

**PREDIKSI KUALITAS UDARA MENGGUNAKAN METODE  
ENSEMBLE PADA MODEL *DECISION TREE* ID.3, *RANDOM FOREST*  
DAN *REGRESI LOGISTIK MULTINOMIAL***

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh gelar Sarjana di  
Jurusan Matematika pada Fakultas MIPA**

**Oleh :**

**MAU'IZATIL RAHMAYANI**

**NIM.08011281823038**



**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS SRIWIJAYA**

**2022**

**HALAMAN PENGESAHAN**

**PREDIKSI KUALITAS UDARA MENGGUNAKAN METODE ENSEMBLE  
PADA MODEL *DECISION TREE* ID.3, *RANDOM FOREST* DAN *REGRESI  
LOGISTIK MULTINOMIAL***

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh gelar Sarjana  
di Jurusan Matematika pada Fakultas MIPA**

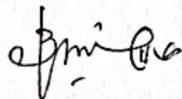
**Oleh :**

**MAU'IZATIL RAHMAYANI**

**08011281823038**

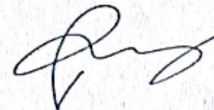
**Indralaya, Juli 2022**

**Pembimbing Pembantu**



**Dra. Ning Elivati, M.Pd**  
**NIP. 195911201991022001**

**Pembimbing Utama**



**Dr. Yulia Resti, M.Si**  
**NIP.197307191997023001**

**Mengetahui**

**Ketua Jurusan Matematika**



**Drs. Sugandi Yahdin, M.M**

**NIP.195807271986031003**



## PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Mau'izatil Rahmayani

NIM : 08011281823038

Jurusan : Matematika

Menyatakan dengan ini saya bersungguh-sungguh bahwa skripsi yang berjudul "Prediksi Kualitas Udara Menggunakan Metode Ensemble Pada Model Decision Tree ID.3, Random Forest Dan Regresi Logistik Multinomial" merupakan karya yang saya susun sendiri dan saya tidak melakukan penjiplakan dari karya manapun serta saya melakukan pengutipan sesuai dengan pedoman keilmuan yang berlaku seperti tertuang dalam Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 17 Tahun 2010 terkait Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi.

Apabila dikemudian hari, terdapat pelanggaran yang ditemukan dalam skripsi saya ataupun adanya pengaduan dari pihak lain terhadap keaslian skripsi saya, maka saya bersedia menanggung sanksi yang dijatuhkan kepada saya.

Demikianlah pernyataan ini dibuat dengan sungguh-sungguh tanpa adanya paksaan dari pihak manapun.

Indralaya, Juli 2022

Yang membuat pernyataan,



Mau'izatil Rahmayani

NIM. 08011281823038

## **HALAMAN PERSEMBAHAN**

### **MOTTO**

"Apapun yang menjadi takdirmu, akan mencari jalannya menemukanmu"

**(Abi bin Abi Thalib)**

"Tidak ada kesuksesan tanpa kerja keras. Tidak ada keberhasilan tanpa kebersamaan. Tidak ada kemudahan tanpa doa."

**(Ridwan Kamil)**

“Jadilah kuat tapi tidak menyakiti. Jadilah baik, tapi tidak lemah.  
Jadilah berani, tapi tidak menakuti, Jadilah rendah hati, tapi tidak rendah.  
Tetap bangga, tapi tidak sombong”

**Skripsi ini kupersembahkan**

**kepada :**

- 1. Allah SWT**
- 2. Bapak dan Ibu**
- 3. Saudaraku**
- 4. Keluarga Besar**
- 5. Dosen**
- 6. Almamater**
- 7. Sahabat dan Temanku**

## KATA PENGANTAR

*Assalamu'alaikum Warahmatullahi Wabarakatuh.*

Puji syukur atas kehadiran Tuhan Yang Maha Esa atas segala limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “**Prediksi Kualitas Udara Menggunakan Metode Ensemble Pada Model Decision Tree ID.3, Random Forest Dan Regresi Logistik Multinomial**”. Shalawat beserta salam semoga selalu tercurahkan kepada kita Nabi Muhammad SAW beserta keluarga, sahabat dan para pengikutnya hingga akhir zaman.

Penulis sadar dalam proses penyusunan skripsi ini masih banyak kekurangan, serta banyaknya rintangan dalam pengerjaannya. Namun dengan kesabaran dan ketekunan yang dilandasi rasa tanggung jawab sehingga penulis dapat menyelesaikan skripsi ini.

