

**PERBANDINGAN METODE *K-NEAREST NEIGHBOR*
DAN *DECISION TREE* UNTUK KLASIFIKASI KUALITAS
UDARA BERDASARKAN *BOOTSTRAP SAMPLING***

SKRIPSI

Oleh :

ANINDITHA ARIANI FAHIRA

NIM. 08011281924035



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

2023

LEMBAR PENGESAHAN

PERBANDINGAN METODE *K-NEAREST NEIGHBOR*
DAN *DECISION TREE* UNTUK KLASIFIKASI KUALITAS UDARA
BERDASARKAN *BOOTSTRAP SAMPLING*

SKRIPSI

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Matematika

Oleh

ANINDITHA ARIANI FAHIRA
NIM. 08011281924035

Pembimbing Kedua



Dra. Ning Eliyati, M.Pd
NIP.195911201991022001

Indralaya, Mei 2023
Pembimbing Utama



Des Alwine Zavanti, S.Si., M.Si
NIP.197012041998022001

Mengetahui,
Ketua Jurusan Matematika



Drs. Sugandi Yahdin, M.M
NIP. 19580727198603 1003

PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Aninditha Ariani Fahira
NIM : 08011281924035
Fakultas/Jurusan : MIPA/Matematika

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai pemenuhan persyaratan untuk memperoleh gelar kesarjanaan starata satu (S1) dari Universitas Sriwijaya maupun perguruan tinggi lain.

Semua informasi yang dimuat dalam skripsi ini berasal dari penulis lain baik yang dipublikasikan atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis secara benar. Semua isi dari skripsi ini sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 3 Juni 2023

A handwritten signature in black ink is written over a red circular stamp. The stamp contains the text '10000' at the top, 'ASISTEN TENGAH' in the middle, and 'B2A1AF.X 75368410' at the bottom. The signature is written in a cursive style, with the initials 'A.F.' being prominent.

Penulis

HALAMAN PERSEMBAHAN

MOTTO

“Jalani semuanya sesuai dengan alurnya, karena manusia hanya mampu berencana dan Allah *Subhanahu Wa Ta’ala* yang menentukan“

“Hati-hati, baca doa, dan niatkan menuntut ilmu sebagai ibadah kepada Allah *Subhanahu Wa Ta’ala*”

-Ayah-

Kupersembahkan skripsi ini untuk :

- 1. Allah Subhanahu Wa Ta’ala,**
- 2. Kedua orang tuaku tersayang,**
- 3. Adikku,**
- 4. Keluarga besarku,**
- 5. Guru dan dosenku,**
- 6. Sahabat-sahabatku,**
- 7. Almamaterku**

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Dengan mengucapkan Alhamdulillah, segala puji bagi Allah Subhanahu wa Ta'ala yang telah memberikan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Perbandingan Metode *K-Nearest Neighbor* dan *Decision Tree* Untuk Klasifikasi Kualitas Udara Berdasarkan *Bootstrap Sampling*”** sebagai salah satu syarat memperoleh gelar sarjana sains studi Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Penulis menyadari bahwa proses pembuatan skripsi ini sebagai proses pembelajaran yang sangat berharga yang tak lepas dari kekurangan dan keterbatasan. Dengan segala hormat dan kerendahan hati, penulis mengucapkan terima kasih dan penghargaan yang tulus penulis sampaikan kepada kedua orang tuaku tercinta, **Nurul Iman Wahyudi, A.md** dan **Dra. Ida Rosidawati**, yang tak pernah lelah mendidik, membimbing, menasehati, dan mendukung serta tak henti untuk mendoakan anaknya. Terima kasih atas segala perjuangan dan pengorbanan hingga detik ini dan sampai kapanpun. Penulis juga menyampaikan terima kasih dan penghargaan 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 Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
3. Ibu **Dr. Dian Cahyawati Sukanda, M.Si** selaku Sekretaris Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

