

**PENERAPAN ALGORITMA MACHINE LEARNING
UNTUK KLASIFIKASI TINGKAT KUALITAS UDARA DI
JAKARTA**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

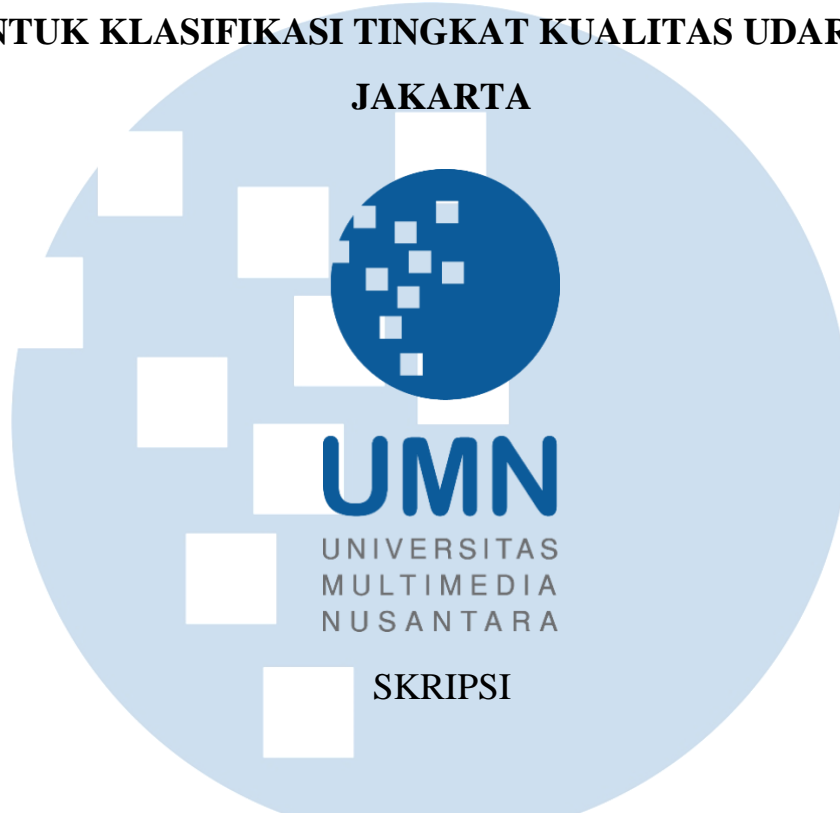
SKRIPSI

**Khadizza Khoirun Nadya
00000047877**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2024

**PENERAPAN ALGORITMA MACHINE LEARNING
UNTUK KLASIFIKASI TINGKAT KUALITAS UDARA DI
JAKARTA**



SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer (S.Kom)

Khadizza Khoirun Nadya

00000047877

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Khadizza Khoirun Nadya

Nomor Induk Mahasiswa : 00000047877

Program studi : Sistem Informasi

Skripsi dengan judul:

**“PENERAPAN ALGORITMA MACHINE LEARNING UNTUK
KLASIFIKASI TINGKAT KUALITAS UDARA DI JAKARTA”**

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, 15 Mei 2024



(Khadizza Khoirun Nadya)

HALAMAN PERSETUJUAN

Skripsi dengan judul

PENERAPAN ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI TINGKAT KUALITAS UDARA DI JAKARTA

Oleh
Nama : Khadizza Khoirun Nadya
NIM : 00000047877
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 15 Mei 2024

Pembimbing


Ir. Raymond Sunardi Oetama, M.C.I.S.
0318046803

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom
313058001

HALAMAN PENGESAHAN

Skripsi dengan judul

PENERAPAN ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI TINGKAT KUALITAS UDARA DI JAKARTA

Oleh
Nama : Khadizza Khoirun Nadya
NIM : 00000047877
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 22 Mei 2024
Pukul 08.00 s.d 10.00 dan dinyatakan
LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang


Ririn Ikana Desanti, S.Kom., M.Kom.
3130580001

Penguji


Dr. Santo Fernandi Wijaya, S.Kom., M.M.
310016902

Pembimbing


Ir. Raymond Sunardi Octama, M.C.I.S
328046803

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.
313058001

LEMBAR PERSETUJUAN PUBLIKASI

KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Khadizza Khoirun Nadya

Nomor Induk Mahasiswa : 00000047877

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah :

**“PENERAPAN ALGORITMA MACHINE LEARNING UNTUK
KLASIFIKASI TINGKAT KUALITAS UDARA DI JAKARTA”**

Menyatakan dengan sesungguhnya bahwa saya bersedia:

Memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.

Saya tidak bersedia, dikarenakan:

Dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) *.

Tangerang, 22 Mei 2024



(Khadizza Khoirun Nadya)

* Jika tidak bisa membuktikan LoA jurnal/HKI selama 6 bulan kedepan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan Skripsi ini dengan judul: "PENERAPAN ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI TINGKAT KUALITAS UDARA DI JAKARTA" dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Strata 1 Jurusan Sistem Informasi Pada Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini.

Mengucapkan terima kasih

1. Dr. Ninok Leksono, M.A., 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. Ir. Raymond Sunardi Oetama, M.C.I.S., 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 bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca

Tangerang, 15 Mei 2024


(Khadizza Khoirun Nadya)

PENERAPAN ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI TINGKAT KUALITAS UDARA DI JAKARTA

Khadizza Khoirun Nadya

ABSTRAK

Udara memiliki peran yang sangat penting dalam mendukung kehidupan di bumi. Dalam menjaga kesehatan makhluk hidup khususnya manusia, Kualitas udara memiliki peran penting untuk menjadi perhatian jangka panjang . Dampak dari polusi udara adalah timbulnya Infeksi Saluran Pernapasan Akut (ISPA), bahkan dapat berujung pada kematian pada manusia .Pemerintah Provinsi DKI Jakarta secara terus-menerus melakukan pemantauan kualitas udara di lingkungan sekitar dan telah menetapkan Indeks Standar Pencemar Udara (ISPU) sesuai dengan ketentuan yang diatur dalam Peraturan Pemerintah (PP) Nomor 41 Tahun 1999. Dalam menangani permasalahan pencemaran udara, tindakan-tindakan seperti mengurangi emisi kendaraan, meningkatkan kontrol di sektor industri, dan mengelola sampah dengan lebih efisien perlu diterapkan secara efektif. Dataset yang digunakan pada penelitian ini adalah data Indeks Standar Pencemaran Udara (ISPU) dengan rentang waktu bulan Januari 2013 hingga bulan Desember 2023. Tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi dalam klasifikasi tingkat kualitas udara di Jakarta dengan menggunakan 6 algoritma sebelum dan sesudah menggunakan optimasi, yaitu KNN, *Decision Tree*, *Naïve Bayes*, SVM, *Random Forest* dan XGBOOST, 5 algoritma hibrida tanpa dan dengan menggunakan optimasi yaitu KNN-XGBoost, *Naïve Bayes*-XGBoost, SVM-XGBoost, *Decision Tree*-XGBoost, dan *Random Forest*-XGBoost . Hasil penelitian ini bahwa model algoritma SVM yang menggunakan *hyperparameter* dan *feature selection* lebih unggul dengan nilai akurasi sebesar 100%. Dapat disimpulkan bahwa model algoritma SVM sangat baik dalam mengklasifikasi tingkat kualitas udara di Jakarta. Menurut data Indeks Standar Pencemaran Udara (ISPU), kualitas udara di Jakarta masuk kategori sedang dengan nilai ISPU 50-100. Ini menunjukkan kondisi udara masih dapat diterima, namun beberapa polutan mungkin berisiko ringan bagi kelompok sensitif.

