

**KOMPARASI ALGORITMA OPTIMASI UNTUK SELEKSI
FITUR KLASIFIKASI PENYAKIT *STROKE* PADA
ALGORITMA BERBASIS POHON KEPUTUSAN**



Skripsi

Gavriel Willy

0000043681

UMN

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

**KOMPARASI ALGORITMA OPTIMASI UNTUK SELEKSI
FITUR KLASIFIKASI PENYAKIT *STROKE* PADA
ALGORITMA BERBASIS POHON KEPUTUSAN**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Sistem Informasi

Gavriel Willy

00000043681

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Gavriel Willy

Nomor Induk Mahasiswa : 00000043681

Program studi : Sistem Informasi

Skripsi dengan judul:

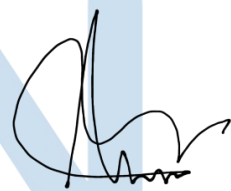
Komparasi Algoritma Optimasi Untuk Seleksi Fitur Klasifikasi Penyakit *Stroke* pada Algoritma Berbasis Pohon Keputusan

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 28 Mei 2024

UMMN



(Gavriel Willy)

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERSETUJUAN

Skripsi dengan judul
Komparasi Algoritma Optimasi Untuk Seleksi Fitur Klasifikasi Penyakit *Stroke*
pada Algoritma Berbasis Pohon Keputusan

Oleh

Nama : Gavriel Willy
NIM : 00000043681
Program Studi : Sistem Informasi
Fakultas : Teknik Informatika

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 14 Mei 2024

Pembimbing



Dinar Ajeng Kristiyanti, S.Kom, M.Kom
0330128801

Ketua Program Studi Sistem Informasi



14.05.2024

Ririn Ikana Desanti, S.Kom, M.Kom
0313058001

HALAMAN PENGESAHAN

Skripsi dengan judul

Komparasi Algoritma Optimasi Untuk Seleksi Fitur Klasifikasi Penyakit *Stroke*
pada Algoritma Berbasis Pohon Keputusan

Oleh

Nama : Gavriel Willy

NIM : 00000043681

Program Studi : Sistem Informasi

Fakultas : Teknik Informatika

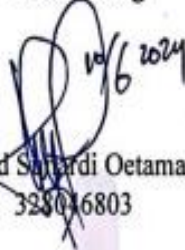
Telah diujikan pada hari Selasa, 28 Mei 2024

Pukul 00.00 s.d 00.00 dan dinyatakan

LULUS

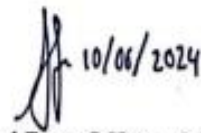
Dengan susunan penguji sebagai berikut.

Ketua Sidang



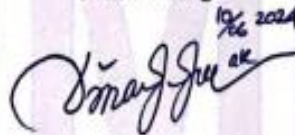
Ir. Raymond Suardi Oetama, M.CIS
323016803

Penguji



Ahmad Faza, S.Kom., M.T.I.
0312019501

Pembimbing



Dinar Ajeng Krishiyanti, S.Kom, M.Kom
0330123801

Ketua Program Studi Sistem Informasi



Ririn Ikana Desanti, S.Kom, M.Kom
0313058001

HALAMAN PERSETUJUAN PUBLIKASI

KARYA ILMIAH MAHASISWA

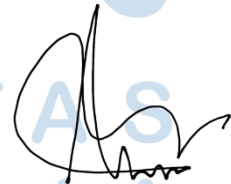
Yang bertanda tangan di bawah ini:

Nama : Gavriel Willy
NIM : 00000043681
Program Studi : Sistem Informasi
Jenjang : D3/S1/S2 (pilih salah satu)
Judul Karya Ilmiah : Komparasi Algoritma Optimasi Untuk Seleksi Fitur Klasifikasi Penyakit *Stroke* pada Algoritma Berbasis Pohon Keputusan

Menyatakan dengan sesungguhnya bahwa saya bersedia (pilih salah satu):

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial. Saya tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: Dalam proses pengajuan penerbitan ke dalam jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)*.