4. Ibu **Des Alwine Zayanti, S.Si., M.Si** selaku Dosen Pembimbing Utama yang telah bersedia meluangkan waktu, tenaga, pikiran untuk memberikan bimbingan dan arahan dengan penuh perhatian, pengertian, dan kesabaran sehingga skripsi ini dapat diselesaikan.
5. Ibu **Dra. Ning Eliyati, M.Pd** selaku Dosen Pembimbing Kedua yang telah bersedia meluangkan waktu, tenaga, pikiran untuk memberikan bimbingan dan arahan dengan penuh perhatian, pengertian, dan kesabaran sehingga skripsi ini dapat diselesaikan.
6. Ibu **Indrawati, S.Si., M.Si** dan **Eka Susanti, S.Si., M.Sc** selaku Dosen Pembahas yang telah memberikan tanggapan, kritik, dan saran yang sangat bermanfaat untuk perbaikan dan penyelesaian skripsi ini.
7. Ibu **Dr. Yulia Resti, S.Si., M.Si** selaku Dosen Pembimbing Akademik yang sangat baik telah memberikan saran, membimbing, membantu, dan mengarahkan urusan akademik penulis setiap semester.
8. **Seluruh Dosen di Jurusan Matematika FMIPA** yang telah memberikan ilmu, nasihat, motivasi, serta bimbingan selama proses perkuliahan.
9. Pak **Irwansyah** selaku admin dan Ibu **Hamidah** selaku pegawai tata usaha Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah banyak membantu penulis selama perkuliahan.
10. **Seluruh guru** yang telah memberikan ilmu yang bermanfaat hingga mengantarkan penulis pada pendidikan ini.

11. Adikku tersayang, **Ananda Nawal Syahidah** yang selalu mendoakan, memberikan semangat, dan tempat bercerita serta berbagi selama ini, beserta keluarga besar yang selalu mendukung penulis.
12. **Sahabat-sahabat terbaik** yang pernah penulis miliki dari masa sekolah hingga saat ini, terima kasih sudah membantu, mengingatkan, dan mendukung dengan tulus selama ini.
13. **Keluarga Matematika 2019, BPH Himastik Gelora Karya dan Akselerasi,** dan **rekan-rekan** selama perkuliahan.
14. Kakak-kakak tingkat Angkatan 2017 dan 2018 serta adik-adik tingkat Angkatan 2020 dan 2021, terima kasih atas segala kebaikan dan bantuan.
15. Semua pihak yang tidak dapat penulis sebutkan satu per satu. Semoga segala kebaikan yang diberikan mendapatkan balasan terbaik dari Allah.

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.

Indralaya, Mei 2023

Penulis

**COMPARISON OF *K-NEAREST NEIGHBOR* METHODS
AND *DECISION TREE* FOR AIR QUALITY CLASSIFICATION BASED ON
*BOOTSTRAP SAMPLING***

By:

ANINDITHA ARIANI FAHIRA

08011281924035

ABSTRACT

Air can change based on factors that affect it and have an impact on air quality. Air quality plays an important role for organisms on the surface of the earth, especially for humans. Air quality must be clean so as not to have harmful impacts. Classification is a way to determine the level of air quality. Classification testing in this study uses two different methods, namely, K-Nearest Neighbor (KNN) and Decision Tree based on air quality data that has gone through a Bootstrap Sampling process. Therefore, the author tried to conduct research by comparing KNN and Decision Tree methods. The final accuracy of the study was measured by accuracy, precision, recall, and Fscore. The accuracy produced by the KNN method is 97%, 97.03%, 84.25%, and 89.60% respectively. The accuracy produced by the Decision Tree method is 95.18%, 89.39%, 80.72%, and 84.45% respectively so that the KNN method shows better accuracy results than the Decision Tree method.

Keywords : Air Quality, KNN, Decision Tree, Bootstrap Sampling

**PERBANDINGAN METODE *K-NEAREST NEIGHBOR*
DAN *DECISION TREE* UNTUK KLASIFIKASI KUALITAS UDARA
BERDASARKAN *BOOTSTRAP SAMPLING***

Oleh:

ANINDITHA ARIANI FAHIRA

08011281924035

ABSTRAK

Udara dapat berubah-ubah berdasarkan faktor-faktor yang memengaruhinya dan berakibat pada kualitas udara. Kualitas udara memegang peranan penting bagi organisme di permukaan bumi terutama untuk manusia. Kualitas udara harus bersih agar tidak memiliki dampak berbahaya. Pengklasifikasian merupakan cara untuk mengetahui tingkatan kualitas udara. Pengujian klasifikasi dalam penelitian ini menggunakan dua metode yang berbeda yaitu, *K-Nearest Neighbor (KNN)* dan *Decision Tree* berdasarkan data kualitas udara yang sudah melalui proses *Bootstrap Sampling*. Oleh karena itu penulis mencoba melakukan penelitian dengan cara membandingkan metode *KNN* dan *Decision Tree*. Ketepatan akhir penelitian diukur oleh *accuracy*, *precision*, *recall*, dan *Fscore*. Ketepatan yang dihasilkan oleh metode *KNN* secara berturut-turut adalah 97%, 97,03%, 84,25%, dan 89,60% Ketepatan yang dihasilkan oleh metode *Decision Tree* secara berturut-turut adalah 95,18%, 89,39%, 80,72%, dan 84,45% sehingga metode *KNN* menunjukkan hasil ketepatan yang lebih baik dibandingkan dengan metode *Decision Tree*.

Kata Kunci : Kualitas Udara, *KNN*, *Decision Tree*, *Bootstrap Sampling*

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	ii
KATA PENGANTAR.....	iii
ABSTRACT.....	vi
ABSTRAK	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	Error! Bookmark not defined.
2.1 Udara	Error! Bookmark not defined.
2.2 Kualitas Udara.....	Error! Bookmark not defined.
2.3 Faktor Kualitas Udara	Error! Bookmark not defined.
2.4 Variabel	Error! Bookmark not defined.
2.5 <i>Statistical Machine Learning</i>	Error! Bookmark not defined.
2.6 Klasifikasi	Error! Bookmark not defined.
2.7 Diskritisasi.....	Error! Bookmark not defined.
2.8 Metode <i>K-Nearest Neighbor</i>	Error! Bookmark not defined.
2.9 Metode <i>Decision Tree ID3</i>	Error! Bookmark not defined.
2.10 <i>Bootstrap Sampling</i>	Error! Bookmark not defined.
2.11 Himpunan <i>Fuzzy</i>	Error! Bookmark not defined.

2.12 Fungsi Keanggotaan.....	Error! Bookmark not defined.
2.13 <i>Confusion Matrix</i>	Error! Bookmark not defined.
BAB III METODE PENELITIAN	Error! Bookmark not defined.
3.1 Tempat.....	Error! Bookmark not defined.
3.2 Waktu	Error! Bookmark not defined.
3.3 Metode Penelitian.....	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN	Error! Bookmark not defined.
4.1 Deskripsi Data	Error! Bookmark not defined.
4.2 <i>Dataset</i> Penelitian	Error! Bookmark not defined.
4.3 <i>Bootstrap Sampling</i>	Error! Bookmark not defined.
4.4 Klasifikasi Menggunakan Metode <i>K-Nearest Neighbor</i>	Error! Bookmark not defined.
4.4.1 Contoh Perhitungan Manual Metode <i>K-Nearest Neighbor</i>	Error! Bookmark not defined.
4.4.2 Evaluasi <i>Machine Learning</i> Metode <i>K-Nearest Neighbor</i>	Error! Bookmark not defined.
4.5 Diskritisasi Data	Error! Bookmark not defined.
4.6 <i>Bootstrap Sampling</i> Data Baru.....	Error! Bookmark not defined.
4.7 Klasifikasi Menggunakan Metode <i>Decision Tree</i> ..	Error! Bookmark not defined.
4.7.1 Contoh Perhitungan Manual Metode <i>Decision Tree</i>	Error! Bookmark not defined.
4.7.2 Evaluasi <i>Machine Learning</i> Metode <i>Decision Tree</i>	Error! Bookmark not defined.
BAB V KESIMPULAN DAN SARAN	Error! Bookmark not defined.
5.1 Kesimpulan.....	Error! Bookmark not defined.
5.1 Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA	7
LAMPIRAN.....	Error! Bookmark not defined.

