

**OPTIMALISASI ALGORITMA *K-NEAREST NEIGHBORS*,
RANDOM FOREST DAN *SUPPORT VECTOR MACHINE*
DENGAN EKPLORASI MODEL *ENSEMBLE VOTING* DAN
STACKING UNTUK PREDIKSI KUALITAS UDARA**

**Tesis
untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-2 Program Studi
Magister Sistem Informasi**



**Aziz Jihadian Barid
30000322410014**

**SEKOLAH PASCASRAJANA
UNIVERSITAS DIPONEGORO
SEMARANG
2024**

HALAMAN PENGESAHAN

TESIS

OPTIMALISASI ALGORITMA *K-NEAREST NEIGHBORS*, *RANDOM FOREST* DAN *SUPPORT VECTOR MACHINE* DENGAN EKPLORASI MODEL *ENSEMBLE VOTING* DAN *STACKING* UNTUK PREDIKSI KUALITAS UDARA

Oleh:

Aziz Jihadian Barid
30000322410014

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 26 Maret 2024 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 26 Maret 2024

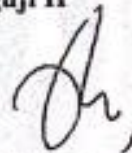
Mengetahui,

Penguji II

Penguji I

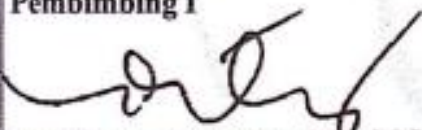


**Prof. Dr. Ir. R. Rizal Isnanto, S.T.,
M.M., M.T., IPU, ASEAN Eng.**
NIP 197007272000121001



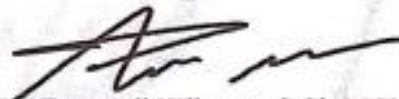
Jatmiko Endro Suseno, S.Si, M.Si, Ph.D
NIP 197211211998021001

Pembimbing I



Prof. Dr. Ir. Hadiyanto, S.T., M.Sc
NIP: 197510281999031004

Pembimbing II



Dr. Eng. Adi Wibowo, S.Si, M.Kom
NIP. 198203092006041002

Mengetahui:

**Dekan Sekolah Pascasarjana
Universitas Diponegoro**



Dr. R. B. Sularto, S.H., M.Hum.
NIP 196704011991031005

**Ketua Program Studi
Magister Sistem Informasi**



Drs. Bayu Surarso, M.Sc., Ph.D.
NIP 196311051988031001

PERNYATAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 01 Maret 2024



METERAI
TEMPEL
SNETGALX124356178

Aziz Jihadian Barid

SEKOLAH PASASARIAN

**NYATAAN PERSETUJUAN
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini :

Nama	:	Aziz Jihadian Barid
NIM	:	30000322410014
Program Studi	:	Magister Sistem Informasi
Program	:	Sekolah Pascasarjana
Jenis Karya	:	Tesis

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul :

**OPTIMALISASI ALGORITMA *K-NEAREST NEIGHBORS*,
RANDOMFOREST DAN *SUPPORT VECTOR MACHINE*
UNTUK PREDIKSI KUALITAS UDARA**

beserta perangkat yang ada. Dengan Hak bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*) merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang

Pada tanggal : 01 Maret 2024

Yang menyatakan



Aziz Jihadian Barid

NIM 30000322410014

KATA PENGANTAR

Segala Puji dan syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa yang telah melimpahkan segala rahmatNya sehingga penulis dapat menyelesaikan tesis dengan judul “Optimalisasi Algoritma *K-Nearest Neighbors*, *Random Forest* Dan *Support Vector Machine* Dengan Ekplorasi Model *Ensemble Voting* Dan *Stacking* Untuk Prediksi Kualitas Udara” guna memenuhi sebagian persyaratan untuk memperoleh gelar magister pada program studi Magister Sistem Informasi, Sekolah Pascasarjana Universitas Diponegoro.

Penulis menyadari kelemahan serta keterbatasan yang ada sehingga dalam menyelesaikan tesis ini memperoleh bantuan dari berbagai pihak, dalam kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. Bapak Dr. R.B. Sularto, S.H., M.Hum. selaku dekan Sekolah Pascasarjana Universitas Diponegoro.
2. Bapak Drs. Bayu Surarso, M.Sc., Ph.D. selaku Ketua Program Studi Magister Sistem Informasi Universitas Diponegoro.
3. Bapak Prof. Dr. Ir. Hadiyanto, S.T., M.Sc. selaku Dosen Pembimbing I, yang telah memberikan bimbingan, arahan, serta kritik kepada penulis hingga bisa selesai dalam penyusunan tesis.
4. Bapak Dr. Eng. Adi Wibowo, S.Si, M.Kom. selaku dosen Pembimbing II, yang sudah membimbing, memberikan arahan, dan masukan yang baik kepada penulis sehingga dapat penulis menyelesaikan laporan tesis ini.
5. Bapak Prof. Dr. Ir. R. Rizal Isnanto, S.T., M.M., M.T., IPU, ASEAN Eng selaku Dosen Ketua Penguji yang telah memberikan banyak saran dan masukan yang baik kepada penulis.
6. Bapak Jatmiko Endro Suseno, S.Si, M.Si, Ph.D. sebagai Dosen Penguji yang telah memberikan banyak saran dan masukan dalam penyusunan tesis.
7. Semua pihak yang tidak dapat disebutkan satu persatu, yang telah membantu sampai dengan terselesaikannya tesis ini.

Penulis menyadari bahwa tesis ini masih banyak kekurangan baik isi maupun susunannya. Semoga tesis ini dapat bermanfaat tidak hanya bagi penulis juga bagi para pembaca.

Semarang, 01 Maret 2024

Aziz Jihadian Barid

DAFTAR ISI

Halaman Judul	i
Halaman Pengesahan	Error! Bookmark not defined.
Halaman Pernyataan Persetujuan Publikasi	iii
Halaman Pernyataan	iv
Kata Pengantar	v
Daftar Isi	vi
Daftar Gambar	viii
Daftar Tabel	ix
Abstrak	xi
Abstract	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian	4
1.3 Manfaat penelitian	4
BAB II LANDASAN TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori	9
2.2.1 <i>Machine learning</i>	9
2.2.2 <i>Random forest</i>	10
2.2.3 <i>K-Nearest Neighbors (KNN)</i>	11
2.2.4 <i>Support Vector Machine (SVM)</i>	12
2.2.5 Parameter	13
2.2.6 <i>Hyperparameter</i>	13
2.2.7 <i>Hyperparameter Tuning</i>	14
2.2.8 <i>Cross validation</i>	15
2.2.9 <i>Matrix MSE & RMSE</i>	16
2.2.10 <i>Ensemble learning</i>	17
2.2.10.1 <i>Ensemble Voting</i>	17
2.2.10.2 <i>Ensemble Stacking</i>	18
2.2.11 <i>Data mining</i>	19
2.2.12 <i>Prediksi</i>	19
2.2.13 <i>Kualitas Udara</i>	19
BAB III METODE PENELITIAN	20
3.1 Bahan dan Alat Penelitian	20
3.2 Prosedur penelitian	21
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	26
4.1.1 Memasukan Library	26
4.1.2 Memasukan Dataset	27

4.1.3	<i>Preprocessing</i> data	28
4.1.4	Supervised Learning	28
4.1.5	Pembagian Data	30
4.1.6	Korelasi Antar <i>Variable</i>	30
4.1.7	<i>Future Importance</i>	31
4.1.8	Visualisasi Data	31
4.2	Evaluasi Model	32
4.2.1	Pengujian Model K-Nearest Neighbor	32
4.2.2	Pengujian Model <i>Support Vector Machine</i>	34
4.2.3	Pengujian Model <i>Random forest</i>	36
4.2.4	Pengujian Model Ensemble Voting	38
4.2.5	Pengujian Model <i>Ensemble Stacking</i>	39
4.3	Komparasi Evaluasi Hasil	41
BAB V KESIMPULAN DAN SARAN		44
5.1	Kesimpulan	44
5.2	Saran	44



SEKOLAH PASCASARJANA

DAFTAR GAMBAR

Gambar 2.1 <i>Random forest</i> (Khan , dkk, 2021)	10
Gambar 2.2 <i>K-Nearest Neighbor</i> (sumber : https://www.javatpoint.com)	11
Gambar 2.3 <i>Support Vector Machine</i> (Sumber : Javapoint.com)	13
Gambar 2.4 <i>K-Fold Cross-Validation</i> (Maredia, n.d.).....	16
Gambar 2.5 <i>Ensemble learning Voting</i>	18
Gambar 2.6 <i>Ensemble learning Stacking</i>	19
Gambar 3. 1 Prosedur penelitian	21
Gambar 3. 2 Dataset Kualitas Udara.....	23
Gambar 4.1 Memasukkan Library pada Python	26
Gambar 4.2 Memasukan dataset pada python.....	27
Gambar 4.3 menampilkan dataset pada python	27
Gambar 4.4 Statistik Sumary	28
Gambar 4.5 Pembagian data	30
Gambar 4.6 Korelasi Antar Variable	30
Gambar 4.7 Future Importance	31
Gambar 4.8 Visualisai rata-rata harian.....	32
Gambar 4.9 Visualisasi prediksi menggunakan KNN <i>hyperparameter</i>	34
Gambar 4.10 Visualisasi prediksi menggunakan SVM <i>hyperparameter</i>	36
Gambar 4.11 Visualisasi prediksi menggunakan RF <i>hyperparameter</i>	37
Gambar 4.12 Visualisasi prediksi <i>ensemble voting hyperparameter</i>	39
Gambar 4.13 Visualisasi prediksi <i>ensemble stacking hyperparameter</i>	41

SEKOLAH PASCASARJANA

DAFTAR TABEL

Tabel 3. 1 Parameter Kualitas Udara.....	20
Tabel 3. 2 Indeks Standar Pencemaran Udara	20
Tabel 4. 1 Dataset polutan (PM10)	29
Tabel 4. 2 Dataset.....	29
Tabel 4. 3 Hasil Evaluasi KNN.....	33
Tabel 4. 4 Hasil Evaluasi SVM.....	34
Tabel 4. 5 Hasil Evaluasi <i>Random forest</i>	36
Tabel 4. 6 Hasil Evaluasi Ensemble Voting.....	38
Tabel 4. 7 Hasil Evaluasi Ensemble Stacking.....	40
Tabel 4. 8 Komparasi evaluasi model menggunakan MSE	42
Tabel 4. 9 Komparasi evaluasi model menggunakan RMSE.....	42



SEKOLAH PASCASARJANA

DAFTAR ARTI LAMBANG DAN SINGKATAN

DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
S	Himpunan Kasus
n	Sample Data
p_i	Proporsi S_i terhadap S
A	Atribut
$ S_1 $	jumlah kasus pada partisi ke-1
X_i	Sampel data atau Data train
Y_i	Data uji atau data testingi
i	Variable Data
d	jarak

DAFTAR SINGKATAN

Singkatan	Kepanjangan Singkatan
<i>KNN</i>	<i>K-Nearest Neighbor</i>
<i>SVM</i>	<i>Support Vector Machine</i>
<i>RF</i>	<i>Random forest</i>
<i>MSE</i>	<i>Mean Squared Error</i>
<i>RMSE</i>	<i>Root Mean Squared Error</i>
<i>DL</i>	<i>Deep Learning</i>

SEKOLAH PASCASARJANA

OPTIMALISASI ALGORITMA *K-NEAREST NEIGHBORS*, *RANDOM FOREST* DAN *SUPPORT VECTOR MACHINE* DENGAN EKPLORASI MODEL *ENSEMBLE VOTING* DAN *STACKING* UNTUK PREDIKSI KUALITAS UDARA

ABSTRAK

Penelitian ini didasari karena perkembangan ekonomi, industrialisasi, dan urbanisasi yang pesat di Indonesia telah menyebabkan tingkat polusi dengan dampak negatif terhadap lingkungan dan kesehatan masyarakat. Tujuan penelitian ini untuk memprediksi kualitas udara menggunakan *machine learning* dengan algoritma *K Nearest Neighbor*, *Support Vector Machine* dan *Random forest*, dari model tersebut digabungkan menjadi *ensemble learning*. Selanjutnya teknik *hyperparameter* diterapkan kedalam proses prediksi untuk meningkatkan hasil. Penelitian ini menggunakan validasi *K-fold cross* untuk mengurangi *overfitting*. Hasil evaluasi menunjukkan bahwa model dengan menggunakan ensemble memiliki kinerja yang lebih baik dibanding dengan single algoritma yang ditunjukkan dengan nilai error yang lebih kecil. Hasil penelitian menunjukkan bahwa penerapan *hyperparameter* membawa peningkatan kinerja model, khusus nya pada algoritma SVM karena dapat menangani fitur yang relevan dalam data secara runtun waktu. Wawasan ini memberikan arahan bagi sistem pemantauan kualitas udara yang efektif dan pengambilan keputusan yang tepat dalam pengelolaan polusi udara.

Kata kunci : prediksi, *K Nearest Neighbor*, *Support Vector Machine*, *Random forest*, *Hyperparameter*

SEMARANG
SEKOLAH PASCASARJANA

**OPTIMIZATION OF *THE K-NEAREST NEIGHBORS* ALGORITHM,
RANDOM FOREST AND *SUPPORT VECTOR MACHINE* WITH
ENSEMBLE VOTING AND STACKING MODEL EXPLORATION FOR
AIR QUALITY PREDICTION**

ABSTRACT

This research is based on the fact that rapid economic development, industrialization and urbanization in Indonesia have caused levels of pollution with negative impacts on the environment and public health. The aim of this research is to predict air quality using machine learning with the K Nearest Neighbor algorithm, Support Vector Machine and *Random forest*, these models are combined into *ensemble learning*. Next, hyperparameter techniques are applied to the prediction process to improve results. This research uses K-fold cross validation to reduce overfitting. The evaluation results show that the model using the ensemble has better performance than the single algorithm as indicated by a smaller error value. The research results show that the application of hyperparameters brings increased model performance, especially the SVM algorithm because it can handle relevant features in time series data. These insights provide direction for effective air quality monitoring systems and informed decision making in air pollution management.

keywords : prediksi, K Nearest Neighbor, Support Vector Machine, *Random forest*, Hyperparameter



SEMARANG

SEKOLAH PASCASARJANA