Tangerang, 11 Juni 2024



(Gavriel Willy)

* Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa karena berkatnya saya dapat menyelesaikan penulisan Skripsi dengan judul Komparasi Algoritma Optimasi Untuk Seleksi Fitur Klasifikasi Penyakit *Stroke* pada Algoritma Berbasis Pohon Keputusan. Pembuatan skripsi ini dilakukan untuk memenuhi syarat mencapai gelar Sarjana Jurusan Sistem Informasi. Saya ingin Mengucapkan terima kasih kepada:

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.kom., M.kom., selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Dinar Ajeng Kristiyanti, M. Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat berguna dan bermanfaat bagi pembaca dan menjadi sumber informasi dan inspirasi untuk membuat karya ilmiah yang lebih baik lagi

Tangerang, 14 Mei 2024



(Gavriel Willy)

U M M N I
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

KOMPARASI ALGORITMA OPTIMASI UNTUK SELEKSI FITUR KLASIFIKASI PENYAKIT *STROKE* PADA ALGORITMA BERBASIS POHON KEPUTUSAN

Gavriel Willy

ABSTRAK

Penyakit *stroke* merupakan gangguan neurologis akut yang terjadi ketika suplai darah ke otak terhenti. Menurut *World Stroke Organization* mencatat sekitar 13 juta orang terkena *stroke* setiap tahunnya dan 5,5 juta di antaranya meninggal. *Stroke* menjadi masalah umum di Asia termasuk Indonesia, dan menjadi penyebab utama kematian dan cacat neurologis. Selain kematian, *stroke* juga menambah beban biaya dan emosional yang mempengaruhi produktivitas. Hal ini mendorong para peneliti untuk memulai penelitian untuk mengantisipasi dan mengelola risiko *stroke*. Dalam penelitian ini, diterapkan teknik *data mining* menggunakan *CRISP-DM* yang terdiri dari tahap *business understanding*, *data understanding*, *data preparation*, *modeling*, dan hanya sampai tahap *evaluation* untuk mengklasifikasikan fitur-fitur yang meningkatkan risiko *stroke* dengan menggunakan algoritma berbasis pohon keputusan yaitu *Decision Tree*, *Xgboost*, dan *Extra tree classifier*. Seleksi fitur digunakan untuk memperoleh fitur yang relevan guna meningkatkan akurasi hasil penelitian. Pendekatan yang digunakan melibatkan algoritma *swarm intelligence* seperti *Particle Swarm Optimization*, dan *Ant Colony Optimization* yang diintegrasikan dengan algoritma berbasis pohon keputusan yaitu *Decision Tree*, *Xgboost*, dan *Extra tree classifier*. Hasil dari penelitian ini akan dikomparasikan untuk melihat perbandingan akurasi, sensitivitas, presisi, dan waktu pemrosesan yang paling efektif dan efisien setelah dan sebelum dilakukannya metode seleksi fitur. Hasil terbaik yang didapat adalah menggunakan seleksi fitur *PSO* pada *Xgboost* berhasil mereduksi fitur sampai 20 fitur terpilih dengan hasil akurasi 97.62% presisi, 97.72%, sensitivitas 97.61% dan mampu menurunkan waktu pemrosesan menjadi 1.4605 detik.

Kata kunci: *Optimization, Stroke, Swarm Intelligence*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

KOMPARASI ALGORITMA OPTIMASI SEBAGAI SELEKSI FITUR PADA ALGORITMA SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI PENYAKIT *STROKE*

Gavriel Willy

ABSTRACT (English)

Stroke disease is an acute neurological disorder that occurs when the blood supply to the brain is interrupted. According to the World Stroke Organization, approximately 13 million people suffer from stroke each year, with 5.5 million of them dying. Stroke is a common problem in Asia, including Indonesia, and is a leading cause of death and neurological disability. Besides death, stroke also adds financial and emotional burdens that affect productivity. This has prompted researchers to initiate studies to anticipate and manage the risk of stroke. In this study, data mining techniques using CRISP-DM are applied, consisting of stages of business understanding, data understanding, data preparation, modeling, and only up to the Evaluation stage to classify factors that increase the risk of stroke using decision tree-based algorithms namely Decision Tree, Xgboost, and Extra tree classifier. Feature selection is used to obtain relevant features to improve the accuracy of the research results. The approach involves swarm intelligence algorithms such as Particle Swarm Optimization, and Ant Colony Optimization integrated with decision tree-based algorithms namely Decision Tree, Xgboost, and Extra tree classifier. The results of this research will be compared to determine the most effective and efficient comparison of accuracy, sensitivity, precision, and processing time after and before feature selection methods are applied. The best result obtained is using PSO feature selection on Xgboost, which successfully reduced features to 20 selected features with an accuracy of 97.62%, precision of 97.72%, sensitivity of 97.61%, and capable of reducing processing time to 1.4605 seconds.

Keywords: Optimization, Stroke, Swarm Intelligence

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	vi
ABSTRAK	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	5
1.4 Tujuan dan Manfaat Penelitian	5
1.4.1 Tujuan Penelitian	5
1.4.2 Manfaat Penelitian	5
1.5 Sistematika Penulisan	6
BAB II LANDASAN TEORI	7
2.1 Penelitian Terdahulu	7
2.2 Tinjauan Teori	19
2.2.1 Penyakit Stroke	19
2.2.2 Seleksi Fitur	19
2.2.3 Swarm Intelligence	20
2.2.4 Shapiro-Wilk test	20
2.2.6 ANOVA	21
2.2.5 ADASYN	22
2.3 Framework dan Algoritma	22

2.3.1 CRISP-DM	22
2.3.2 Decision Tree	24
2.3.3 Xgboost.....	26
2.3.4 Extra Tree Classifier	28
2.3.5 Particle Swarm Optimization.....	30
2.3.6 Ant Colony Optimization	32
2.4 Tools.....	35
2.4.1 Python	35
2.4.2 Jupyter notebook	35
BAB III METODOLOGI PENELITIAN	36
3.1 Gambaran Umum Objek Penelitian	36
3.2 Metode Penelitian	37
3.2.1 Metode Data Mining.....	37
3.2.2 Alur Penelitian.....	40
3.3 Teknik Pengumpulan Data.....	48
3.3.1 Populasi dan Sampel.....	48
3.4 Variabel Penelitian	48
3.4.1 Variabel Dependen	48
3.4.2 Variabel Independen	49
3.5 Teknik Analisis Data	51
BAB IV ANALISIS DAN HASIL PENELITIAN	52
4.1 Business understanding	52
4.2 Data Understanding	53
4.3 Data Preparation	54
4.4 Modelling.....	62
4.4.1 Decision Tree	62
4.4.2 Optimasi PSO dalam Seleksi Fitur pada Decision Tree.....	64
4.4.3 Optimasi ACO dalam Seleksi Fitur pada Decision Tree	72
4.4.4 Extra Tree classifier.....	80
4.4.5 Optimasi PSO dalam Seleksi Fitur pada Extra tree classifier.....	82
4.4.6 Optimasi ACO dalam Seleksi Fitur pada Extra tree classifier	87
4.4.7 Xgboost	92

4.4.8 Optimasi PSO dalam Seleksi fitur pada Xgboost	94
4.4.9 Optimasi ACO dalam seleksi fitur pada Xgboost.....	99
4.5 Evaluation	103
4.5.1 Statistical Test	106
4.6 Hasil dan diskusi.....	109
4.7 Diskusi	110
BAB V SIMPULAN DAN SARAN	113
5.1 Simpulan.....	113
DAFTAR PUSTAKA	116
LAMPIRAN	124