DAFTAR TABEL

Tabel 2.1	<i>Confusion Matrix Multiclass</i>	20
Tabel 2.2	Kategori Evaluasi Kinerja Model	22
Tabel 3.1	Variabel Data Kualitas Udara	24
Tabel 3.2	Kategori Variabel.....	25
Tabel 4.1	Keterangan Variabel	29
Tabel 4.2	Dataset Kualitas Udara.....	30
Tabel 4.3	Data Contoh untuk Perhitungan Manual <i>Bootstrap Sampling</i>	31
Tabel 4.4	Pengambilan Sampel Secara Acak.....	32
Tabel 4.5	Data Baru pada Data Contoh <i>Bootstrap Sampling</i>	33
Tabel 4.6	Dataset Setelah <i>Bootstrap Sampling</i>	34
Tabel 4.7	Data Contoh untuk <i>KNN</i>	35
Tabel 4.8	Normalisasi Data Contoh <i>KNN</i>	36
Tabel 4.9	Klasifikasi Data Contoh ke-14.....	38
Tabel 4.10	<i>Confusion Matrix Multiclass</i> Metode <i>KNN</i>	39
Tabel 4.11	Diskritisasi Data.....	46
Tabel 4.12	Hasil Diskritisasi Data	50
Tabel 4.13	<i>Bootstrap Sampling</i> Data Baru	51
Tabel 4.14	Data Contoh untuk <i>Decision Tree</i>	52
Tabel 4.15	Hasil Perhitungan <i>Entropy</i> masing-masing Variabel.....	54
Tabel 4.16	Nilai <i>Gain</i> pada Contoh Data.....	56
Tabel 4.17	<i>Confusion Matrix Multiclass</i> Metode <i>Decision Tree</i>	58
Tabel 4.18	Hasil Visualisasi <i>Confusion Matrix</i>	59

Tabel 4.19 Evaluasi <i>Confusion Matrix</i>	59
Tabel 4.20 Hasil Ketepatan.....	60

DAFTAR GAMBAR

Gambar 2.1 Representasi Linear Naik	17
Gambar 2.2 Representasi Linear Turun	18
Gambar 2.3 Representasi Kurva Segitiga	19
Gambar 2.3 Representasi Kurva Trapesium	20
Gambar 4.1 Ilustrasi X_6 sebagai <i>Root Node</i>	57

BAB I

PENDAHULUAN

1.1 Latar Belakang

Klasifikasi merupakan proses pembangunan model atau aturan klasifikasi tertentu dengan menggunakan *data training* sebagai tahapan pembelajaran dan *data testing* sebagai tahapan pengujian sehingga klasifikasi termasuk kedalam metode *supervised learning*. Pembangunan model tersebut berdasarkan ciri-ciri yang dimiliki oleh setiap objek yang ada dan mengelompokan objek ke dalam kelas yang sesuai (Wibawa *et al.*, 2018).

Decision Tree dan *K-Nearest Neighbor* merupakan metode yang termasuk klasifikasi. *Decision Tree* merupakan diagram alir yang berbentuk seperti pohon dan setiap cabang dari pohon yang terbentuk menyatakan suatu hasil yang dilakukan. Metode ini memiliki *node* pohon yang merepresentasikan variabel yang telah diuji dan cabang yang terbentuk merupakan suatu pembagian hasil uji serta *node* daun (*leaf*) merepresentasikan kelompok kelas tertentu (Setio *et al.*, 2020). *Decision Tree* dibangun secara *top-down* (dari atas ke bawah), dimana tahap awal yang dilakukan adalah mengevaluasi seluruh variabel yang ada dengan menggunakan suatu ukuran statistik untuk mengukur keefektifan suatu variabel dalam mengklasifikasikan sekumpulan sampel data (Kasih, 2019).

K-Nearest Neighbor (*KNN*) adalah salah satu metode klasifikasi yang memiliki prinsip kerja berdasarkan kedekatan jarak atau lokasi suatu data dengan data yang lain (Elly Pusporani *et al.*, 2019). Penyelesaian *KNN* dipengaruhi oleh beberapa hal seperti

nilai K dan memilih ukuran jarak. Memilih nilai K dengan menetapkan konstanta tetap ke setiap data pengujian untuk disetiap titik data pengujian (Zhang *et al.*, 2017).

Penelitian ini dilakukan dengan membandingkan metode *KNN* dan *Decision Tree* berdasarkan data kualitas udara yang sudah melalui proses *Bootstrap Sampling*. *Bootstrap Sampling* adalah suatu metode yang digunakan untuk membangun interval kepercayaan untuk statistik ketika ukuran sampel kecil dan tidak terdistribusi dengan baik. *Bootstrap Sampling* dapat juga mengatasi dataset yang besar perlu dilakukan sampel data secara acak agar data yang diproses menjadi lebih kecil (Setiawan & Karomi, 2017).

Peranan udara dalam mempertahankan kehidupan makhluk hidup sangat penting terutama bagi manusia. Udara dapat diartikan sebagai campuran gas tidak berwarna yang terdapat pada permukaan bumi. Secara fisik, udara tidak terlihat, tidak berbau, dan tidak memiliki rasa. Udara merupakan salah satu jenis sumber daya alam yang digunakan oleh makhluk hidup. Komponen yang terdapat dalam udara pun memiliki fungsi yang berbeda, oksigen (O_2) dibutuhkan manusia agar proses respirasi berlangsung (Purba & Harefa, 2019), karbon dioksida (CO_2) untuk proses fotosintesis, dan ozon (O_3) yang memiliki fungsi menahan sinar *ultraviolet* dari matahari. Udara dapat berubah-ubah berdasarkan faktor-faktor yang memengaruhinya, seperti faktor cuaca yang dapat memengaruhi kualitas udara sehingga berakibat terjadinya pencemaran udara hingga menyebabkan penurunan kualitas udara. Faktor lain yang memengaruhi penurunan kualitas udara adalah pertumbuhan industri serta

perkembangan pesat kendaraan bermotor dan peningkatan pertumbuhan penduduk yang tidak sebanding dengan pertumbuhan ruang terbuka hijau dan pelestarian kawasan hijau, pada dasarnya semakin baik kualitas udara maka semakin baik pula dampaknya terhadap kesehatan makhluk hidup, begitupun sebaliknya (Prayudha *et al.*, 2018).

Cuaca merupakan keadaan udara yang memiliki beberapa unsur yang menempati suatu tempat tertentu dengan jangka waktu tertentu. Cuaca disetiap wilayah tidak selalu sama dan dapat berubah-ubah dalam waktu singkat (Puspita & Yulianti, 2016). Unsur-unsur yang memengaruhi cuaca antara lain suhu atau temperatur udara, tekanan udara, kelembaban udara, curah hujan, intensitas cahaya matahari, dan angin. Unsur-unsur pada cuaca tersebut juga memengaruhi pengklasifikasian kualitas udara, dengan kata lain kondisi cuaca sangat penting dalam kehidupan makhluk hidup.

Nurdiana & Algifari (2020), melakukan studi komparasi menggunakan *Decision Tree* dalam pengklasifikasian diabetes mellitus berdasarkan data penderita diabetes yang dimiliki tenaga kesehatan. Dari penelitian yang dilakukan didapatkan nilai ketepatan klasifikasi sebesar 74%. Penelitian lain yang dilakukan oleh T.P & Indah Sari (2017), untuk memperkirakan cuaca di kota Malang dengan menggunakan metode *Decision Tree* mendapatkan nilai ketepatan sebesar 73,74 %. kedua penelitian sudah masuk kedalam Batasan cukup yang dihasilkan pada sebuah penelitian.

Amalia *et al.* (2022), melakukan prediksi kualitas udara di Daerah Khusus Ibukota Jakarta yang didasari dengan Indeks Standar Pencemaran Udara (ISPU) dengan menggunakan *KNN* memperoleh hasil yang berbeda berdasarkan *K* yang

digunakan, hasil pengujian terbesar didapatkan dari $K = 7$ dengan hasil 96%. Penelitian lain yang dilakukan oleh Siswanto *et al.* (2018), untuk mengklasifikasikan suatu desa berdasarkan Indeks Pembangunan Desa (IPD) yang diterbitkan oleh Badan Pusat Statistik (BPS) tahun 2014 menghasilkan ketepatan nilai sebesar 92,22% dengan $K=3$.

Agustian & Bisri (2019), melakukan penelitian menggunakan *Decision Tree* dengan data pengajuan kredit menghasilkan ketepatan nilai sebesar 85,99% dan setelah melakukan *Bootstrap Sampling* menghasilkan nilai yang lebih baik yaitu sebesar 88,16%. Penelitian lain yang dilakukan oleh Setiawan & Karomi (2017) dengan menggunakan *KNN* pada dataset berdimensi tinggi menghasilkan ketepatan nilai sebesar 91,52% dan setelah melakukan *Bootstrap Sampling* menghasilkan nilai yang lebih baik yaitu sebesar 96,87%.

Penelitian yang melibatkan metode *KNN* dan *Decision Tree* mendapatkan hasil dengan tingkat ketepatan cukup sampai dengan baik. Pemilihan metode *KNN* dan *Decision Tree* digunakan karena setiap metode memiliki riwayat hasil ketepatan yang baik. Oleh karena itu penulis mencoba melakukan penelitian dengan cara membandingkan metode *KNN* dan *Decision Tree* berdasarkan *Bootstrap Sampling*. Ukuran ketepatan yang diperoleh adalah nilai *Accuracy*, *Precision*, *Recall*, dan *Fscore* untuk menentukan tingkat ketepatan klasifikasi kualitas udara.

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana hasil *accuracy*, *precision*, *recall*, dan *fscore* klasifikasi kualitas udara berdasarkan *Bootstrap Sampling* menggunakan metode *KNN* ?
2. Bagaimana hasil *accuracy*, *precision*, *recall*, dan *fscore* klasifikasi kualitas udara berdasarkan *Bootstrap Sampling* menggunakan metode *Decision Tree* dengan data yang telah didiskritisasi ?
3. Bagaimana perbandingan metode *KNN* dan *Decision Tree* berdasarkan hasil *accuracy*, *precision*, *recall*, dan *fscore* untuk klasifikasi kualitas udara?

1.3 Batasan Masalah

Adapun Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan data kualitas udara yang bersumber dari *Kaggle.com* kota Shanghai, China pada tahun 2014-2021. Variabel yang digunakan sebanyak 20 variabel dari total 21 variabel yang terdapat pada data.
2. Data kualitas udara yang telah melalui proses *Bootstrap Sampling* diolah menggunakan *KNN*.
3. Data kualitas udara yang telah didiskritisasi dan *Bootstrap Sampling* diolah menggunakan metode *Decision Tree Iterative Dichotomiser 3 (ID3)*.
4. Tingkat ketepatan klasifikasi dibatasi oleh nilai *accuracy*, *precision*, *recall*, dan *fscore*.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut :

1. Mendapatkan hasil *accuracy*, *precision*, *recall*, dan *fscore* klasifikasi kualitas udara dengan metode *KNN* dan telah melalui proses *Bootstrap Sampling*.
2. Mendapatkan hasil *accuracy*, *precision*, *recall*, dan *fscore* klasifikasi kualitas udara menggunakan metode *Decision Tree ID3* dengan data yang telah didiskritisasi dan melalui proses *Bootstrap Sampling*.
3. Membandingkan metode *KNN* dan *Decision Tree* berdasarkan hasil *accuracy*, *precision*, *recall*, dan *fscore* untuk klasifikasi kualitas udara.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Sebagai media pembelajaran peneliti dan pembaca dalam meningkatkan ilmu pengetahuan tentang kualitas udara dan metode klasifikasi menggunakan *KNN* dan *Decision Tree*.
2. Sebagai bahan referensi peneliti lain yang memiliki topik terkait pengklasifikasian kualitas udara.

DAFTAR PUSTAKA

- Afandi, M. A., Purnama, S. I., & Crisianti, R. F. (2020). Implementasi Metode Deteksi Tepi Laplacian dan Jarak Euclidean untuk Identifikasi Tanda Tangan. *Jurnal Nasional Teknik Elektro*, 9(1), 34. <https://doi.org/10.25077/jnte.v9n1.756.2020>
- Agustian, A. A., & Bisri, A. (2019). Data Mining Optimization Using Sample Bootstrapping and Particle Swarm Optimization in the Credit Approval Classification. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(1), 18–27. <https://doi.org/10.24014/ijaidm.v2i1.6299>
- Amalia, A., Zaidiah, A., & Isnainiyah, I. N. (2022). Prediksi Kualitas Udara Menggunakan Algoritma K Nearest Neighbor. *JUPI*, 7(2), 496–507. <https://doi.org/10.33387/jiko.v4i2.2871>
- Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, 1(2), 29–33. <https://doi.org/10.33096/ijodas.v1i2.11>
- Batta, M. (2018). Machine Learning Algorithms - A Review. (*IJSR*), 18(8), 381–386. <https://doi.org/10.21275/ART20203995>
- Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap* (1st Editio). Chapman and Hall/CRC. <https://doi.org/https://doi.org/10.1201/9780429246593>
- Elly Pusporani, Qomariyah, S., & Irhamah. (2019). *Klasifikasi Pasien Penderita Penyakit Liver*. 2(March).
- Fikri, M. N. (2020). The Impact of Fuzzy Discretization's Output on Classification Accuracy of Random Forest Classifier. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 3950–3956. <https://doi.org/10.30534/ijatcse/2020/218932020>
- Johar, A., Yanosma, D., & Anggriani, K. (2021). Implementasi Metode K-Nearest Neighbor (Knn) Dan Simple Additive Weighting (Saw) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka (Studi Kasus : Dinas Pemuda dan Olahraga Provinsi Bengkulu). *Pseudocode, Jurnal Nomor, Volume III Kasus, Studi Pemuda, Dinas Bengkulu, Provinsi, III(0065)*, 98–112.
- Kasih, P. (2019). Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara. *INNOVATICS*, 1(2), 63–69. <https://doi.org/10.37058/innovatics.v1i2.918>

- Kusuma, S. R., Hartati, R. S., & Sukerayasa, I. W. (2020). Metode Fuzzy Logic Terhadap Hasil Peramalan Beban Listrik Jangka Panjang. *Jurnal SPEKTRUM*, 7(1), 18–24.
- Li, M., Xu, H., Liu, X., & Lu, S. (2018). Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technology and Health Care*, 26(S1), S509–S519. <https://doi.org/10.3233/THC-174836>
- Mishra, S., Vanli, O. A., Huffer, F. W., & Jung, S. (2016). Regularized discriminant analysis for multi-sensor decision fusion and damage detection with Lamb waves. *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2016*, 9803(March), 98032H. <https://doi.org/10.1117/12.2217959>
- Muslihudin, M., & Oktafianto, D. (2020). Jurnal simada. *Jurnal Sistem Informasi & Manajemen Basis Data (SIMADA)*, 03(01), 68.
- Nikmatur, R. (2017). Proses Penelitian, Masalah, Variabel dan Paradigma Penelitian. *Jurnal Hikmah*, 14(1), 63.
- Nugroho, A. K., & Iskandar, D. (2015). Algoritma Iterative Dichotomizer 3 (ID3) Pengambilan Keputusan Decision Making Using Iterative Dichotomizer 3 Algorithm. *Dinamika Rekayasa*, 11(2), 43–47.
- Nurdiana, N., & Algifari, A. (2020). Studi Komparasi Algoritma Id3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *INFOTECH Journal*, 6(2), 18–23.
- Nurjanah, D. W. (2023). *Rancang Bangun Pendeteksi dan Penetralisir Asap Rokok Dalam Ruangan Menggunakan Sensor Mq-2 Dan Metode Fuzzy Logic. IV*, 7–14.
- 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>
- Purba, L. S. L., & Harefa, N. (2019). Pengaruh Kandungan Oksigen Udara Sekolah terhadap Konsentrasi Belajar Siswa SMA N 9 Jakarta Timur Leony Sanga Lamsari Purba, Nelius Harefa. *Seminar Nasional Pendidikan (Sendika)*, 3(November), 9–16.
- Puspita, E. S., & Yulianti, L. (2016). *Perancangan Sistem Peramalan Cuaca Berbasis Logika Fuzzy. 12*(1).
- Rachman, F. P. A. putra, Goejantoro, R., & Hayati, M. N. (2018). Penentuan Jumlah Replikasi Bootstrap Menggunakan Metode Pretest Pada Independent Sampel T

- Test (Pendapatan Asli Daerah Kabupaten / Kota di Provinsi Kalimantan Timur dan Kalimantan Utara Tahun 2015). *Jurnal Eksponensial*, 9(1), 35–40.
- Rahmawan, H. (2020). Penentuan Rekomendasi Pelatihan Pengembangan Diri Bagi Pegawai Negeri Sipil Menggunakan Algoritma C4.5 Dengan Principal Component Analysis Dan Diskritisasi. *Jurnal Tekno Kompak*, 14(1), 5. <https://doi.org/10.33365/jtk.v14i1.531>
- Rahmayani, M. (2022). *Prediksi Kualitas Udara Menggunakan Metode Ensemble Pada Model Decision Tree Id.3, Random Forest Dan Regresi Logistik Multinomial*.
- Rindengan, A. J., & Langi, Y. A. R. (2019). *Sistem Fuzzy*.
- Setiawan, T. A., & Karomi, M. A. A. (2017). Penerapan Metode Sample Bootstrapping untuk Meningkatkan Performa kNearest Neighbor pada Dataset Berdimensi Tinggi. *Jurnal STMIK IC-Tech*, XII(1), 9–14. <http://jurnal.stmik-wp.ac.id>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Siswanto, E., Suprapedi, & Purwanto. (2018). Metode Sample Bootstrapping pada K-Nearest Neighbor untuk Klasifikasi Status Desa. *Jurnal Teknologi Informasi*, 14(1), 13–23.
- T.P, B. P., & Indah Sari, R. D. (2017). Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (Id3). *Jouticla*, 2(2), 101–108. <https://doi.org/10.30736/jti.v2i2.68>
- Thind, M. P. S., Heath, G., Zhang, Y., & Bhatt, A. (2022). Characterization factors and other air quality impact metrics: Case study for PM2.5-emitting area sources from biofuel feedstock supply. *Science of the Total Environment*, 822, 153418. <https://doi.org/10.1016/j.scitotenv.2022.153418>
- Toha, A., Purwono, P., & Gata, W. (2022). Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV. 4(1), 12–21. <https://doi.org/10.12928/biste.v4i1.6079>
- Waluyo, S. H., & Prihandoko. (2017). Klasifikasi Pemanfaat Program Beras Sejahtera (RASTRA) Berdasarkan Tingkat Kemiskinan Dengan Menggunakan Algoritma Decision Tree C4 . 5 Berbasis Particle Swarm Optimization. *Jurnal Ilmiah Ilmu-Ilmu Teknik*, 7(2), 19–24.

- Wang, X., Zhou, C., Wang, X., Zhou, C., & Xu, X. (2019). *ScienceDirect ScienceDirect Application of C4 . 5 decision tree for scholarship evaluations Application of C4 . 5 decision tree for scholarship evaluations*. 00(2018). <https://doi.org/10.1016/j.procs.2019.04.027>
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1), 134.
- Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for kNN Classification. *ACM Transactions on Intelligent Systems and Technology*, 8(3). <https://doi.org/10.1145/2990508>
- Zhao, Z., Wu, J., Cai, F., Zhang, S., & Wang, Y. G. (2022). A statistical learning framework for spatial-temporal feature selection and application to air quality index forecasting. *Ecological Indicators*, 144(August), 109416. <https://doi.org/10.1016/j.ecolind.2022.109416>