Dengan penuh rasa hormat penulis juga mengucapkan banyak terimakasih yang sebesar-besarnya kepada orang tua tercinta, yaitu **Bapak Herawadi** dan **Ibu Dismawati** yang tidak pernah lupa mendoakan, merawat dan memberikan kasih sayang, semangat serta restunya kepada penulis. Penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak oleh karena itu penulis ingin mengucapkan terima kasih kepada:

1. Bapak **Prof. Hermansyah, S.Si., M.Si., Ph.D.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
2. Bapak **Drs. Sugandi Yahdin, M.M** selaku Ketua Jurusan Matematika Fakultas MIPA Universitas Sriwijaya atas ilmu yang telah diberikan

3. Ibu **Dr. Dian Cahyawati Sukanda, M.Si** selaku Sekretaris Jurusan yang telah membantu dalam proses administrasi pendaftaran seminar, serta ilmu yang diberikan.
4. Ibu **Des Alwine Zayanti, M.Si.** selaku Dosen Pembimbing Akademik yang telah memberikan bimbingan, pengarahan, dan saran selama masa perkuliahan di Jurusan Matematika FMIPA Universitas Sriwijaya.
5. Ibu **Dr. Yulia Resti, M.Si** selaku Dosen Pembimbing Utama yang telah memberikan masukan, saran dan membimbing serta meluangkan waktu kepada penulis sehingga penulis dapat menyelesaikan skripsi dengan baik.
6. Ibu **Dra. Ning Eliyati, M.Pd.** selaku Dosen Pembimbing Pembantu yang telah memberikan masukan, saran dan membimbing serta meluangkan waktu kepada penulis sehingga penulis dapat menyelesaikan skripsi dengan baik.
7. Ibu **Endang Sri Kresnawati, M.Si** selaku Dosen Pembahas Pertama yang telah memberikan tanggapan, masukan dan saran dalam penyelesaian skripsi ini.
8. Bapak **Drs. Endro Setyo C, M.Si** selaku Dosen Pembahas Kedua yang telah memberikan tanggapan, masukan dan saran dalam penyelesaian skripsi ini.
9. Ibu **Anita Desiani M.Kom** selaku Ketua Seminar skripsi yang telah memberikan arahan, saran, membimbing penulis serta mengatur jalannya seminar sehingga dapat berjalan dengan baik.
10. Bapak **Dr. Bambang Suprihatin M.Si** selaku Sekretaris Seminar skripsi yang telah membantu dan memberikan catatan masukan yang sangat berguna bagi penulis.

11. **Seluruh Dosen** di Jurusan Matematika FMIPA Universitas Sriwijaya yang telah memberikan ilmu yang bermanfaat kepada penulis selama menempuh pendidikan.
12. Bapak **Irwansyah** dan Ibu **Hamidah** yang telah banyak membantu dalam proses administrasi.
13. Saudara-saudaraku **Ana, Ami** dan **Shoib** yang selalu memberikan perhatian, kasih sayang, dukungan kepada penulis sampai saat ini.
14. Sahabat sejak SMA **Novita Yulinda** yang selalu mendukung dan memberi semangat dalam suka maupun duka serta selalu ada dan mengerti setiap keadaan penulis.
15. Teman seperjuangan di bangku kuliah **Nadya, Nabel, Indi, Imam, Kare, Cenny, Dwi, Udin, Adinda, Dineks, gita** dan seluruh teman-teman angkatan **2018** yang selalu menemani keseharian penulis dengan penuh kebahagiaan.
16. Teman seperjuangan dalam perskripsian **Hasma, Rara, Anggi, Desi** dan **Tasya** yang telah kebersamai, membantu dan memberikan semangat dalam penyelesaian tugas akhir ini.
17. Teman yang selalu mensupport penulis dalam penyelesaian tugas akhir ini yaitu **Virgo**.
18. Kakak tingkat Angkatan **2016** dan **2017** yang telah memberikan ilmu dan pengalamannya serta adik tingkat Angkatan **2019** dan **2020** yang telah memberikan dukungan dan doa serta semangat.
19. Semua Pihak yang tidak dapat penulis sebutkan satu persatu. Semoga segala kebaikan yang diberikan mendapat balasan dari Allah SWT.

Akhir kata semoga skripsi ini dapat menambah pengetahuan dan bermanfaat bagi mahasiswa/mahasiswi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya dan seluruh pihak yang membutuhkan.  
*Wassalamu'alaikum Warahmatullahi Wabarakatuh.*

**Indralaya, Juli 2022**

**Penulis**



**PREDICTION OF AIR QUALITY USING ENSEMBLE METHOD ON  
DECISION TREE ID.3 MODEL, RANDOM FOREST AND  
MULTINOMIAL LOGISTIC REGRESSION**

**By:**

**Mau'izatil Rahmayani**

**08011281823038**

**ABSTRACT**

Air quality is one of the important components for living things on the earth's surface, especially for humans. Air quality must be maintained and maintained so as not to experience a decrease in quality caused by weather factors. Air quality is one of the causes of human health problems. Therefore, the purpose of this study is to predict air quality using the Ensemble method using three classification models. The use of the ensemble method aims to minimize errors in classification and get a better level of accuracy. The data used in this study has 21 variables with a total of 2502 data. The classification uses the Ensemble Majority vote method based on three algorithm models Decision Tree, Random Forest and Multinomial Logistics Regression. The results of this study indicate that the level of accuracy of air quality prediction using the Ensemble Majority Vote method obtained an accuracy value of 99.31%, macro precision of 78.45%, and macro recall of 78.63%, macro fscore of 78.54%, for precision, recall and micro fscore have the same value, namely 98.28%.