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	7
Tabel 2. 2 Pseudocode of Decision Tree Algorithm.....	25
Tabel 2. 3 Pseudocode Xboost Algorithm	27
Tabel 2. 4 Extra-Trees splitting algorithm (for numerical attributes).....	28
Tabel 2. 5 pseudo-code of the PSO algorithm (Kennedy & Eberhart, 1995)	31
Tabel 2. 6 Pseudocode Ant Colony Optimization.....	33
Tabel 4. 1 Tabel Summary Data Preparation	61
Tabel 4. 2 Tabel perbandingan performa model klasifikasi stroke	103
Tabel 4. 3 Validasi fitur terpilih dari metode seleksi fitur	105
Tabel 4. 4 Hasil Uji Shapiro-Wilk Test dan ANOVA Test	106
Tabel 4. 5 Hasil Uji Tukey	107
Tabel 4. 6 Perbandingan hasil penelitian dengan penelitian terdahulu	110

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

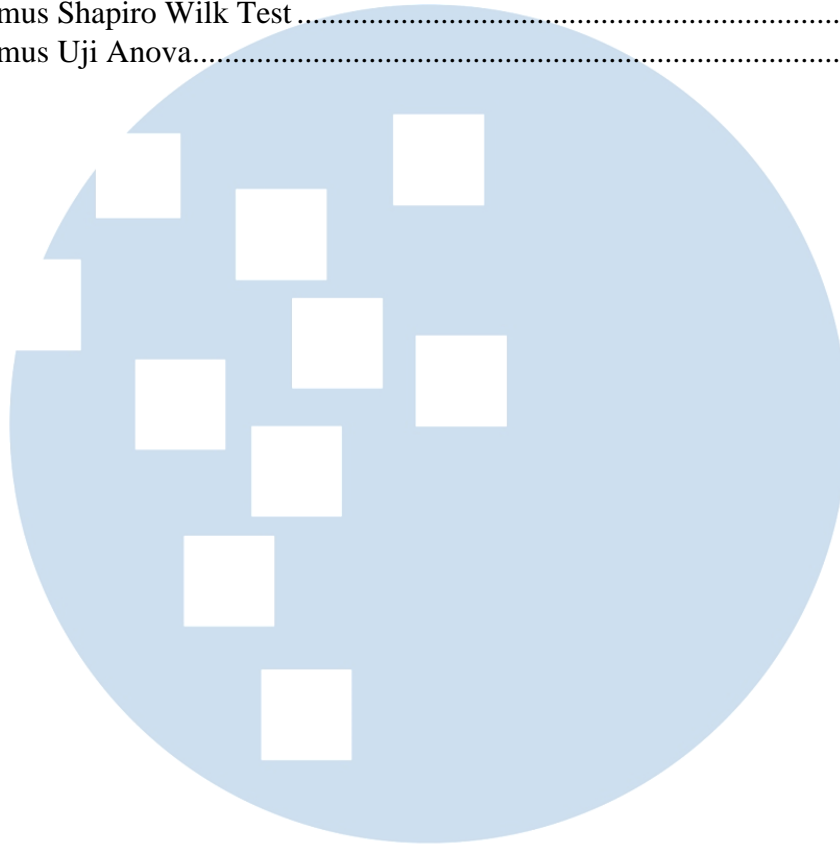
DAFTAR GAMBAR

Gambar 3. 1 Flowchart CRISP-DM.....	38
Gambar 3. 2 Alur Penelitian.....	40
Gambar 4. 1 Membaca dataset dan mengatasi nilai data yang hilang dan terduplikat	54
Gambar 4. 2 Mengisi nilai yang hilang dengan nilai acak.....	55
Gambar 4. 3 Visualisasi boxplot variabel SleepHours, HeightInMeters, WeightInKilograms, MentalHealthDays, dan PhysicalHealthDays	56
Gambar 4. 4 Proses pembuangan outlier variabel SleepHours	57
Gambar 4. 5 Visualisasi boxplot variabel SleepHours, HeightInMeters, WeightInKilograms, MentalHealthDays, dan PhysicalHealthDays setelah pembuangan outlier.....	58
Gambar 4. 6 Proses transformasi variabel kategorik menjadi numerik	59
Gambar 4. 7 Proses normalisasi menggunakan StandarScaler	59
Gambar 4. 8 Perbandingan kelas mayoritas dan minortas Hadstroke.....	60
Gambar 4. 9 Proses resample menggunakan teknik ADASYN.....	61
Gambar 4. 10 Pemisahan data menjadi 70 % untuk data training dan 30 % untuk data testing	62
Gambar 4. 11 Klasifikasi menggunakan Decision Tree.....	63
Gambar 4. 12 Kode Visualisasi Decision Tree	63
Gambar 4. 13 Visualisasi Decision Tree.....	64
Gambar 4. 14 Fungsi evaluasi model error_rate	65
Gambar 4. 15 fungsi fun dan update PSO.....	65
Gambar 4. 16 Inisialisasi fungsi bagian dari PSO	67
Gambar 4. 17 Kode fungsi jfs Seleksi fitur algoritma PSO	69
Gambar 4. 18 Melatih model Decision Tree dengan fitur terpilih.....	70
Gambar 4. 19 Hasil seleksi fitur PSO pada Decision Tree	70
Gambar 4. 20 Nama Fitur Terpilih PSO pada Decision Tree dan tingkat korelasinya	71
Gambar 4. 21 Proses pelatihan model menggunakan fitur terpilih.....	72
Gambar 4. 22 Kode fungsi fitness.....	72
