



**PENERAPAN KLASIFIKASI *RANDOM FOREST* TERHADAP
DATA GANGGUAN SPEKTRUM AUTISME (ASD) PADA
ANAK – ANAK MENGGUNAKAN SELEKSI FITUR
*PRINCIPAL COMPONENT ANALYSIS***

SKRIPSI

LUTHFIYAH AMATULLAH

1810511008

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022



**PENERAPAN KLASIFIKASI *RANDOM FOREST* TERHADAP
DATA GANGGUAN SPEKTRUM AUTISME (ASD) PADA
ANAK – ANAK MENGGUNAKAN SELEKSI FITUR
*PRINCIPAL COMPONENT ANALYSIS***

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Komputer**

LUTHFIYAH AMATULLAH

1810511008

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Luthfiyah Amatullah

NIM : 1810511008

Tanggal : 19 Juli 2022

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 19 Juli 2022

Yang Menyatakan,



(Luthfiyah Amatullah)

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Luthfiyah Amatullah

NIM : 1810511008

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (Non-exclusive Royalty Free Right) atas skripsi saya yang berjudul:

**Penerapan Klasifikasi *Random Forest* Terhadap Data Gangguan Spektrum
Autisme (ASD) Pada Anak-Anak Menggunakan Seleksi Fitur *Principal
Component Analysis***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 19 Juli 2022

Yang menyatakan,



(Luthfiyah Amatullah)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Luthfiyah Amatullah

NIM : 1810511008

Program Studi : S1 Informatika 2018

Judul Skripsi : Penerapan Klasifikasi *Random Forest* Terhadap Data Gangguan Spektrum Autisme (ASD) Pada Anak-Anak Menggunakan Seleksi Fitur *Principal Component Analysis*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Iin Ernawati, S.Kom., M.Si.

Penguji I



Helena Nurramdhani L., S.Pd., M.Kom.

Penguji II



Yuni Widiastiwi, S.Kom., M.Si.

Pembimbing I



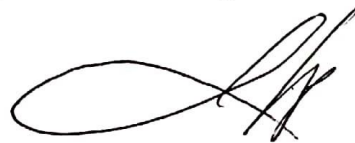
Nurul Chamidah, S.Kom., M.Kom.

Pembimbing II



Dr. Ermatita, M.Kom.

Dekan



Desta Sandya Prasvita, S.Kom., M.Kom.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Pengesahan : 19 Juli 2022



PENERAPAN KLASIFIKASI RANDOM FOREST TERHADAP DATA GANGGUAN SPEKTRUM AUTISME (ASD) PADA ANAK-ANAK MENGGUNAKAN SELEKSI FITUR *PRINCIPAL COMPONENT ANALYSIS*

Luthfiyah Amatullah

Abstrak

Autistic Spectrum Disorder (ASD) merupakan gangguan perkembangan fungsi otak yang kompleks dan sangat bervariasi. Kelainan ini secara signifikan berpengaruh terhadap komunikasi verbal, non-verbal serta interaksi sosial. Gejala umum gangguan ini biasanya akan terlihat pada anak-anak mulai sejak usia dua tahun. Tujuan penelitian ini untuk menerapkan metode seleksi fitur *Principal Component Analysis* (PCA) pada data penelitian *Autistic Spectrum Disorder Screening Data for Children Data Set* yang diperoleh dari *University of California Irvine* (UCI) *Machine Learning Data Repository* dimana PCA berfungsi untuk mereduksi atau menyederhanakan dimensi data dari dataset serta mengklasifikasikannya menggunakan pemodelan klasifikasi *Random Forest*. Selain itu, untuk mengetahui bagaimana hasil evaluasi (*confusion matrix*) serta bagaimana perbedaan yang terdapat pada hasil evaluasi tersebut terhadap dataset yang menggunakan metode seleksi fitur *Principal Component Analysis* (PCA) dengan yang tidak menggunakan metode tersebut. Evaluasi dari hasil penelitian terhadap Data Gangguan Spektrum Autisme Pada Anak - anak yang menggunakan seleksi fitur PCA yaitu menghasilkan nilai akurasi sebesar 98%, *precision* sebesar 96%, *recall* sebesar 100% dan *specificity* sebesar 96%. Sedangkan hasil evaluasi yang tanpa melalui proses PCA terlebih dahulu menghasilkan nilai akurasi sebesar 91%, *precision* sebesar 92%, *recall* sebesar 84% dan *specificity* sebesar 100%.

Kata kunci : ASD, Gangguan Spektrum Autisme, Klasifikasi, *Random Forest*, *Principal Component Analysis*.

**APPLICATION OF RANDOM FOREST CLASSIFICATION TO AUTISM
SPECTRUM DISORDER (ASD) DATA IN CHILDREN USING
PRINCIPAL COMPONENT ANALYSIS FEATURE SELECTION**

Luthfiyah Amatullah

Abstract

Autistic Spectrum Disorder (ASD) is a complex and highly variable developmental disorder of brain function. This disorder significantly affects verbal, non-verbal communication and social interaction. General symptoms of this disorder will usually be seen in children starting from the age of two years. The purpose of this research is to apply the Principal Component Analysis (PCA) feature selection method to research data Autistic Spectrum Disorder Screening Data for Children Data Set obtained from the University of California Irvine (UCI) Machine Learning Data Repository where PCA serves to reduce or simplify dimensions data from the dataset and classify it using the Random Forest classification modeling. In addition, to find out how the results of the evaluation (confusion matrix) are and how the differences in the results of the evaluation are to datasets that use the Principal Component Analysis (PCA) feature selection method and those that do not use this method. Evaluation of the results of research on Autism Spectrum Disorder Data in Children using PCA feature selection, which resulted in an accuracy value of 98%, precision of 96%, recall of 100% and specificity of 96%. While the evaluation results without going through the PCA process first resulted in an accuracy value of 91%, precision of 92%, recall of 84% and specificity of 100%.

Keywords : ASD, Autistic Spectrum Disorder, Classification, Random Forest, Principal Component Analysis.

KATA PENGANTAR

Puji syukur dan ucapan terima kasih yang tidak henti - hentinya penulis panjatkan kepada kehadiran Allah SWT yang mana telah memberikan kesehatan, kekuatan dan segala karunia-Nya kepada penulis sehingga skripsi ini berhasil diselesaikan. Tidak lupa juga penulis mengucapkan shalawat dan salam kepada junjungan Nabi Besar Muhammad SAW. Skripsi ini disusun oleh penulis guna memenuhi salah satu syarat untuk menyelesaikan pendidikan Sarjana Komputer pada Program Studi S1 Informatika di Fakultas Ilmu Komputer (FIK) Universitas *Pembangunan Nasional Veteran Jakarta*. *Judul yang dipilih pada penelitian ini adalah Penerapan Klasifikasi Random Forest Terhadap Data Gangguan Spektrum Autisme (ASD) Pada Anak-Anak Menggunakan Seleksi Fitur Principal Component Analysis*. Penulis berharap, dengan adanya skripsi ini dapat menambah referensi para pembaca. Penyelesaian tulisan ini tidak terlepas bantuan dari berbagai pihak, terutama dan teristimewa penulis ucapkan Terima kasih juga kepada:

1. Ayah dan Ibu penulis, serta seluruh keluarga yang selalu memberikan penulis dukungan, semangat dan doa hingga penulis dapat menyelesaikan skripsi ini.
2. Bu Yuni Widiastiwi, S.Kom., M.Si. dan Bu Nurul Chamidah, S.Kom., M.Kom. selaku dosen pembimbing satu dan pembimbing dua penulis yang telah memberikan saran dan arahan dalam proses penyusunan skripsi penulis.
3. Bapak/Ibu dosen Universitas Pembangunan Nasional Veteran Jakarta, terkhusus Fakultas Ilmu Komputer yang telah memberikan ilmu yang bermanfaat sehingga penulis dapat menyelesaikan skripsi ini berkat seluruh ilmu yang penulis dapatkan selama ini.
4. Teman-teman penulis, dan seluruh pihak yang telah berkontribusi baik langsung dan tidak langsung, yang tidak dapat penulis sebutkan satu - persatu dalam menyelesaikan skripsi ini.

Penulis sadar bahwa skripsi ini masih memiliki banyak kekurangan. Semua kesalahan dan kekurangan yang ada dalam skripsi ini berasal dari penulis. Penulis

harap dengan semua kekurangan tersebut, skripsi ini dapat tetap memberikan manfaat.

Jakarta, 19 Juli 2022

Penulis,

Luthfiyah Amatullah

DAFTAR ISI

PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
Abstrak	vi
Abstract	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
DAFTAR SIMBOL.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup.....	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Luaran yang Diharapkan	5
1.7 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 <i>Data Mining</i>	7
2.2 Klasifikasi	7
2.3 Gangguan Spektrum Autisme (ASD)	8
2.4 Praproses Data.....	8
2.5 Seleksi Fitur	10
2.6 Klasifikasi Random Forest.....	16
2.7 Evaluasi	18
2.8 Review Penelitian Terkait	20

BAB III METODE PENELITIAN.....	26
3.1 Tahapan Penelitian.....	26
3.2 Identifikasi Masalah.....	27
3.3 Studi Literatur	27
3.4 Akuisisi Data.....	27
3.5 Praproses Data.....	29
3.6 <i>Principal Component Analysis (PCA)</i>	31
3.7 Klasifikasi Random Forest.....	34
3.8 Evaluasi	35
3.9 Perangkat Penelitian.....	35
3.10 Jadwal Penelitian.....	36
BAB IV PEMBAHASAN.....	37
4.1 Akuisisi Data.....	37
4.2 Praproses Data.....	39
4.3 Persiapan data untuk proses PCA	45
4.4 Klasifikasi Random Forest.....	77
BAB V PENUTUP.....	85
5.1 Kesimpulan	85
5.2 Saran.....	86
DAFTAR PUSTAKA	88
RIWAYAT HIDUP.....	91
LAMPIRAN.....	92

DAFTAR TABEL

Tabel 1 Confusion Matrix	19
Tabel 2 Perbandingan Penelitian Terkait	23
Tabel 3 Keterangan fitur pada dataset	27
Tabel 4 Contoh imputasi modus.....	30
Tabel 5 Contoh menghapus fitur	30
Tabel 6 Label Encoding.....	32
Tabel 7 Jadwal Penelitian.....	36
Tabel 8 ASD-child dataset kolom id – A8_Score	37
Tabel 9 ASD-child dataset kolom A9_Score – austim.....	37
Tabel 10 ASD-child dataset kolom contry_of_res – Class/ASD	37
Tabel 11 Informasi dataframe	38
Tabel 12 Jumlah missing value masing – masing fitur	39
Tabel 13 Sebelum Imputasi	42
Tabel 14 Sesudah Imputasi.....	43
Tabel 15 Dataframe hasil praproses data	44
Tabel 16 Dataframe setelah Class/ASD dihapus.....	46
Tabel 17 Sebelum Label Encoding.....	48
Tabel 18 Sesudah Label Encoding	49
Tabel 19 Informasi dataframe siap digunakan pada proses PCA.....	49
Tabel 20 Contoh hasil standarisasi data	51
Tabel 21 Hasil matriks kovarian (Covariance Matrix).....	52
Tabel 22 Hasil vektor eigen.....	56
Tabel 23 Nilai eigen setelah diurutkan dalam urutan menurun.....	57
Tabel 24 Jumlah PC yang digunakan beserta nilai eigennya	58
Tabel 25 Matriks Proyeksi (Projection Matrix) dalam bentuk dataframe	59
Tabel 26 Hasil transformasi data (hasil akhir dari proses PCA)	61
Tabel 27 Data sampel	62
Tabel 28 Rata – rata dari masing – masing fitur (data sampel).....	63
Tabel 29 Hitung standar deviasi (data sampel)	63
Tabel 30 Hasil perhitungan standar deviasi masing – masing fitur (data sampel).....	63
Tabel 31 Hitung standarisasi data (data sampel).....	64
Tabel 32 Hasil perhitungan standarisasi data (data sampel)	65
Tabel 33 Gambaran hitung perkalian matriks kovarians (data sampel).....	65
Tabel 34 Hasil nilai rata – rata terhadap masing – masing fitur (data sampel)....	66
Tabel 35 Hasil perhitungan mencari nilai varians (data sampel)	66
Tabel 36 Hasil perhitungan mencari nilai kovarians (data sampel)	67
Tabel 37 Hasil akhir perhitungan nilai kovarians (data sampel).....	67
Tabel 38 Hasil perhitungan manual matriks kovarians (data sampel)	68
Tabel 39 Vektor eigen (eigenvectors)	74
Tabel 40 Hasil nilai eigen data sampel diurutkan secara menurun (data sampel).....	75
Tabel 41 Jumlah PC yang digunakan beserta nilai eigennya (data sampel).....	75
Tabel 42 Matriks proyeksi (data sampel)	75
Tabel 43 Hasil akhir proses PCA (data sampel).....	76
Tabel 44 Confusion Matrix dengan PCA	81

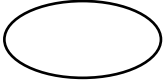
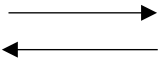
Tabel 45 <i>Confusion Matrix</i> tanpa PCA	82
Tabel 46 Perbandingan hasil evaluasi	83

DAFTAR GAMBAR

Gambar 1	Model konseptual untuk efektifitas penggunaan seleksi fitur PCA ...	12
Gambar 2	Cara Kerja Random Forest	17
Gambar 3	Tahapan Penelitian.....	26
Gambar 4	Hasil pengecekan missing value dalam bentuk subplot	40
Gambar 5	Fitur age	41
Gambar 6	Fitur ethnicity	42
Gambar 7	Hasil tangkapan layar Confusion Matrix terhadap percobaan pada dataset menggunakan seluruh fitur.....	43
Gambar 8	Distribusi variabel target / kelas ‘Class/ASD’	46
Gambar 9	Matriks kovarian dengan standarisasi data dalam bentuk heatmap....	54
Gambar 10	Matriks kovarian tanpa standarisasi data dalam bentuk heatmap.....	55
Gambar 11	Ilustrasi alur klasifikasi random forest	78

DAFTAR SIMBOL

1. Elemen – elemen pada *Flowchart* (diagram alir)

No	Simbol	Nama	Keterangan
1		<i>Terminals (Start, End)</i>	Simbol ini digunakan untuk menandakan awal dan akhir pada suatu proses.
2		<i>Processing</i>	Simbol yang menggambarkan deskripsi terhadap proses yang dijalankan.
3		<i>Flow Direction</i>	Simbol yang menghubungkan antar simbol dan menyatakan jalannya proses dalam sistem.

DAFTAR LAMPIRAN

Lampiran 1 Dataset ASD-child	93
Lampiran 2 Data hasil proses PCA	95