**Keywords:** Air quality, Ensemble Majority Vote, Decision Tree, Random Forest  
Multinomial Logistics Regression.

**PREDIKSI KUALITAS UDARA MENGGUNAKAN METODE  
ENSEMBLE PADA MODEL *DECISION TREE* ID.3, *RANDOM FOREST*  
DAN *REGRESI LOGISTIK MULTINOMIAL***

**Oleh :**

**Mau'izatil Rahmayani**

**NIM.08011281823038**

**ABSTRAK**

Kualitas Udara merupakan salah satu komponen penting bagi makhluk hidup dipermukaan bumi terutama untuk manusia. Kualitas udara harus dijaga dan dipelihara agar tidak mengalami penurunan dalam kualitasnya yang disebabkan oleh adanya faktor cuaca. Kualitas udara menjadi salah satu penyebab gangguan kesehatan manusia. Oleh karena itu tujuan dalam penelitian ini adalah memprediksi kualitas udara menggunakan metode *Ensemble* dengan tiga model klasifikasi. Penggunaan metode ensemble bertujuan untuk meminimalkan kesalahan pada pengklasifikasian dan mendapatkan nilai tingkat ketepatan lebih baik. Data yang digunakan dalam penelitian memiliki 21 variabel dengan jumlah data sebanyak 2502 data. Klasifikasi menggunakan metode *Ensemble Majority vote* berdasarkan tiga model algoritma *Decision Tree*, *Random Forest* Dan *Regresi Logistik Multinomial*. Hasil penelitian ini menunjukkan bahwa tingkat ketepatan prediksi kualitas udara menggunakan Metode *Ensemble Majority Vote* didapat nilai *accuracy* sebesar 99,31%, *precision makro* sebesar 78,45%, dan *recall makro* sebesar 78,63%, *fscore makro* sebesar 78,54 %, untuk *presisi*, *recall* dan *fscore mikro* memiliki nilai yang sama yaitu 98,28%.

Kata Kunci : Kualitas udara, *Ensemble Majority Vote*, *Decision Tree*, *Random Forest* *Regresi Logistik Multinomial*.

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>ii</b>
<b>HALAMAN PERSEMBAHAN.....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>ABSTRACT .....</b>	<b>viii</b>
<b>ABSTRAK .....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xiii</b>
<b>DAFTAR GAMBAR .....</b>	<b>xiv</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	6
1.3 Batasan Masalah .....	6
1.4 Tujuan Penelitian .....	7
1.5 Manfaat Penelitian .....	7
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>8</b>
2.1 Udara .....	8
2.2 Kualitas Udara .....	8
2.3 Faktor Kualitas Udara .....	9
2.4 Machine Learning .....	9
2.5 Statistical Machine Learning .....	10
2.6 Klasifikasi .....	10
2.7 Metode Ensemble .....	11
2.8 Tabel Kontingensi .....	11
2.9 Analisis Regresi .....	12
2.10 Regresi Logistik .....	12
2.11 <i>Regresi Logistik Multinomial</i> .....	13
2.12 Estimasi Parameter .....	15

2.13 Pengujian Parameter .....	19
2.13.1 Uji Simultan .....	19
2.13.2 Uji Parsial .....	19
2.14 Uji Kesesuaian Model .....	20
2.15 Odds Ratio .....	21
2.16 <i>Decision Tree</i> .....	21
2.16.1 Algoritma ID.3 .....	22
2.17 <i>Random Forest</i> .....	23
2.18 <i>Confusion Matrix</i> .....	24
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>27</b>
3.1 Tempat .....	27
3.2 Waktu .....	27
3.3 Alat .....	27
3.4 Metode Penelitian .....	27
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>32</b>
4.1 Deskripsi Data .....	32
4.2 Diskritisasi Data .....	34
4.3 Ringkasan Data .....	37
4.4 Partisi Data .....	41
4.5 Model Regresi Logistik Multinomial .....	42
4.5.1 Estimasi Parameter.....	42
4.5.2 Uji Serentak .....	45
4.5.3 Uji Parsial .....	46
4.5.4 Uji Kesesuaian Model .....	48
4.5.5 Odds Ratio .....	49
4.5.6 Model Terbaik .....	51
4.5.7 Probabilitas .....	51
4.6 Algoritma <i>Decision Tree</i> .....	55
4.7 Algoritma <i>Random Forest</i> .....	63
4.8 Metode Ensemble .....	70
4.9 Analisa Hasil .....	73

<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>75</b>
5.1 Kesimpulan .....	75
5.2 Saran .....	76
<b>DAFTAR PUSTAKA .....</b>	<b>77</b>
<b>LAMPIRAN .....</b>	<b>81</b>

## DAFTAR TABEL

Tabel 2.1 Tabel Kontigensi .....	11
Tabel 2.2 <i>Confusion Matrix</i> .....	25
Tabel 4.1 Keterangan Variabel .....	33
Tabel 4.2 Diskritisasi Data .....	34
Tabel 4.3 Data Train .....	42
Tabel 4.4 Data Test .....	42
Tabel 4.5 Hasil Uji Serentak .....	46
Tabel 4.6 Tabel Uji Wald .....	47
Tabel 4.7 Tabel Uji Kesesuaian Model .....	48
Tabel 4.8 Tabel Nilai Odds Ratio .....	50
Tabel 4.9 Hasil Prediksi Regresi Logistik Multinomial .....	52
Tabel 4.10 Hasil Klasifikasi Regresi Logistik Multinomial .....	53
Tabel 4.11 Confusion Matriks Regresi Logistik Multinomial .....	53
Tabel 4.12 Nilai Entrophy Algoritma Decision Tree .....	56
Tabel 4.13 Nilai Gain Algoritma Decision Tree .....	58
Tabel 4.14 Hasil Prediksi Algoritma Decision Tree .....	60
Tabel 4.15 Hasil Klasifikasi Algoritma Decision Tree.....	60
Tabel 4.16 Confusion Matrix Algoritma Decision Tree.....	61
Tabel 4.17 Nilai Entrophy Algoritma Random Forest .....	64
Tabel 4.18 Nilai Gain Algoritma Random Forest .....	65
Tabel 4.19 Hasil Prediksi Algoritma Random Forest .....	67
Tabel 4.20 Hasil Klasifikasi Algoritma Random Forest .....	68
Tabel 4.21 Confusion Matrix Algoritma Random Forest.....	68
Tabel 4.22 Hasil Prediksi Metode Ensemble .....	70
Tabel 4.23 Hasil Klasifikasi Metode Ensemble .....	71
Tabel 4.24 Confusion Matrix Metode Ensemble .....	71
Tabel 4.25 Analisa Hasil .....	73



## DAFTAR GAMBAR

Gambar 4.1 Ringkasan Data Variabel Dependen .....	37
Gambar 4.2 Ringkasan Data Variabel Independen .....	41
Gambar 4.3 Pohon Keputusan Node 1.1.1 .....	59
Gambar 4.4 Pohon Keputusan Node 1.1.1.1.1 .....	66

## DAFTAR LAMPIRAN

Lampiran 1 Iterasi New Rapshon metode regresi logistic multinomial .....	81
Lampiran 2 Hasil Estimasi Parameter .....	81
Lampiran 3 Hasil Estimasi Parameter, Statistik Uji wald, dan Odds Ratio pada Metode Regresi Logistik Multinomial.....	84
Lampiran 4 Pohon Keputusan Decision Tree.....	92
Lampiran 5 Pohon Keputusan Random Forest.....	105
Lampiran 6 Hasil Prediksi.....	116

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Udara merupakan campuran gas tidak berwarna yang terdapat pada permukaan bumi. Udara tidak tampak mata, tidak berbau, dan tidak ada rasanya. Udara hanya dapat dilihat dari adanya angin yang menggerakkan suatu benda (Mayssara A & Supervised, 2014). Udara termasuk salah satu jenis sumber daya alam karena memiliki banyak fungsi bagi makhluk hidup. Udara terdapat oksigen (O<sub>2</sub>) untuk bernafas, karbon dioksida (CO<sub>2</sub>) untuk proses fotosintesis, dan ozon (O<sub>3</sub>) untuk menahan sinar ultraviolet dari matahari (Isramadhanti, 2019)

Udara berperan dalam mempertahankan kehidupan makhluk hidup terutama bagi manusia. Udara dapat berubah-ubah dengan cepat karena adanya faktor cuaca. Faktor cuaca dapat mempengaruhi kualitas udara yang berakibat terjadinya pencemaran udara sehingga menimbulkan penurunan kualitas udara yang ada. Penurunan kualitas udara meningkat karena semakin berkembangnya pertumbuhan industri, pertumbuhan penduduk dan perkembangan kendaraan bermotor namun tidak sebanding dengan adanya penghijauan dan pelestarian pepohonan. Semakin baik tingkat kualitas udara maka tidak memberikan efek buruk bagi kesehatan makhluk hidup, begitupun sebaliknya jika kualitas udaranya buruk maka dapat berdampak buruk bagi kesehatan makhluk hidup (Prayudha et al., 2018).

Cuaca merupakan keadaan udara pada saat tertentu di wilayah yang relative sempit, dimana sifatnya dapat berubah-ubah dalam jangka waktu yang singkat (Puspita & Yulianti, 2016). Cuaca sulit untuk diprediksi disebabkan oleh

banyaknya unsur-unsur cuaca yang mempengaruhi. Ada beberapa unsur-unsur cuaca misalnya tekanan udara, kecepatan angin, kelembaban, curah hujan, suhu, dan fenomena atmosfer sebagai komponennya. Kondisi cuaca sangat penting dan tidak lepas dari kehidupan manusia (Siregar, 2020).

Klasifikasi merupakan proses pengelompokan suatu data menjadi beberapa kelompok yang saling keterkaitan satu sama lain dengan masalah yang ada hingga mendapatkan tujuan akhir dalam menemukan suatu model atau fungsi yang bisa menggambarkan dan membedakan setiap kelompok data sehingga dapat memprediksikan suatu kelompok data (Wicahyo et al., 2021).

Penelitian dilakukan dengan cara memprediksi kualitas udara menggunakan proses *statistical machine learning* dengan metode klasifikasi. Metode yang digunakan pada penelitian ini adalah Metode Ensemble dengan model *Decision Tree* ID.3, *Random Forest* dan *Regresi Logistik Multinomial*. Penggunaan metode ensemble bertujuan untuk meminimalkan kesalahan pada pengklasifikasian dan mendapatkan nilai tingkat ketepatan lebih baik. Dataset yang digunakan dari website *Kaggle.com*. Data berhubungan dengan kualitas udara di Shanghai, China. Data terdiri dari 2502 observasi dan 21 variabel yang digunakan selama 8 tahun.