Gambar 4. 23 Parameter dan Variabel Ant colony optimization	73
Gambar 4. 24 Proses iterasi algoritma Ant Colony Optimization	75
Gambar 4. 25 Penggunaan fitur terbaik dari proses seleksi fitur Ant Colony Optimization.....	77
Gambar 4. 26 Hasil akurasi dan fitur terpilih dari seleksi fitur ACO pada Decision Tree	78
Gambar 4. 27 Nama Fitur Terpilih ACO pada Decision Tree dan tingkat korelasinya	79
Gambar 4. 28 Pelatihan Decision Tree menggunakan fitur terpilih.....	80
Gambar 4. 29 Pelatihan model menggunakan Extra tree classifier Model.....	80
Gambar 4. 30 Kode visualisasi Extra Tree Classifier	81

Gambar 4. 31 Visualisasi Extra Tree	81
Gambar 4. 32 Inisialisasi fungsi dalam seleksi fitur PSO pada Extra tree classifier	82
Gambar 4. 33 Proses Seleksi fitur menggunakan PSO	84
Gambar 4. 34 Hasil akurasi dan fitur terpilih PSO pada Extra Tree Classifier	85
Gambar 4. 35 Nama Fitur Terpilih PSO pada Extra Tree dan tingkat korelasinya	86
Gambar 4. 36 Pelatihan Model Extra Tree Classifier menggunakan fitur terpilih	87
Gambar 4. 37 Fungsi fitness dan parameter seleksi fitur ACO pada Extra Tree Classifier	88
Gambar 4. 38 Proses optimasi fitur.....	89
Gambar 4. 39 Hasil seleksi fitur dan akurasi dari Extra tree classifier yang dilatih dengan best feature.....	90
Gambar 4. 40 Nama fitur terpilih ACO pada Extra Tree dan tingkat korelasinya	90
Gambar 4. 41 Pelatihan model menggunakan hasil seleksi fitur menggunakan ACO pada Extra Tree Classifier	91
Gambar 4. 42 Pelatihan model menggunakan Xgboost.....	92
Gambar 4. 43 Kode visualisasi Xgboost.....	93
Gambar 4. 44 Visualisasi Xgboost.....	93
Gambar 4. 45 Inisialisasi fungsi dan proses seleksi fitur PSO pada Xgboost.....	94
Gambar 4. 46 inisialisasi fungsi jfs dan proses iterasi seleksi fitur PSO pada Xgboost	96
Gambar 4. 47 hasil akurasi dan fitur terpilih dari seleksi fitur PSO pada Xgboost	97
Gambar 4. 48 Nama Fitur Terpilih PSO pada Xgboost dan tingkat korelasinya..	97
Gambar 4. 49 pelatihan model hasil seleksi fitur PSO pada Xgboost	98
Gambar 4. 50 Inisialisasi fungsi fitness, parameter, dan variabel ACO pada Xgboost	99
Gambar 4. 51 Proses Iterasi dan pelatihan model dari seleksi fitur menggunakan ACO pada Xgboost	100
Gambar 4. 52 Hasil seleksi fitur dan akurasi dari algoritma ACO pada Xgboost	101
Gambar 4. 53 Nama Fitur Terpilih ACO pada Xgboost dan tingkat korelasinya	101
Gambar 4. 54 pelatihan model menggunakan hasil seleksi fitur menggunakan algoritma ACO pada Xgboost.....	102

DAFTAR RUMUS

2.1 Rumus Shapiro Wilk Test.....	21
2.2 Rumus Uji Anova.....	21

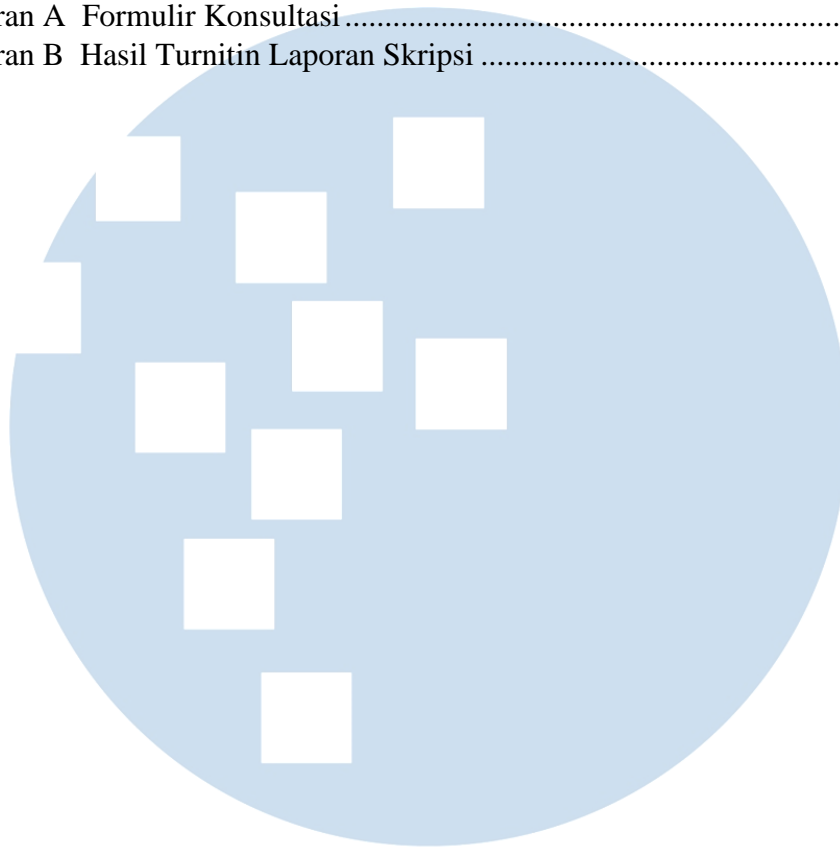


UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Formulir Konsultasi 124
Lampiran B Hasil Turnitin Laporan Skripsi 126



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA