5	96,6 ± 1,0	3,3 ± 1,0	96,2 ± 1,3	96,3 ± 1,3
6	96,4 ± 1,0	3,5 ± 1,0	96,2 ± 1,1	96,3 ± 1,1
7	96,3 ± 0,9	3,6 ± 0,9	96,3 ± 1,0	96,4 ± 1,0
8	96,1 ± 0,8	3,8 ± 0,8	96,0 ± 0,8	96,1 ± 0,8
9	95,9 ± 1,2	4,0 ± 1,2	95,7 ± 1,0	95,8 ± 1,0
10	95,5 ± 1,2	4,4 ± 1,2	95,3 ± 1,2	95,4 ± 1,2
11	95,5 ± 1,2	4,4 ± 1,2	95,5 ± 1,4	95,6 ± 1,3
12	94,7 ± 1,5	5,2 ± 1,5	94,8 ± 1,5	94,9 ± 1,6
13	94,2 ± 1,5	5,7 ± 1,5	94,2 ± 1,6	94,3 ± 1,6

Pada Tabel 4.5 dapat dilihat kinerja yang dihasilkan dari proses seleksi fitur dari menghilangkan 1 fitur hingga menghilangkan K-fitur. Dapat dilihat pada saat menghilangkan 1 fitur nilai kinerja yang dihasilkan yaitu sensitivitas 96,2% dan spesifisitas 96,3% nilai kinerja ini masih belum mencapai 2% dibawah kinerja *baseline*. Oleh karena itu akan dilakukan proses selanjutnya dengan menghilangkan fitur lebih banyak lagi. Pada proses penghilangan 2 hingga 12 fitur nilai kinerja sensitivitas dan spesifisitas masih belum memiliki nilai 2% di bawah nilai *baseline* yaitu dengan nilai antara 96,6% hingga 94,8%. Pada saat mereduksi 13 fitur nilai kinerja yang dihasilkan yaitu sensitivitas 94,2% dan spesifisitas 94,3%. Nilai kinerja saat mengilangkan fitur sebanyak 13 ini sudah 2% dibawah nilai kinerja *baseline*, karena nilai kinerja yang dihasilkan sudah 2% lebih rendah dari nilai *baseline* maka proses reduksi fitur dihentikan.

4.2.6 Pengujian Jaringan Syaraf Tiruan Tanpa Fitur yang Telah Direduksi

1. Pengujian dengan tidak mengikutsertakan fitur yang direduksi lebih dari 1 kali

Pada proses seleksi fitur menghasilkan fitur yang direduksi secara dominan. Fitur-fitur yang direduksi berbeda beda tiap menghilangkan K-fitur. Pada Tabel 4.6 dapat dilihat fitur mana saja yang dominan direduksi dari pengujian seleksi fitur sebanyak K = 13 kali.

Tabel 4.6 Tabel fitur yang dominan direduksi saat seleksi fitur

Jumlah Fitur yang Direduksi	Direduksi Sebanyak	Fitur ke
1	7	32
2	4	31
3	7	32
4	9	32
5	9	32
6	7	3
7	13	24
8	11	32
9	11	32
10	11	29
11	12	24, 28, 32
12	12	25
13	13	27

Pada Tabel 4.6 diketahui bahwa pada setiap penghilangan fitur sebanyak K terdapat fitur-situr yang dominan direduksi. Fitur-fitur tersebut adalah fitur 3 yaitu fitur ke 3 dari LPC, lalu fitur 24, 25, 27 yang secara berurut merupakan fitur ke 2, 3, 5 dari fitur *formants frequencies*, fitur 28 merupakan fitur ZCR, fitur 29 merupakan fitur LogE, fitur 31 merupakan fitur *kurt* dan fitur 32 merupakan fitur *pitch*. Dapat dilihat bahwa fitur 32 dan fitur 24 menjadi fitur yang paling dominan direduksi. Dari K = 13 kali pengujian fitur 32 dihilangkan sebanyak 6 kali. Sedangkan pada fitur 24 dihilangkan sebanyak 2 kali. Maka pada tahap ini dilakukan pengujian dengan tidak mengikutkan fitur 32 dan 24 secara bergantian.

Tabel 4.7 Kinerja pengujian klasifikasi dengan tidak mengikutsertakan fitur yang direduksi lebih dari 1 kali

Direduksi Sebanyak	Tanpa Fitur ke	Kinerja		
		Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)
6	32	96,4 ± 0,8	96,5 ± 0,8	96,7 ± 0,8
2	24	96,1 ± 0,9	96,3 ± 0,8	96,4 ± 0,8

Dapat dilihat pada Tabel 4.7 kinerja yang didapatkan pada proses klasifikasi dengan tidak mengikutsertakan fitur ke 32 dan 24 secara bergantian masih cukup baik. Nilai kinerja yang didapatkan pada saat tidak mengikut sertakan fitur ke 32 yaitu akurasi 96,4%, sensitivitas 96,5% dan spesifisitas 96,7%, selain itu nilai standar deviasinya masih cukup rendah yaitu 0,8% pada akurasi, sensitivitas dan spesifisitas. Pada pengujian dengan tidak mengikutsertakan fitur ke 24 nilai kinerja yang didapatkan yaitu akurasi 96,1%, sensitivitas 96,3% dan spesifisitas 96,4%, dan pada nilai standar deviasinya juga masih cukup rendah yaitu akurasi 0,9% dan 0,8% pada sensitivitas dan spsifistasnya.

2. Pengujian dengan tidak mengikutsertakan fitur yang pernah direduksi dalam jumlah yang sama

Dari Tabel 4.6 dapat dilihat bahwa terdapat fitur yang direduksi sebanyak jumlah yang sama. Fitur ke 3, 6, 25, 27, 28, 29, 31 direduksi secara dominan sebanyak 1 kali. Tahap ini akan melakukan pengujian dengan tidak mengikut sertakan fitur-fitur tersebut.

Tabel 4.8 Kinerja pengujian klasifikasi dengan tidak mengikutsertakan fitur yang direduksi sebanyak jumlah yang sama

Direduksi Sebanyak	Tanpa Fitur ke	Kinerja		
		Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)
1	3, 25, 27, 28, 29, 31	95,9 ± 1,0	96,1 ± 1,1	96,2 ± 1,1

Pada Tabel 4.8 dapat dilihat bahwa kinerja yang dihasilkan dari pengujian dengan tidak mengikutsertakan fitur yang pernah direduksi dalam jumlah yang sama. Pada pengujian ini terdapat 7 fitur yang direduksi saat pengujian. Kinerja yang dihasilkan memiliki hasil yang cukup baik dengan akurasi 95,9%, sensitivitas 96,1% dan spesifisitas 96,2%. Nilai standar deviasi yang dihasilkan juga masih cukup rendah yaitu akurasi 1,0% dan 1,1% pada sensitivitas dan spesifisitas.

3. Pengujian dengan tidak mengikutsertakan semua fitur yang menjadi fitur dominan direduksi

Pengujian ini akan dilakukan dengan tidak mengikutsertakan semua fitur yang dominan direduksi oleh algoritme seleksi fitur. Dapat dilihat pada Tabel 4.6 terdapat fitur-fitur yang dominan direduksi saat seleksi fitur. Fitur tersebut merupakan fitur 3, 6, 24, 25, 27, 28, 29, 31, 32. Pada tahap ini akan dilakukan proses klasifikasi dengan tidak mengikutsertakan semua fitur yang pernah menjadi fitur yang dominan direduksi dari proses penghilangan fitur ke-K.

Tabel 4.9 Kinerja pengujian klasifikasi dengan tidak mengikutsertakan semua fitur yang pernah direduksi

Tanpa Fitur ke	Kinerja		
	Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)
3, 24, 25, 27, 28, 29, 31, 32	96,4 ± 0,6	96,5 ± 0,6	96,6 ± 0,6

Pada Tabel 4.9 dapat dilihat bahwa hasil klasifikasi dari pengujian ini memiliki nilai kinerja yang cukup baik. Ini dibuktikan dengan menghilangkan 8 fitur kinerjanya masih cukup tinggi yaitu akurasi 96,4%, sensitivitas 96,5% dan spesifisitas 96,6%. Nilai standar deviasinya juga masih cukup rendah yaitu 0,6%. Jika nilai kinerja yang dihasilkan pada tahap ini dibandingkan dengan nilai kinerja *baseline*, maka nilai kinerjanya memiliki nilai signifikansi yang cukup rendah yaitu akurasi 0,00011, sensitivitas 0,00002 dan spesifisitas 0,00001. Mengingat bahwa pada pengujian ini tanpa menggunakan 8 fitur atau 25% dari keseluruhan fitur yang digunakan pada penelitian ini.

Pada penelitian yang dilakukan ini kinerja yang didapatkan pada proses klasifikasi suara batuk basah kering anak-anak dengan menggunakan metode LRM masih memiliki nilai yang rendah jika dibandingkan dengan dengan kinerja klasifikasi suara batuk basah kering anak-anak dengan menggunakan jaringan syaraf tiruan. Metode LRM yang digunakan pada penelitian ini memiliki hasil kinerja yang sangat rendah karena memiliki nilai kinerja dibawah 50% dapat pada Tabel 4.1, dan memiliki hasil yang lebih rendah juga jika dibandingkan dengan penelitian yang telah dilakukan sebelumnya. Pada penelitian sebelumnya tersebut memiliki kinerja yaitu sensitivitas 91% dan spesifisitas 92%.

Hasil klasifikasi suara batuk basah kering anak-anak dengan menggunakan metode LRM terjadi peningkatan kinerja setelah dilakukan seleksi fitur, namun nilai kinerja yang dihasilkan masih cukup rendah karena berada dibawah 90% seperti pada Tabel 4.3, jika dibandingkan dengan penelitian sebelumnya yang memiliki hasil sensitivitas 87% dan spesifisitas 88%. Pada saat proses seleksi fitur menggunakan algoritme genetika kinerja yang didapatkan memiliki hasil yang sangat baik, hasil yang didapatkan memiliki kinerja diatas 90% dapat dilihat pada Tabel 4.9. Oleh karena itu dapat dikatakan bahwa proses seleksi menggunakan algoritme genetika memiliki hasil yang lebih baik jika dibandingkan dengan menggunakan metode LRM.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan beberapa pengujian yang telah dilakukan pada penelitian ini dapat disimpulkan bahwa. Kinerja dari klasifikasi sebelum dilakukan seleksi fitur dengan LRM yaitu akurasi $49,00\% \pm 6,94\%$, sensitivitas $49,05\% \pm 6,74\%$ dan spesifisitas $49,09\% \pm 6,74\%$. Hasil kinerja ini masih sangat rendah jika dibandingkan hasil kinerja dari klasifikasi menggunakan jaringan syaraf tiruan sebelum seleksi fitur yaitu nilai akurasi, sensitivitas, dan spesifisitas berurut adalah $96,42\% \pm 0,75\%$, $96,68\% \pm 0,79\%$, $96,78\% \pm 0,79\%$. Nilai kinerja tersebut dijadikan sebagai nilai *baseline*. Seleksi yang dilakukan dengan LRM dapat meningkatkan kinerja dari klasifikasi, namun nilai kinerja yang didapatkan dari hasil klasifikasi ini masih cukup rendah yaitu akurasi $54,36\% \pm 5,22\%$, sensitivitas $49,45\% \pm 2,94\%$ dan spesifisitas $49,48\% \pm 2,95\%$. Peningkatan nilai kinerja yang terjadi $\pm 3,45\%$ dari kinerja sebelum dilakukan seleksi. Nilai kinerja ini dikatakan rendah karena masih dibawah 90%. Algoritme genetika mampu digunakan untuk mereduksi 8 fitur dengan mempertahankan kinerja klasifikasi suara batuk basah/kering pada anak-anak. Fitur yang diusulkan untuk dihilangkan adalah fitur 3 yaitu fitur ke 3 dari LPC, fitur 24, 25, 27 secara berurut adalah fitur ke 3,4 dan 5 dari *formant frequencies*, fitur 28 adalah fitur ZCR, fitur 29 adalah fitur LogE, fitur 31 adalah fitur *kurt* dan fitur 32 yang merupakan fitur *pitch*. Dengan mereduksi fitur tersebut nilai kinerja klasifikasi yang dihasilkan yaitu dengan nilai akurasi $96,41\% \pm 0,6\%$, sensitivitas $96,55\% \pm 0,6\%$ dan spesifisitas $96,65\% \pm 0,6\%$.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan ini, peneliti memberikan beberapa saran yang dapat dilakukan untuk penelitian yang selanjutnya

1. Saat melakukan *screening* data suara batuk sebaiknya di konsultasikan dengan orang yang ahli dibidang tersebut (dokter) agar algoritme yang dibuat lebih andal.
2. Menambah dataset/subjek yang digunakan untuk penelitian sehingga sistem yang dihasilkan menjadi lebih baik lagi.
3. Menambah fitur yang digunakan pada penelitian, karena masih banyak fitur-fitur yang dapat digunakan pada penelitian suara batuk yang lain. Sehingga dapat mengetahui apakah fitur-fitur yang lain memiliki kontribusi yang besar atau tidak dalam proses klasifikasi suara batuk.

DAFTAR PUSTAKA

- [1] V. Swarnkar, U. R. Abeyratne, A. B. Chang, Y. A. Amrulloh, A. Setyati, and R. Triasih, "Automatic Identification of Wet and Dry Cough in Pediatric Patients with Respiratory Diseases," vol. 41, no. 5, pp. 1016–1028, 2013.
- [2] V. Swarnkar and U. R. Abeyratne, "Automated Algorithm for Wet/Dry Cough Sounds Classification," pp. 3147–3150, 2012.
- [3] A. B. Chang, "Cough : are children really different to adults?," vol. 15, pp. 1–15, 2005.
- [4] U. R. Abeyratne, V. Swankar, A. Setyati, and R. Triasih, "Cough Sound Analysis Can Rapidly Diagnose Childhood Pneumonia," 2013.
- [5] A. B. Chang, L. I. Landau, P. P. Van Asperen, N. Glasgow, C. F. Robertson, J. M. Marchant, and C. M. Mellis, "Cough in children: Definitions and clinical evaluation," no. May, 2006.
- [6] A. B. Chang, "State of the Art Chronic Wet Cough : Protracted Bronchitis , Chronic Suppurative Lung Disease and Bronchiectasis," vol. 531, no. January, pp. 519–531, 2008.
- [7] T. L. Wiley, R. Chappell, L. Carmichael, D. M. Nondahl, and J. Karen, "Changes in Hearing Thresholds over 10 Years in Older Adults," vol. 19, no. 4, 2010.
- [8] Suyanto, *Algoritma Genetika Dalam MATLAB*. Yogyakarta: Andi Offset, 2005.
- [9] Y. A. Amrulloh, D. A. R. Wati, F. Pratiwi, and R. Triasih, "A Novel Method for Wet / Dry Cough Classification in Pediatric Population," pp. 125–129, 2016.
- [10] H. Chatrzarrin, A. Arcelus, R. Goubran, and F. Knoefel, "Feature Extraction for the Differentiation of Dry and Wet Cough Sounds," pp. 0–4, 2011.
- [11] K. F. Chung, *Cough: Causes, Mechanisms and Therapy*. .
- [12] F. De Blasio, J. C. Virchow, M. Polverino, A. Zanasi, P. K. Behrakis, G. Kiliñç, R. Balsamo, G. De Danieli, and L. Lanata, "Cough management : a practical approach," pp. 1–12, 2011.
- [13] R. Palmer, J. B. Anon, and P. Gallagher, "Pediatric cough : what the otolaryngologist needs to know," 2011.
- [14] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data : A Fast Correlation-Based Filter Solution," 2003.

- [15] F. Tan, X. Fu, Y. Zhang, and A. G. Bourgeois, "A genetic algorithm-based method for feature subset selection," pp. 111–120, 2008.
- [16] H. A. Saputro, W. F. Mahmudy, and C. Dewi, "Implementasi Algoritma Genetika Untuk Optimasi Penggunaan Lahan Pertanian," no. 12, 2015.
- [17] J. Zhong, X. Hu, M. Gu, and J. Zhang, "Comparison of Performance between Different Selection Strategies on Simple Genetic Algorithms," vol. 1, 2005.
- [18] J. K. F.J. Ferry, V. Kadirkamanthan, "Feature Subset Search Using Genetic Algorithm."
- [19] H. D. Purnomo, *Cara Mudah Belajar Metode Optimasi Metaheuristik menggunakan MatLab*. Yogyakarta: Gava Media, 2014.
- [20] A. Kristanto, *Jaringan Syaraf Tiruan: Konsep Dasar, Algoritma dan Aplikasi*. Yogyakarta: Gava Media, 2004.
- [21] D. A. R. Wati, *Sistem Kendali Cerdas*. Yogyakarta: Graha Ilmu, 2011.
- [22] S. Sasithra, "Review On Classification Based On Artificial Neural Network," vol. 2, no. 4, pp. 11–18, 2014.
- [23] J. Zurada, *Introduction to Artificial Neural Systems*. 1992.
- [24] J. Korbicz, J. M. Koscielny, Z. Kowalczyk, and W. Cholewa, *Fault Diagnosis*. Berlin, Germany: Springer-Verlag, 2004.
- [25] S. Drs, D. Dorran, M. Gainza, and E. Coyle, "Linear Prediction," no. August, 2008.
- [26] Y. A. Amrulloh, U. R. Abeyratne, V. Swarnkar, R. Triasih, and A. Setyati, "Automatic cough segmentation from non-contact sound recordings in pediatric wards," *Biomed. Signal Process. Control*, vol. 21, pp. 126–136, 2015.
- [27] E. Presetyo, *DATA MINING Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: Andi Offset, 2014.
- [28] "Linear regression," 2018. [Online]. Available: https://en.wikipedia.org/wiki/Linear_regression. [Accessed: 20-Jan-2018].

LAMPIRAN

Lampiran 1 Fitur-fitur yang dihilangkan pada proses seleksi fitur

		Jumlah Fitur Yang Dihilangan Pada Proses Seleksi Fitur GA				
		1	2	3	4	5
Kombinasi Data	1	32	16, 32	12, 23, 32	3, 32, 24, 28	31, 17, 5, 30, 8
	2	32	29, 18	32, 8, 11	9, 26, 22, 17	23, 7, 5, 22, 26
	3	31	23, 2	27, 7, 5	31, 20, 32, 21	4, 23, 27, 28, 9
	4	32	23, 20	27, 7, 5	31, 9, 12, 7	25, 7, 26, 32, 6
	5	10	31, 7	6, 31, 10	13, 10, 9, 31	25, 7, 26, 32, 6
	6	30	31, 24	4, 8, 18	2, 25, 1, 32	23, 6, 15, 25, 24
	7	32	31, 6	9, 18, 3	2, 25, 1, 32	25, 7, 26, 32, 6
	8	32	5, 32	25, 26, 27	22, 24, 23, 12	3, 8, 4, 6, 7
	9	32	9, 5	26, 13, 30	25, 8, 3, 32	7, 4, 26, 2, 25
	10	10	9, 5	25, 11, 30	32, 7, 27, 31	31, 23, 5, 6, 29
	11	30	1, 28	32, 20, 9	32, 7, 27, 31	31, 23, 5, 6, 29
	12	32	25, 22	3, 32, 16	28, 27, 32, 25	29, 10, 4, 9, 32
	13	10	32, 6	22, 5, 32	23, 24, 14, 3	9, 8, 14, 7, 23
	14	10	21, 29	3, 32, 16	2, 27, 10, 22	25, 7, 26, 32, 6
	15	10	31, 7	3, 32, 16	2, 25, 1, 32	9, 16, 32, 14, 6

		Jumlah Fitur Yang Dihilangkan Pada Proses Seleksi Fitur GA			
		6	7	8	9
Kombinasi Data	1	8, 22, 4, 11, 1, 26	31, 27, 11, 18, 25, 23, 32	24, 7, 10, 27, 5, 25, 32, 1	18, 17, 5, 26, 9, 32, 27, 28, 25
	2	6, 4, 32, 26, 8, 29	11, 31, 24, 21, 29, 3, 23	3, 29, 24, 1, 17, 23, 25, 32	2, 7, 27, 29, 5, 22, 10, 31, 25
	3	25, 30, 22, 23, 31, 4	11, 31, 24, 21, 29, 3, 23	6, 9, 26, 24, 32, 4, 30, 25	32, 16, 27, 25, 26, 31, 24, 2, 23
	4	24, 23, 17, 6, 30, 9	2, 3, 4, 32, 9, 7, 24	3, 29, 24, 1, 17, 23, 25, 32	12, 10, 6, 23, 1, 5, 24, 11, 2
	5	3, 9, 31, 30, 4, 26	24, 23, 2, 25, 27, 28, 31	4, 1, 9, 3, 15, 5, 7, 23	9, 31, 8, 3, 10, 32, 5, 24, 29
	6	2, 32, 3, 7, 20, 24	27, 29, 32, 24, 28, 26, 8	4, 1, 9, 3, 15, 5, 7, 23	32, 14, 11, 19, 6, 16, 2, 9, 1
	7	2, 32, 3, 7, 20, 24	24, 27, 32, 30, 31, 7, 3	3, 29, 24, 1, 17, 23, 25, 32	32, 8, 3, 25, 24, 4, 9, 30, 16
	8	5, 10, 28, 22, 25, 3	27, 29, 32, 24, 28, 26, 8	8, 22, 27, 10, 7, 4, 28, 32	32, 8, 3, 25, 24, 4, 9, 30, 16
	9	27, 8, 28, 11, 6, 7	27, 29, 32, 24, 28, 26, 8	31, 7, 32, 25, 27, 1, 17, 21	2, 7, 27, 29, 5, 22, 10, 31, 25
	10	2, 32, 3, 7, 20, 24	27, 29, 32, 24, 28, 26, 8	31, 7, 32, 25, 27, 1, 17, 21	27, 11, 16, 25, 26, 32, 30, 4, 7
	11	3, 6, 19, 32, 31, 2	2, 3, 4, 32, 9, 7, 24	31, 7, 32, 25, 27, 1, 17, 21	24, 13, 8, 32, 1, 29, 3, 15, 5
	12	29, 7, 12, 5, 31, 10	9, 27, 26, 24, 17, 31, 7	23, 30, 28, 22, 32, 5, 21, 3	21, 32, 26, 15, 28, 5, 27, 17, 2
	13	29, 7, 12, 5, 31, 10	25, 28, 8, 24, 27, 5, 17	3, 31, 4, 26, 1, 24, 10, 23	27, 11, 16, 25, 26, 32, 30, 4, 7
	14	2, 4, 21, 23, 6, 32	24, 27, 32, 30, 31, 7, 3	26, 27, 29, 6, 32, 8, 4, 11	21, 12, 2, 24, 23, 27, 10, 3, 26
	15	17, 3, 22, 29, 10, 27	5, 31, 25, 18, 12, 32, 2	22, 6, 4, 30, 7, 26, 8, 24	24, 2, 32, 13, 14, 5, 29, 1, 10

		Jumlah Fitur Yang Dihilangkan Pada Proses Seleksi Fitur GA			
		10	11	12	13
Kombinasi Data	1	24, 22, 4, 16, 9, 18, 1, 32, 21, 29	10, 18, 30, 25, 14, 27, 32, 9, 28, 24, 6	6, 24, 12, 32, 2, 10, 25, 23, 31, 7, 17, 19	27, 29, 31, 32, 2, 15, 30, 14, 9, 18, 10, 5, 28
	2	2, 7, 9, 27, 29, 4, 23, 12, 31, 25	2, 32, 4, 20, 3, 31, 29, 16, 28, 10, 30	11, 27, 24, 17, 1, 29, 12, 25, 18, 28, 31, 10	10, 29, 25, 28, 11, 4, 27, 18, 6, 30, 26, 7, 9
	3	30, 7, 5, 14, 11, 6, 8, 9, 2, 32	7, 26, 25, 28, 32, 2, 4, 5, 27, 3, 6	31, 9, 7, 32, 25, 17, 26, 23, 2, 22, 5, 28	30, 1, 29, 7, 9, 3, 28, 25, 31, 24, 8, 22, 18
	4	20, 16, 32, 31, 5, 27, 17, 7, 21, 10	2, 29, 10, 11, 24, 32, 21, 15, 5, 9, 28	26, 1, 5, 30, 23, 6, 14, 17, 3, 15, 28, 25	4, 15, 11, 26, 5, 8, 28, 7, 24, 14, 27, 1, 21
	5	13, 26, 4, 6, 1, 23, 29, 9, 21, 5	18, 4, 9, 6, 3, 31, 8, 28, 29, 24, 30	1, 4, 12, 9, 30, 2, 11, 31, 6, 25, 18, 3	5, 31, 27, 15, 3, 24, 9, 8, 25, 1, 2, 6, 32
	6	4, 11, 7, 30, 6, 2, 29, 24, 9, 3	16, 28, 22, 30, 8, 5, 9, 4, 6, 10, 24	18, 27, 14, 23, 11, 5, 22, 26, 13, 32, 24, 25	5, 31, 27, 15, 3, 24, 9, 8, 25, 1, 2, 6, 32
	7	4, 6, 27, 20, 9, 24, 23, 32, 18, 3	32, 27, 11, 29, 24, 28, 1, 2, 23, 5, 17	20, 9, 24, 6, 30, 1, 32, 26, 3, 10, 4, 18	27, 19, 26, 23, 29, 32, 30, 24, 3, 2, 10, 6, 14
	8	30, 22, 3, 29, 24, 27, 8, 25, 32, 2	32, 27, 11, 29, 24, 28, 1, 2, 23, 5, 17	22, 20, 24, 8, 12, 31, 5, 7, 32, 28, 30, 25	30, 1, 29, 7, 9, 3, 28, 25, 31, 24, 8, 22, 18
	9	30, 22, 3, 29, 24, 27, 8, 25, 32, 2	7, 26, 25, 28, 32, 2, 4, 5, 27, 3, 6	22, 20, 24, 8, 12, 31, 5, 7, 32, 28, 30, 25	5, 31, 27, 15, 3, 24, 9, 8, 25, 1, 2, 6, 32
	10	30, 22, 3, 29, 24, 27, 8, 25, 32, 2	7, 8, 24, 21, 4, 26, 13, 3, 25, 6, 32	8, 27, 20, 17, 10, 32, 21, 3, 24, 25, 31, 5	32, 19, 31, 27, 26, 10, 12, 5, 6, 24, 25, 11, 22
	11	23, 27, 30, 29, 31, 28, 6, 32, 5, 21	18, 4, 9, 6, 3, 31, 8, 28, 29, 24, 30	12, 28, 15, 21, 1, 31, 17, 3, 18, 2, 29, 32	8, 25, 5, 31, 26, 3, 27, 6, 9, 16, 23, 2, 4
	12	23, 8, 3, 9, 25, 14, 31, 29, 6, 15	10, 2, 16, 30, 22, 27, 4, 1, 6, 24, 32	31, 9, 7, 32, 25, 17, 26, 23, 2, 22, 5, 28	8, 25, 5, 31, 26, 3, 27, 6, 9, 16, 23, 2, 4
	13	2, 7, 9, 27, 29, 4, 23, 12, 31, 25	13, 32, 12, 25, 23, 3, 6, 26, 27, 1, 24	8, 27, 20, 17, 10, 32, 21, 3, 24, 25, 31, 5	6, 2, 26, 3, 14, 31, 20, 22, 24, 27, 10, 7, 23
	14	8, 32, 16, 9, 6, 4, 7, 29, 13, 1	32, 27, 11, 29, 24, 28, 1, 2, 23, 5, 17	20, 9, 24, 6, 30, 1, 32, 26, 3, 10, 4, 18	1, 4, 12, 9, 27, 2, 11, 32, 6, 23, 17, 3, 28
	15	8, 5, 4, 7, 32, 24, 6, 10, 22, 23	32, 6, 9, 26, 24, 13, 28, 4, 1, 8, 22	11, 17, 23, 24, 16, 7, 25, 15, 9, 3, 6, 17	24, 6, 7, 20, 26, 18, 31, 32, 27, 11, 19, 1, 30