Alasan memilih data kualitas udara menggunakan fitur cuaca karena cuaca sangat berpengaruh akibat adanya aktifitas manusia. Data ini memberikan gambaran penyebab terjadinya perubahan kualitas udara. Variabel yang dituju itu seperti suhu, ketebalan salju, angin, kelembaban, tekanan, curah hujan, kecepatan angin, hembusan angin dan sebagainya.

Metode Ensemble merupakan pengabungan beberapa algoritma klasifikasi untuk mendapatkan hasil kinerja yang lebih baik dibandingkan dengan menggunakan satu algoritma (Maulia et al., 2021). Metode Random Forest (RF) adalah pengembangann dari metode *Classification and Regression Tree* (CART) yang menerapkan metode *bootstrap aggregating* (bagging) dan *random feature selection*. Random Forest merupakan salah satu metode yang digunakan pada pengklasifikasian suatu data dengan membangun banyak pohon keputusan. Hasil pengklasifikasian dari setiap pohon yang didapat akan diakumulasikan dan dipilih berdasarkan hasil klasifikasi yang paling banyak muncul. Metode random forest terdiri dari root node (akar), internal node(cabang), dan leaf node(simpul akhir) (Sandag, 2020).

*Decision Tree* merupakan diagram alir yang berbentuk seperti pohon untuk setiap cabang dari pohon menyatakan suatu hasil dari pengujian yang dilakukan. *Decision Tree* sering disebut pohon keputusan. Metode *decision tree* terdiri dari beberapa atribut dalam setiap pohon yang disebut node. Node yang paling atas disebut root node(node akar), setiap cabang dari atribut disebut internal node, dan hasil akhir dari setiap label kelas atribut disebut leaf node (Kasih, 2019). Regresi logistik adalah salah satu metode yang digunakan dalam mencari hubungan setiap variabel independen yang bersifat dichotomous atau polychotomous. Regresi logistik terdiri dari beberapa jenis diantaranya adalah *Regresi Logistik Multinomial* yang bersifat polichotomous atau multinomial dimana memiliki jumlah variabel dependen lebih dari dua kategori (Subekti, 2014).

Penelitian oleh Sadik (2019), melakukan penelitian tentang survei pemilihan kontrasepsi pada perempuan yang menikah baik sedang hamil maupun sedang tidak hamil menggunakan pengukuran ketepatan klasifikasi pada model regresi logistik multinomial. Dari hasil penelitian yang dilakukan didapatkan nilai ketepatan klasifikasi adalah sebesar 52,3%.

Selanjutnya Penelitian oleh Hamami & Dahlan (2022), penelitian yang dilakukan menggunakan metode *Random Forest* untuk mengklasifikasikan cuaca dengan teknik oversampling untuk menangani jumlah data di kelas minoritas. Dari hasil penelitian yang dilakukan didapatkan hasil akurasi sebesar 70%

Penelitian oleh Setiawati et al. (2019), melakukan penelitian tentang penentuan klasifikasi penyakit liver yang paling berpengaruh dengan menggunakan metode *decision tree*. Dari penelitian tersebut menunjukkan hasil akurasi sebesar 72,67%.

Penelitian oleh Sang et al (2021), melakukan penelitian Untuk klasifikasi data kualitas udara DKI jakarta menggunakan algoritma decision tree dan support vector machine, didapatkan nilai akurasi pada algoritma Decision Tree mendapatkan nilai Precision sebesar 99,02%, Recall 99,73%, F1-Measure 99,37%, Akurasi 99,40% dan pada algoritma SVM mendapatkan nilai Precision sebesar 95,82%, Recall 88,89%, F1-Measure 92,22% dan Akurasi 94,93%.

Penelitian oleh Yu et al (2016), melakukan penelitian data publik dalam sistem penginderaan perkotaan, model memprediksi kualitas udara wilayah di Shenyang menggunakan algoritma random forest untuk memprediksi semua



wilayah yang tidak ditemukan di pusat kota. Dari penelitian yang dilakukan menghasilkan nilai presisi sebesar 81%.

Penelitian lain oleh Siregar (2020), melakukan penelitian tentang klasifikasi Prediksi Cuaca Menggunakan *Ensemble Learning*. Dari penelitian yang dilakukan didapat nilai akurasi untuk metode *naive bayes* sebesar 77,22%, *decision tree* sebesar 79%, *random forest* sebesar 82,38%, *deep learning* sebesar 82,92%, dan *generalized linier model* sebesar 84,05%. Untuk metode ensemble dengan bagging sebesar 81,21% dengan MSE 18,79%. Hasil metode ensemble pada penelitian tersebut lebih rendah dibanding dengan performance dari *Generalized linier model*, dilihat dari hasil akurasi *decision tree* lebih rendah, hasil akurasi *deep learning* diatas rata rata *ensambel learning* dan untuk *naive bayes* dibawah rata rata *ensambel learning*.