Kata kunci: Decision Tree, Feature Selection, Hyperparameter, ISPU, K- Nearest Neighbors, Kualitas Udara, Naïve Bayes, Random Forest, Support Vector Machine, XGboost

IMPLEMENTATION OF MACHINE LEARNING ALGORITHMS FOR CLASSIFICATION AIR QUALITY LEVELS IN JAKARTA

Khadizza Khoirun Nadya

ABSTRACT (English)

Air plays a crucial role in supporting life on Earth. In maintaining the health of living beings, especially humans, air quality plays a vital role and is a long-term concern. The impact of air pollution is the onset of Acute Respiratory Infections (ARI), which can even lead to death in humans. The Jakarta Provincial Government continuously monitors air quality in the surrounding environment and has established the Air Pollution Standard Index (ISPU) by the provisions outlined in Government Regulation No. 41 of 1999. To address air pollution issues, actions such as reducing vehicle emissions, enhancing control in the industrial sector, and managing waste more efficiently need to be effectively implemented. The dataset used in this research consists of Air Pollution Standard Index (ISPU) data from January 2013 to December 2023. This study aims to determine the accuracy level in classifying air quality levels in Jakarta using six algorithms before and after optimization, namely KNN, Decision Tree, Naïve Bayes, SVM, Random Forest, and XGBOOST, as well as five hybrid algorithms without and with optimization, namely KNN-XGBoost, Naïve Bayes-XGBoost, SVM-XGBoost, Decision Tree-XGBoost, and Random Forest-XGBoost. The results of this study show that the SVM algorithm model using hyperparameters and feature selection excels with an accuracy rate of 100%. It can be concluded that the SVM algorithm model is excellent in classifying air quality levels in Jakarta. According to the Air Pollution Standard Index (ISPU) data, Jakarta's air quality is classified as moderate, with ISPU values between 50-100. This indicates that the air quality is generally acceptable, although some pollutants may pose a minor health risk to sensitive groups.

Keywords: Air Quality, Decision Tree, Feature Selection, Hyperparameter, ISPU, K- Nearest Neighbors, Machine Learning, Naïve Bayes, Random Forest, Support Vector Machine, XGboost

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
LEMBAR PERSETUJUAN PUBLIKASI	v
KARYA ILMIAH MAHASISWA	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT (English)	viii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR RUMUS	xvi
DAFTAR LAMPIRAN	xvii
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	6
1.4.1 Tujuan Penelitian	6
1.4.2 Manfaat Penelitian	6
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Penelitian Terdahulu	9
2.2 Tinjauan Pustaka	21
2.2.1 Kualitas Udara	21
2.2.2 Polusi Udara	21
2.2.3 Indeks Standar Pencemar Udara (ISPU)	21
2.3 Framework , Algoritma Klasifikasi dan Metode Evaluasi	22
2.3.1 Framework	22
2.3.1.1 Klasifikasi	22

2.3.1.2	<i>Knowledge Discovery in Database</i>	22
2.3.2	Algoritma Klasifikasi	24
2.3.2.1	<i>K-Nearest Neighbor (KNN)</i>	24
2.3.2.2	<i>Decision Tree</i>	26
2.3.2.3	<i>Naïve bayes</i>	27
2.3.2.4	<i>Support Vector Machine</i>	28
2.3.2.5	<i>Random Forest</i>	28
2.3.2.6	<i>XGBoost</i>	30
2.3.2.7	Hibrida	31
2.3.2.8	<i>Hyperparameter</i>	31
2.3.3	Metode Evaluasi	35
2.3.3.1	<i>Confusion Matrix</i>	35
2.4	Tools	37
2.4.1	<i>Jupyter Notebook</i>	37
2.4.2	<i>Python</i>	37
BAB III METODOLOGI PENELITIAN		38
3.1	Gambaran Umum Objek Penelitian	38
3.2	Alur Penelitian	38
3.2.1	Pre-KDD	39
3.2.2	<i>Data Selection</i>	39
3.2.3	<i>Data Pre-processing</i>	40
3.2.4	<i>Data Transformation</i>	40
3.2.5	Data Mining	40
3.2.6	<i>Interpretation/ Evaluation</i>	41
3.3	Metode Penelitian	41
3.4	Teknik Pengumpulan Data	44
3.4.1	Data Collection	44
3.5	Teknik Analisis Data	44
BAB IV ANALISIS DAN HASIL PENELITIAN		48
4.1	Pre-KDD	48
4.1.1	<i>Exploratory Data Analysis (EDA)</i>	50
4.1	<i>Data Selection</i>	51