Penelitian-penelitian lain melihat bahwa metode *Random Forest*, *Decision Tree* dan *Regresi Logistik Multinomial* mendapatkan hasil klasifikasi dengan tingkat ketepatan yang baik. Pemilihan model yang digunakan karena dari setiap model memiliki riwayat hasil ketepatan yang baik. Oleh karena itu penulis mencoba melakukan penelitian dengan cara mencari prediksi terbaik menggunakan Metode Ensemble. Untuk mendapatkan hasil prediksi pada metode ensemble dilakukan secara voting dengan mengambil hasil prediksi terbanyak dari setiap model. Hasil prediksi terakhir yang akan diperoleh menggunakan Metode Ensemble dapat menghasilkan nilai *Accuracy*, *Precision Makro*, *Precision Mikro*, *Recall Makro*, *Recall Mikro*, *Fscore Makro* dan *Fscore Mikro* untuk menentukan tingkat ketepatan klasifikasi kualitas udara.

## 1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah :

1. Bagaimana hasil prediksi kualitas udara berdasarkan metode *Decision Tree ID.3, Random Forest dan Regresi Logistik Multinomial* ?
2. Bagaimana hasil prediksi kualitas udara menggunakan Metode Ensemble dengan model *Decision Tree ID.3, Random Forest dan Regresi Logistik Multinomial* ?
3. Bagaimana perbandingan tingkat ketepatan dalam memprediksi kualitas udara menggunakan Metode Ensemble dengan model *Decision Tree ID.3, Random Forest dan Regresi Logistik Multinomial*?

## 1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah:

1. Data kualitas udara menggunakan *dataset* dari *kaggle.com* kota Shanghai, China pada tahun 2014-2021. Data terdiri dari 21 variabel diantaranya adalah tanggal, *maxtemp, mintemp, totalsnow, sunhour, uvindex, moon illumination, dew point, feelslike, heat index, wind chill, wind gust, cloud cover, humidity, precip, pressure, temp, visibility, winddir degree, wind speed* dan *air quality index*.
2. Data yang digunakan sebanyak 2502 observasi, validasi data train tahun 2014-2018 dan data test tahun 2019-2021.

3. Tingkat ketepatan klasifikasi dibatasi oleh nilai *Accuracy*, *Precision Makro*, *Precision Mikro*, *Recall Makro*, *Recall Mikro*, *Fscore Makro* dan *Fscore Mikro*

#### **1.4 Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah :

1. Memprediksi kualitas udara berdasarkan model *Decision Tree ID.3*, *Random Forest* dan *Regresi Logistik Multinomial*.
2. Memprediksi kualitas udara menggunakan Metode Ensemble dengan model *Decision Tree ID.3*, *Random Forest* dan *Regresi Logistik Multinomial*.
3. Membandingkan tingkat ketepatan dalam memprediksi kualitas udara menggunakan Metode Ensemble dengan model *Decision Tree ID.3*, *Random Forest* dan *Regresi Logistik Multinomial*.

#### **1.5 Manfaat Penelitian**

Adapun manfaat dari penelitian ini adalah:

1. Dapat membantu dan mempermudah pembaca dalam mencari informasi terkait pengklasifikasian kualitas udara.
2. Sebagai bahan referensi penelitian lain yang juga membahas informasi terkait pengklasifikasian kualitas udara.
3. Sebagai media pembelajaran penulis dan pembaca dalam meningkatkan ilmu pengetahuan tentang pengklasifikasian kualitas udara.

## DAFTAR PUSTAKA

- Agus, A., Ahmad, M., Kusumaningtyas, S. D. A., & ... (2019). Analisis Dampak Diterapkannya Kebijakan Working From Home Saat Pandemi Covid-19 Terhadap Kondisi Kualitas Udara Di Jakarta. *Jurnal Meteorologi ...*, 6(3), 6–14. <https://jurnal.stmkg.ac.id/index.php/jmkg/article/view/141>
- Adnyani, L. P. (2019). Kajian Meteorologi dan Oseanografi Perairan Balikpapan-Tarakan sebagai Rute Health Boat untuk Pelayanan Kesehatan Pulau Terpencil Kalimantan. *SPECTA Journal of Technology*, 1(1), 11–21. <https://doi.org/10.35718/specta.v1i1.71>
- Brawijaya, J., Djohan, F., & Suryaningrum, K. M. (2020). Aplikasi Pendeteksi Dan Analisa Cuaca Menggunakan Metode K-Nearest Neighbor Berbasis Android. *JSTIE (Jurnal Sarjana Teknik Informatika) (E-Journal)*, 8(2), 43. <https://doi.org/10.12928/jstie.v8i1.14623>
- Budyandra, B., & Azzahra, G. N. (2017). Penerapan Regresi Logistik Ordinal Proportional Odds Model pada Analisis Faktor-Faktor yang Mempengaruhi Kelengkapan Imunisasi Dasar Anak Balita di Provinsi Aceh Tahun 2015. *Media Statistika*, 10(1), 37. <https://doi.org/10.14710/medstat.10.1.37-47>
- Dewi, N. K., Mulyadi, S. Y., & Syafitri, U. D. (2012). Penerapan Metode Random Forest Dalam Driver Analysis. *Forum Statistika Dan Komputasi*, 16(1), 35–43. <http://journal.ipb.ac.id/index.php/statistika/article/view/5443>
- Dr. Kusman Sadik. (2019). Regresi Logistik Multinomial. *Departemen Statistika IPB*, 1–19.
- Fajar Nugroho. (n.d.). *UNIKOM\_Fajar Nugroho\_BAB II*. 6–27.
- Hadjar, I. (2018). Regresi Logistik. *Phenomenon : Jurnal Pendidikan MIPA*, 7(2), 137.
- Hamami, F., & Dahlan, I. A. (2022). Klasifikasi Cuaca Provinsi Dki Jakarta Menggunakan Algoritma Random Forest Dengan Teknik Oversampling. *Jurnal Teknoinfo*, 16(1), 87. <https://doi.org/10.33365/jti.v16i1.1533>
- Hartono, S. (2020). Journal Of Technology Information. *Journal Of Technology Information*, 5(1), 25–30. <http://jurnal.kampuswiduri.ac.id/index.php/infoteh/article/view/79/67>
- Hosmer, D.W. dan S. Lemeshow, (2000) Applied Logistic Regression. Second Edition, John Willey & Sons, New York.
- Ii, B. A. B., Pustaka, T., & Dasar, D. A. N. (2018). *Bab Ii Tinjauan Pustaka Dan Dasar Teori 2.1 Tinjauan Pustaka Penelitian terkait dengan Metode*. 2015, 5–

11.

- Isramadhanti, H. W. (2019). Gambaran Kualitas Udara di Kota Yogyakarta Berdasarkan Pemantauan Air Quality Monitoring System tahun 2019-2020. *Skripsi. Politeknik Kesehatan Kemenkes Yogyakarta*, 30–48.
- Jaya, B. (2021). Estimasi Parameter Regresi Logistik Multinomial Menggunakan Maksimum Likelihood. *Repositori Universitas Sumatera Utara*, 45. [https://repositori.usu.ac.id/bitstream/handle/123456789/30765/160803079.pdf?sequence=1&isAllowed=y#:~:text=Regresi logistik multinomial atau disebut,satu atau lebih variabel bebas.](https://repositori.usu.ac.id/bitstream/handle/123456789/30765/160803079.pdf?sequence=1&isAllowed=y#:~:text=Regresi%20logistik%20multinomial%20atau%20disebut,satu%20atau%20lebih%20variabel%20bebas.)
- Kasih, P. (2019). Modeling Data Mining Decision Tree with Classification Error for Selection of Candidates for Choir Team Members. *Innovation in Research of Informatics (INNOVATICS)*, 1(2), 63–69.
- Komunikasi, F., Surakarta, U. M., Yani, J. A., & Pos, T. (2017). Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest. *Jurnal Teknik Elektro*, 9(1), 24–29. <https://doi.org/10.15294/jte.v9i1.10452>
- Kurniawan, A. B. (2020). Pendekatan random forest untuk memprediksi nasabah yang berpotensi membuka tabungan deposito. *Repository.Usd.Ac.Id*. [https://repository.usd.ac.id/37794/2/165314039\\_full.pdf](https://repository.usd.ac.id/37794/2/165314039_full.pdf)
- LLC, B. (2022). Resilience Metrics: Engineering Reference — EnergyPlus 9.4. <https://bigladdersoftware.com/epx/docs/9-4/engineering-reference/resilience-metrics.html>
- Lufritayanti, & Annisa. (2013). Disusun Oleh : Disusun Oleh : *Pengetahuan Dan Sikap Dalam Penelitian Kesehatan*, 11150331000034, 1–147.
- Lutfia Rohmi, A. (2017). *Analisis regresi logistik multinomial pada jenis pelanggaran lalu lintas di kota surabaya*. 40.
- MAULIA, O. R. I., Resti, Y., & Zayanti, D. A. (2021). ... *Jantung Koroner Dengan Metode Ensemble Menggunakan Algoritma Naive Bayes, Decision Tree C4. 5 Dan Regresi Logistik Biner*. [https://repository.unsri.ac.id/55008/0Ahttps://repository.unsri.ac.id/55008/3/RAMA\\_44201\\_08011181722056\\_0019077302\\_0004127001\\_01\\_front\\_ref.pdf](https://repository.unsri.ac.id/55008/0Ahttps://repository.unsri.ac.id/55008/3/RAMA_44201_08011181722056_0019077302_0004127001_01_front_ref.pdf)
- Mayssara A. Abo Hassanin Supervised, A. (2014). Udara. *Paper Knowledge . Toward a Media History of Documents*, 6–35.
- Nordhausen, K. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition by Trevor Hastie, Robert Tibshirani, Jerome Friedman. *International Statistical Review*, 77(3), 482–482.

[https://doi.org/10.1111/j.1751-5823.2009.00095\\_18.x](https://doi.org/10.1111/j.1751-5823.2009.00095_18.x)

- Nugroho, E. S. (2009). *Analisis Kualitas Udara Di Daerah Istimewa Yogyakarta (Diy) Tahun 2002 – 2008 Sebagai Sumber Belajar Siswa Kelas Vii Sltp/Mts.* 1–27. [http://digilib.uin-suka.ac.id/3684/1/BAB I%2C V%2C DAFTAR PUSTAKA.pdf](http://digilib.uin-suka.ac.id/3684/1/BAB%20V%20DAFTAR%20PUSTAKA.pdf)
- Prayudha, J., Pranata, A., & Al Hafiz, A. (2018). Implementasi Metode Fuzzy Logic Untuk Sistem Pengukuran Kualitas Udara Di Kota Medan Berbasis Internet of Things (Iot). *Jurteksi*, 4(2), 141–148. <https://doi.org/10.33330/jurteksi.v4i2.57>
- Priyam, A., Gupta, R., Rathee, A., & Srivastava, S. (2013). *Comparative Analysis of Decision Tree Classification Algorithms.* 334–337.
- Probabilistik Curah Hujan 20 mm (tiap 24 jam) | BMKG. (2022). [https://www.bmkg.go.id/cuaca/probabilistik-curah-hujan.bmkg\\_](https://www.bmkg.go.id/cuaca/probabilistik-curah-hujan.bmkg_) diakses pada tanggal 18 Maret 2022
- Puspita, E. S., & Yulianti, L. (2016). Perancangan Sistem Peramalan Cuaca Berbasis Logika Fuzzy. *Jurnal Media Infotama*, 12(1). <https://doi.org/10.37676/jmi.v12i1.267>
- Putri, N. I., & Budyanra. (2019). Penerapan Regresi Logistik Ordinal Dengan Proportional Odds Model Pada Determinan Tingkat Stres Akademik Mahasiswa ( Studi Kasus pada Mahasiswa Tingkat I Politeknik Statistika STIS Tahun Akademik 2018 / 2019 ). *Seminar Nasional Official Statistics*, 368–378.
- Portal Statistik Sektor Provinsi Dki Jakarta (2022). Indeks Standar Kualitas Udara Jakarta Selama Masa PSBB. <https://statistik.jakarta.go.id/indeks-standar-pencemaran-udara-ispu-selama-masa-psbb-di-dki-jakarta/>
- Radical air mass changes ahead for Minnesota. (2022). <https://www.mprnews.org/story/2018/06/12/fresh-to-oppressive-radical-air-mass-changes-ahead-for-minnesota>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Sandag, G. A. (2020). Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest. *CogITO Smart Journal*, 6(2), 167. <https://doi.org/10.31154/cogito.v6i2.270.167-178>
- Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). *Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma*



*Decision Tree Dan Support Vector Machine Data Mining Analysis For Classification Of Air Quality Data Dki Jakarta Using Decision Tree Algorithm And Support Vector.* 8(5), 8954–8963.

- Scott, A. J., Hosmer, D. W., & Lemeshow, S. (1991). Applied Logistic Regression. In *Biometrics* (Vol. 47, Issue 4). <https://doi.org/10.2307/2532419>
- Setiawati, I., Wibowo, A. P., & Hermawan, A. (2019). Pendahuluan Tinjauan Pustaka Penelitian Sebelumnya Klasifikasi. *Jurnal of Information System Management*, 1(1), 13–17.
- Siregar, A. M. (2020). Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning. *Petir*, 13(2), 138–147. <https://doi.org/10.33322/petir.v13i2.998>
- Subekti, P. (2014). Model regresi logistik multinomial untuk menentukan pilihan sekolah lanjutan tingkat atas pada siswa SMP. *Cauchy*, 3(2), 91. <https://doi.org/10.18860/ca.v3i2.2577>
- True, K. (2022). Saratoga-Weather.org - UV Index., <https://saratoga-weather.org/uv-index.php>
- Tyasti, A. E., Ispriyanti, D., & Hoyyi, A. (2015). Algoritma Iterative Dichotomiser 3 (ID3) untuk Mengidentifikasi Data Rekam Medis. *Gaussian*, 4(Dm), 237–246.
- Utami, D. D. (2018). *Penerapan Regresi Logistik Multinomial Terhadap Pemilihan Alat Kontrasepsi Pada PUS Di Desa Telaga Sari Kecamatan Tanjung Morawa Tahun 2018.*
- Weather Spark (2022) <https://id.weatherspark.com/y/116847/Cuaca-Rata-rata-pada-bulan-in-Jakarta-Indonesia-Sepanjang-Tahun>
- Wicahyo, A., Pudoli, A., & Kusumaningsih, D. (2021). *Penggunaan Algoritma Naive Bayes dalam klasifikasi Pengaruh Pencemaran Udara.* 20, 103–108.
- Yu, R., Yang, Y., Yang, L., Han, G., & Move, O. A. (2016). RAQ—A random forest approach for predicting air quality in urban sensing systems. *Sensors (Switzerland)*, 16(1). <https://doi.org/10.3390/s16010086>
- Zahroh, Z. Z., & Zain, I. (2019). Analisis Regresi Logistik Multinomial Pada Faktor-Faktor Yang Mempengaruhi Sumber Air Bersih Rumah Tangga Di Jawa Timur. *Jurnal Sains Dan Seni ITS*, 7(2). <https://doi.org/10.12962/j23373520.v7i2.34701>