4.2	Data Preprocessing	51
4.3	Transformation	52
4.4	Data Mining	54
4.5.1	Algoritma KNN	54
4.5.2	Algoritma Naïve Bayes	57
4.5.3	Algoritma SVM	59
4.5.4	Algoritma Decision Tree	60
4.5.5	Algoritma Random Forest	62
4.5.6	Algoritma XGBoost	63
4.5.7	Hibrida KNN-XGBoost	65
4.5.7.1	Tanpa Menggunakan Optimasi	65
4.5.7.2	Menggunakan Optimasi	65
4.5.8	Hibrida Naïve Bayes-XGBoost	66
4.5.8.1	Tanpa Menggunakan Optimasi	66
4.5.8.2	Menggunakan Optimasi	67
4.5.9	Hibrida SVM-XGBoost	67
4.5.9.1	Tanpa Menggunakan Optimasi	67
4.5.9.2	Menggunakan Optimasi	68
4.5.10	Hibrida Decision Tree-XGBoost	69
4.5.10.1	Tanpa Menggunakan Optimasi	69
4.5.10.2	Menggunakan Optimasi	70
4.5.11	Hibrida Random Forest-XGBoost	70
4.5.11.1	Tanpa Menggunakan Optimasi	70
4.5.11.2	Menggunakan Optimasi	71
4.5	Interpreation/Evaluation	71
4.6.1	Visualisasi Perbandingan Akurasi	72
4.6.1.1	Tanpa Menggunakan Optimasi	72
4.6.1.2	Menggunakan Optimasi	74
4.6.2	Visualisasi Confusion Matrix	77
4.6.2.1	Tanpa Menggunakan Optimasi	77
A.	KNN	77
B.	Naïve Bayes	78

C. SVM.....	78
D. Decision Tree	79
E. Random Forest	80
F. XGBoost	81
G. KNN – XGBoost	82
4.6.2.1 Menggunakan Optimasi.....	86
4.6 Hasil dan Diskusi.....	94
BAB V SIMPULAN DAN SARAN.....	98
5.1 Simpulan.....	98
5.2 Saran.....	98
DAFTAR PUSTAKA.....	101
LAMPIRAN.....	110

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	9
Tabel 3. 1 Variabel Dataset ISPU DKI Jakarta.....	39
Tabel 3. 2 Perbandingan 6 Algoritma	42
Tabel 3. 3 Perbandingan antara KDD, CRISP-DM,dan SEMMA.....	44
Tabel 3. 4 Perbandingan Python dan R.....	45
Tabel 3. 5 Perbandingan Jupyter Notebook dan R Studio	46
Tabel 4. 1 Perbandingan Akurasi Seluruh Model	94
Tabel 4. 2 Feature Selection Berdasarkan Algoritma	95

A large, light blue watermark logo of Universitas Multimedia Nusantara (UMMN) is centered on the page. It features a stylized globe with a grid of white squares and the letters 'UMMN' in a bold, rounded font.

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 2. 1 Tahapan KDD	22
Gambar 2. 2 Prosedur K-Nearest Neighbor	24
Gambar 2. 3 Perhitungan Confusion Matrix	25
Gambar 3. 1 Alur Penelitian KDD	38
Gambar 3. 2 Data Mentah ISPU	44
Gambar 4. 1 Data Mentah	48
Gambar 4. 2 Package Import	49
Gambar 4. 3 Dataset Describe	49
Gambar 4. 4 Grafik Distribusi Kategori Kualitas Udara di Jakarta	50
Gambar 4. 5 Dataset Import	51
Gambar 4. 6 Data Null	52
Gambar 4. 7 Hasil Penghapusan Data	52
Gambar 4. 8 Dictionary Mapping String dan Integer	53
Gambar 4. 9 Split Data	53
Gambar 4. 10 Algoritma KNN tanpa Optimasi	54
Gambar 4. 11 Output KNN	55
Gambar 4. 12 Algoritma KNN Optimasi	56
Gambar 4. 13 Output KNN Optimasi	57
Gambar 4. 14 Output Naive Bayes	58
Gambar 4. 15 Output Naive Bayes Optimasi	58
Gambar 4. 16 Output SVM	59
Gambar 4. 17 Output SVM Optimasi	60
Gambar 4. 18 Output Decision Tree	61
Gambar 4. 19 Output Decision Tree	61
Gambar 4. 20 Output Random Forest	62
Gambar 4. 21 Output Random Forest Optimasi	63
Gambar 4. 22 Output XGBoost	64
Gambar 4. 23 Output XGBoost Optimasi	64
Gambar 4. 24 Output Hibrida KNN-XGBoost	65
Gambar 4. 25 Output Hibrida KNN-XGBoost Optimasi	66
Gambar 4. 26 Output Hibrida Naive Bayes-XGBoost	66
Gambar 4. 27 Output Hibrida Naive Bayes-XGBoost Optimasi	67
Gambar 4. 28 Output Hibrida SVM-XGBoost	68
Gambar 4. 29 Output Hibrida SVM-XGBoost Optimasi	68
Gambar 4. 30 Output Hibrida Decision Tree-XGBoost	69
Gambar 4. 31 Output Hibrida Decision Tree-XGBoost Optimasi	70
Gambar 4. 32 Output Hibrida Random Forest-XGBoost	70
Gambar 4. 33 Output Hibrida Random Forest-XGBoost Optimasi	71
Gambar 4. 34 Script Code Visualisasi Tanpa Optimasi	72

Gambar 4. 35 Hasil Visualisasi Pemodelan Tanpa Optimasi	73
Gambar 4. 36 Hasil Visualisasi Pemodelan Hibrida Tanpa Optimasi	74
Gambar 4. 37 Hasil Script Code Visualisasi Pemodelan Optimasi	75
Gambar 4. 38 Hasil Visualisasi Pemodelan Optimasi	75
Gambar 4. 39 Hasil Visualisasi Pemodelan Hibrida Optimasi	76
Gambar 4. 40 Confusion Matrix KNN.....	77
Gambar 4. 41 Confusion Matrix Naive Bayes	78
Gambar 4. 42 Confusion Matrix SVM.....	78
Gambar 4. 43 Confusion Matrix Decision Tree.....	79
Gambar 4. 44 Confusion Matrix Random Forest.....	80
Gambar 4. 45 Confusion Matrix XGBoost	81
Gambar 4. 46 Confusion Matrix KNN-XGBoost	82
Gambar 4. 47 Confusion Matrix Naive Bayes-XGBoost.....	83
Gambar 4. 48 Confusion Matrix SVM-XGBoost	84
Gambar 4. 49 Confusion Matrix Decision Tree-XGBoost	85
Gambar 4. 50 Confusion Matrix Random Forest-XGBoost	85
Gambar 4. 51 Confusion Matrix KNN Optimasi	86
Gambar 4. 52 Confusion Matrix Naive Bayes Optimasi	87
Gambar 4. 53 Confusion Matrix Naive Bayes Optimasi	87
Gambar 4. 54 Confusion Matrix Naive Bayes Optimasi	88
Gambar 4. 55 Confusion Matrix Random Forest Optimasi	89
Gambar 4. 56 Confusion Matrix XGBoost Optimasi	89
Gambar 4. 57 Confusion Matrix KNN-XGBoost Optimasi	90
Gambar 4. 58 Confusion Matrix Naive Bayes-XGBoost Optimasi.....	91
Gambar 4. 59 Confusion Matrix SVM-XGBoost Optimasi	92
Gambar 4. 60 Confusion Matrix Decision Tree-XGBoost Optimasi.....	93
Gambar 4. 61 Confusion Random Forest-XGBoost Optimasi.....	94

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

DAFTAR RUMUS

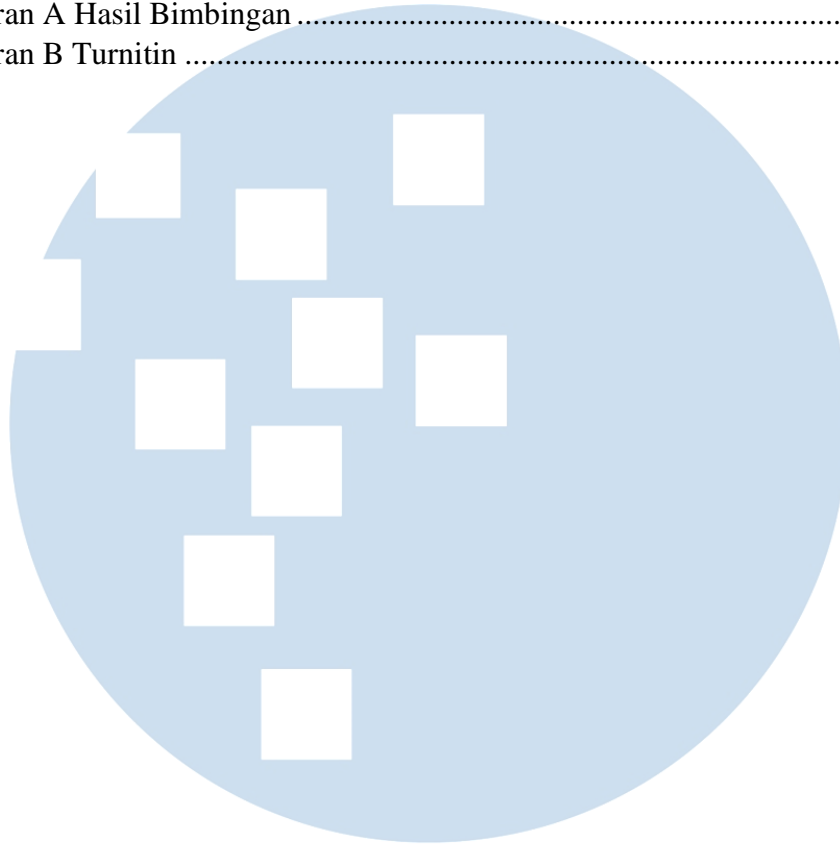
Rumus 2. 1 Perhitungan normalisasi min-max	25
Rumus 2. 2 Rumus Euclidean	25
Rumus 2. 3 Rumus Information Gain	26
Rumus 2. 4 Perhitungan Entropy	27
Rumus 2. 5 Rumus Teorema Bayes	27
Rumus 2. 6 Rumus SVM	28
Rumus 2. 7 Rumus Enthropy Random Forest.....	29
Rumus 2. 8 Rumus Gain Random Forest.....	29
Rumus 2. 9 Rumus Persamaan Fungsi Objektif	30
Rumus 2. 10 Regularisasi.....	30
Rumus 2. 11 Rumus Akurasi	36
Rumus 2. 12 Rumus Presisi	36
Rumus 2. 13 Rumus Recall.....	36
Rumus 2. 14 Rumus Recall.....	36

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Hasil Bimbingan 110
Lampiran B Turnitin 112



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA