

**DETEKSI ANOMALI DATA IKLIM MIKRO PERKEBUNAN KELAPA
SAWIT DENGAN MENGGUNAKAN ALGORITMA *ISOLATION FOREST***

(Skripsi)

**Oleh
Siti Nafisha Meidina**



**PROGRAM STUDI TEKNIK ELEKTRO
JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
2025**

ABSTRAK

DETEKSI ANOMALI DATA IKLIM MIKRO PERKEBUNAN KELAPA SAWIT DENGAN MENGGUNAKAN ALGORITMA *ISOLATION FOREST*.

Oleh

SITI NAFISHA MEIDINA

Perubahan iklim mikro di kebun kelapa sawit dapat berdampak pada produktivitas dan kesehatan tanaman, sehingga pemantauan serta deteksi awal terhadap anomali cuaca sangatlah penting. Penelitian ini bertujuan untuk mendeteksi anomali pada data iklim mikro dengan menggunakan algoritma *Isolation Forest*. Data yang diperoleh dari sensor cuaca selama periode tujuh bulan (1 Juni–31 Desember 2024) mencakup suhu, kelembapan, curah hujan, kecepatan angin, arah angin, serta intensitas radiasi matahari. Proses pra-pemrosesan mencakup interpolasi linier untuk menangani nilai yang hilang serta pengambilan sampel data dari interval 1 menit ke 30 menit. Model *Isolation Forest* dinilai menggunakan *Mean Absolute Error (MAE)*, *Root Mean Square Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)* yang berbasis *moving average*. Hasil menunjukkan kinerja yang sangat baik pada suhu (MAE=1,77; RMSE=2,41; MAPE=5,53%), cukup baik pada kelembapan (MAE=6,81; RMSE=9,02; MAPE=11,24%) dan curah hujan (MAE=0,003; RMSE=9,02; MAPE=10,68%), serta kurang akurat pada data yang berfluktuasi seperti arah angin (MAE=38,7; RMSE=61,94; MAPE=50,22%), kecepatan angin (MAE=0,101; RMSE=0,188; MAPE=24,88%), dan intensitas radiasi matahari (MAE=225,205; RMSE=264,06; MAPE=37,45%). Ketidakkonsistenan dalam evaluasi pada fitur tertentu mungkin disebabkan oleh: (1) ketidakcocokan *moving average* sebagai *baseline*; (2) ketidakcocokan RMSE untuk model *Isolation Forest*; (3) ketidakcocokan model dengan tipe data tersebut; atau (4) kualitas data yang buruk karena *noise* atau kesalahan sensor. Oleh karena itu, diperlukan penelitian lebih lanjut dengan model deteksi berbeda, pendekatan tambahan atau pra-proses lanjutan untuk meningkatkan kualitas data sebelum proses deteksi anomali dilakukan.

Kata kunci: *Deteksi Anomali, Evaluasi Model, Iklim Mikro, Isolation Forest, Perkebunan Kelapa Sawit.*

ABSTRACT

ANOMALY DETECTION IN OIL PALM PLANTATION MICROCLIMATE DATA USING THE ISOLATION FOREST ALGORITHM

By

SITI NAFISHA MEIDINA

Microclimate changes in oil palm plantations can affect to productivity and plant health, making weather anomaly monitoring and early detection crucial. This study aims to detect anomalies in microclimate data using the Isolation Forest algorithm. The data, collected from weather sensors over a seven-month period (June 1–December 31, 2024), includes temperature, humidity, rainfall, wind speed, wind direction, and solar radiation intensity. The preprocessing stage involved linear interpolation to handle missing values and resampling the data from 1-minute to 30-minute intervals. The Isolation Forest model was evaluated using three metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), with a moving average used as the baseline. The results demonstrated excellent performance for temperature (MAE = 1.77; RMSE = 2.41; MAPE = 5.53%), good performance for humidity (MAE = 6.81; RMSE = 9.02; MAPE = 11.24%) and rainfall (MAE = 0.003; RMSE = 9.02; MAPE = 10.68%), but lower accuracy for more fluctuating features such as wind direction (MAE = 38.7; RMSE = 61.94; MAPE = 50.22%), wind speed (MAE = 0.101; RMSE = 0.188; MAPE = 24.88%), and pyrano (MAE = 225.205; RMSE = 264.06; MAPE = 37.45%). The inconsistency in evaluation results for certain features may be due to: (1) the moving average basis used is not compatible with the type of wind speed and solar intensity data; (2) the evaluation method used is not suitable with the Isolation Forest model; (3) the Isolation Forest model for anomaly detection is not proper with the type of data for these two parameters; (4) the quality of wind speed and solar intensity data is not well mannered due to noise or sensor reading errors. Therefore, further research is needed with different detection models, additional approaches, or further preprocessing are needed to improve data quality prior to anomaly detection.

Keywords: *Anomaly Detection, Isolation Forest, Model Evaluation, Microclimate, Oil Palm Plantation.*

**DETEKSI ANOMALI DATA IKLIM MIKRO PERKEBUNAN KELAPA
SAWIT DENGAN MENGGUNAKAN ALGORITMA *ISOLATION FOREST***

Oleh

Siti Nafisha Meidina

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA TEKNIK**

Pada

**Jurusan Teknik Elektro
Fakultas Teknik Universitas Lampung**



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2025

Judul Skripsi

: **DETEKSI ANOMALI DATA IKLIM MIKRO
PERKEBUNAN KELAPA SAWIT DENGAN
MENGGUNAKAN ALGORITMA *ISOLATION*
*FOREST***

Nama Mahasiswa

: **Siti Nafisha Meidina**

Nomor Pokok Mahasiswa : 1815031047

Jurusan

: Teknik Elektro

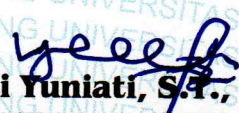
Fakultas

: Teknik



Pembimbing Utama

Pembimbing Pendamping


Yetti Yuniati, S.T., M.T.

NIP 19800113 200912 2 002


Dr. Ing. Ardian Ulvan, S.T., M.Sc.

NIP 19731128 199903 1 005

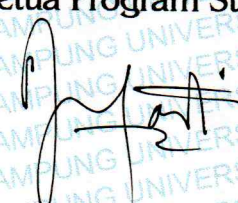
2. Mengetahui

Ketua Jurusan Teknik Elektro

Ketua Program Studi Teknik Elektro


Herlinawati, S.T., M.T.

NIP 19710314 199903 2 001


Sumadi, S.T., M.T.

NIP 19731104 200003 1 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Yetti Yuniati, S.T., M.T.**

Sekretaris : **Dr. Ing. Ardian Ulvan, S.T., M.Sc.**

Penguji : **Ing. Melvi, S.T., M.T., Ph.D.**

2. Dekan Fakultas Teknik

Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc.

NIP 19750928 200112 1 002

Tanggal Lulus Ujian Skripsi : **09 Mei 2025**

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi yang saya buat dengan judul “Deteksi Anomali Data Iklim Mikro Perkebunan Kelapa Sawit Dengan Menggunakan Algoritma *Isolation Forest*” dibuat tidak berdasarkan karya yang pernah dilakukan oleh orang lain. Bahwa karya ini tidak terdapat karya lain atau pendapat yang ditulis atau diterbitkan orang lain, kecuali tertulis diacu dalam naskah ini sebagaimana yang disebutkan di dalam Daftar Pustaka. Selain itu saya menyatakan pula bahwa skripsi ini dibuat oleh saya sendiri.

Apabila pernyataan saya tidak benar, maka saya bersedia dikenai sanksi sesuai dengan hukum yang berlaku.

Bandar Lampung, 06 Juni 2025



Siti Nafisha Meidina

NPM. 1815031047

RIWAYAT HIDUP



Penulis lahir di Bandar Lampung, Lampung, pada tanggal 09 Mei 2000 sebagai anak kedua dari 3 bersaudara, anak dari bapak Wawan Hermawan dan ibu Sri Sulastri. Pendidikan Sekolah Dasar penulis diselesaikan di SDN 1 Kupang Teba pada tahun 2012, Sekolah Menengah Pertama di SMPN 16 Bandar Lampung pada tahun 2015 dan Sekolah Menengah Atas di SMAN 4 Bandar Lampung pada tahun 2018.

Penulis terdaftar sebagai mahasiswa Jurusan Teknik Elektro Universitas Lampung pada tahun 2018 melalui jalur PMPAP. Selama menjadi mahasiswa penulis aktif mengikuti organisasi Himpunan Mahasiswa Teknik Elektro sebagai Anggota Divisi Sosial dan Kewirausahaan selama dua periode kepengurusan. Penulis mengambil konsentrasi Telekomunikasi dan Teknik Informasi pada semester 5 dan secara aktif mengikuti kegiatan penelitian di Laboratorium Telekomunikasi. Penulis juga merupakan asisten Laboratorium Pengukuran Besaran Listrik pada semester 5. Selama menjadi asisten Laboratorium Pengukuran Besaran Listrik penulis menjadi asisten praktikum pada mata kuliah Fisika Dasar Teknik Informatika semester Genap TA 2020/2021, mata kuliah Pengukuran Besaran Listrik semester Genap TA 2020/2021 dan semester Ganjil TA 2021/2022, dan mata kuliah Rangkaian Listrik semester Ganjil TA 2021/2022.

PERSEMBAHAN



Dengan segala syukur bagi Allah SWT atas berkah Rahmat dan karunia-Nya,
Kupersembahkan Karya ini dengan rasa Syukur, hormat dan kasih sayang:

Kepada Bapak dan Ibu

Atas semua dukungan, kepercayaan, pengorbanan dan doa selama ini kepada Nafisha sehingga Nafisha selalu kuat dan bisa selalu dalam lindungan Allah SWT selama menempuh perkuliahan.

Kepada Saudara Tercinta,

Siti Ulfa Nabila, Wahyu Hidayat, dan Siti Fatia Ramadhanti
Yang telah memberi motivasi, dan menjadi teman yang selalu mengarahkan serta mendukung penulis dalam pengerjaan skripsi ini.

Kepada Dosen Pembimbing dan Penguji

Atas kesempatan, ilmu serta kesabaran yang telah diberikan kepada Nafisha selama perkuliahan. Terimakasih atas bimbingan, motivasi dan inspirasi kepada Nafisha.

Kepada Sahabat

As'ad Humam dan Risa Zannati yang selalu jadi orang yang memotivasi penulis dan menjadi tempat berkeluh kesah. Terimakasih atas saran dan dukungan yang telah diberikan.

MOTTO HIDUP

“Hadapi rasa takut itu, karena rasa takut itu datang dari ketidaktahuan”

SANWACANA

Alhamdulillah segala puji bagi Allah subhanahu wa ta'ala atas limpahan rahmat dan karunia-Nya lah penulis dapat menyelesaikan skripsi yang berjudul “Deteksi Anomali Data Iklim Mikro Perkebunan Kelapa Sawit Dengan Menggunakan Algoritma *Isolation Forest*” sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Jurusan Teknik Elektro Fakultas Teknik Universitas Lampung. Shalawat serta salam senantiasa penulis haturkan kepada Nabi Muhammad SAW, semoga kita termasuk golongan yang mendapatkan syafaatnya di Yaumul Qiyamah kelak, Aamiin.

Penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, kritk, saran, dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini sebagai wujud rasa hormat, penulis mengucapkan banyak terima kasih kepada:

1. Bapak dan ibu tercinta serta seluruh keluarga penulis yang tiada hentinya mendoakan serta memberikan motivasi, dorongan semangat dan juga materi.
2. Bapak Dr. Eng. Helmy Fitriawan, S.T., M.Sc., selaku Dekan Fakultas Teknik Universitas Lampung;
3. Ibu Herlinawati, S.T., M.T. selaku Ketua Jurusan Teknik Elektro Universitas Lampung;
4. Ibu Yetti Yuniati, S.T., M.T. selaku Dosen pembimbing Utama juga selaku Dosen Pembimbing Akademik yang telah memberikan ilmu, bimbingan, bantuan, masukan, dan pandangan serta berbagai pola pikir kepada penulis di setiap kesempatan dengan baik dan menarik;
5. Bapak Dr. Ing. Ardian Ulvan, S.T., M.Sc. selaku Dosen Pembimbing Pendamping yang telah memberikan ilmu, bimbingan, masukan, motivasi, dan sudut pandang lain serta berbagai pola pikir kepada penulis disetiap kesempatan dengan baik dan menarik;
6. Ibu Ing. Melvi, S.T., M.T., Ph.d. selaku Dosen Penguji yang telah memberikan kritik dan saran serta penyelesaian masalah yang membangun kepada penulis;

7. Segenap dosen di Jurusan Teknik Elektro yang telah memberikan ilmu yang bermanfaat kepada penulis selama menempuh pendidikan perkuliahan;
8. Segenap staff di Jurusan Teknik Elektro yang telah membantu penulis baik dalam hal administrasi dan lain-lain;
9. As'ad Humam dan Risa Zannati yang selalu jadi orang yang memotivasi penulis dan menjadi tempat berkeluh kesah;
10. Seluruh pihak yang tidak dapat disebutkan satu per satu yang telah membantu penulis.

Penulis mengucapkan banyak terima kasih kepada semua pihak yang telah membantu penulis. Semoga Allah subhanahu wa ta'ala membalas segala bentuk kebaikan yang telah Bapak, Ibu, dan rekan-rekan semua berikan. Penulis berharap semoga skripsi ini dapat bermanfaat untuk kita semua, Aamiin.

Bandar Lampung, 06 Juni 2025

Penulis,

Siti Nafisha Meidina

NPM. 1815031047

DAFTAR ISI

	Halaman
DAFTAR ISI	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Tujuan Skripsi.....	2
1.3. Manfaat Skripsi.....	3
1.4. Rumusan Masalah.....	3
1.5. Batasan Masalah	3
1.6. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Pustaka pada Penelitian yang Berkaitan	5
2.2 <i>Exploratory Data Analysis (EDA)</i>	15
2.3 <i>Data Cleaning dan Penanganan Missing value</i>	16
2.4 <i>Interpolasi Data</i>	17
2.5 <i>Anomaly Detection (AD)</i>	18
2.6 <i>Machine learning</i>	20
2.7 <i>Isolation forest</i>	24
2.8 <i>Evaluation Metrics</i>	28
2.8.1 <i>Mean Absolute Error (MAE)</i>	28

2.8.2	<i>Root Mean Squared Error (RMSE)</i>	29
2.8.3	<i>Mean Absolute Percentage Error (MAPE)</i>	30
BAB III METODOLOGI PENELITIAN.....		31
3.1.	Waktu dan Tempat Penelitian.....	31
3.2.	Alat dan Bahan	31
3.3.	Metode Penelitian	31
3.4.	Diagram Alir Pengerjaan Penelitian	34
3.5.	Indikator Keberhasilan.....	42
BAB IV HASIL DAN PEMBAHASAN		43
4.1.	<i>Capture Data</i>	43
4.2.	<i>Exploratory Data Analysis (EDA)</i>	44
4.3.	<i>Pre-Processing Data</i>	49
4.3.1.	Penggabungan Dataset	49
4.3.2.	Pembersihan dan Pemeriksaan <i>Missing Value</i>	49
4.3.3.	Penanganan <i>Missing Value</i>	50
4.3.4.	<i>Resampling Data</i>	51
4.4.	Pemilihan Model Algoritma	52
4.5.	Deteksi Anomali dan Evaluasi.....	53
4.5.1.	Hasil Deteksi Anomali	53
4.5.2.	Evaluasi Hasil Deteksi Anomali	57
BAB V KESIMPULAN DAN SARAN.....		58
DAFTAR PUSTAKA		60

DAFTAR GAMBAR

Gambar 2. 1 <i>Roadmap</i> Penelitian Terkait.....	15
Gambar 2. 2 Cara Kerja <i>Machine learning</i>	21
Gambar 2. 3 Kategori <i>Machine learning</i> dan Algoritma	22
Gambar 2. 4 Cara Kerja <i>Isolation Forest</i>	25
Gambar 3. 1 Diagram Alir Penelitian.....	34
Gambar 3. 2 Tahap Identifikasi.....	35
Gambar 3. 3 Tahap Akuisisi Data	36
Gambar 3. 4 Tahap <i>Pre-Processing Data</i>	38
Gambar 3. 5 Tahap Penyesuaian Algoritma <i>Isolation Forest</i>	39
Gambar 3. 6 Tahap Identifikasi Anomali.....	40
Gambar 3. 7 Tahap Penulisan Laporan Akhir.....	41
Gambar 4. 1 Visualisasi Hasil EDA 2 Atribut (Data Kecepatan dan Arah Angin)	47
Gambar 4. 2 Visualisasi Hasil EDA 5 Atribut (<i>Temperature, Humidity, Rainfall,</i> <i>dan Pyrano</i>).....	48
Gambar 4. 3 Cuplikan <i>Source Code</i> Model <i>Isolation Forest</i>	53
Gambar 4. 4 Hasil Deteksi Anomali Gabungan 6 Atribut (<i>Temperature, Humidity,</i> <i>Rainfall, Wind_Speed</i> dan <i>Pyrano</i>).....	54
Gambar 4. 5 Hasil Deteksi Anomali Gabungan 2 Atribut (<i>Wind_Speed</i> dan <i>Direction</i>).....	56

DAFTAR TABEL

Tabel 2. 1. Penelitian Terkait	10
Tabel 4. 1 <i>Capture Data</i> Iklim Mikro Bulan Juni 2024 – Desember 2024.....	43
Tabel 4. 2 <i>Exploratory Data Analysis (EDA)</i> Data Gabungan	44
Tabel 4. 3 <i>Exploratory Data Analysis (EDA)</i> Data Tiap Bulan (<i>Missing Values</i> dan % <i>Missing Values</i>).....	45
Tabel 4. 4 <i>Exploratory Data Analysis (EDA)</i> Data Tiap Bulan (<i>Outliers</i> dan % <i>Outliers</i>)	46
Tabel 4. 5 Evaluasi Hasil Deteksi Anomali	57

BAB I

PENDAHULUAN

1.1. Latar Belakang

Seiring perkembangan zaman, kecenderungan konsumsi minyak nabati terus meningkat. Minyak sawit merupakan salah satu produk unggulan dan pemberi kontribusi paling besar pada perekonomian Indonesia. Ekspor minyak sawit menjadi salah satu sektor inti dalam perekonomian Indonesia dengan nilai ekspor mencapai 24,3 juta ton atau USD 16,2 miliar pada tahun 2016 [1]. Indonesia menjadi salah satu penguasa pasar minyak nabati dikarenakan menjadi salah satu negara yang mampu menghasilkan minyak sawit terbesar di dunia yang mencapai 54% konsumsi global [2].

Program perkebunan kelapa sawit juga menjadi salah satu program yang berhasil meningkatkan pertumbuhan ekonomi, kesejahteraan dan pemberdayaan masyarakat di pedesaan [3]. Pengembalian nilai ekonomi yang tinggi, pengelolaan yang mudah (sistem pertanian yang cukup dikontrol oleh kelompok masyarakat), kemudahan dalam pemasaran, dukungan perusahaan kelapa sawit yang ada di sekitar desa, dan pohon kelapa sawit dapat tumbuh ditanah marjinal seperti lahan gambut menjadi faktor pendorongnya [4].

Oleh karena pentingnya sektor perkebunan kelapa sawit di Indonesia, kestabilan lingkungan pertumbuhan pohon sawit juga menjadi salah satu hal yang harus diperhatikan. Sehingga pemantauan lingkungan perkebunan kelapa sawit harus dilakukan secara berkala. Salah satu bentuk hasil pemantauan perkebunan kelapa sawit dapat disajikan dalam bentuk dataset yang disebut dengan data iklim mikro.

Data iklim mikro menggambarkan kondisi lingkungan yang spesifik di sekitar suatu lokasi. Dalam arsitektur dan desain lingkungan, data ini sangat penting untuk

memantau kenyamanan dan kualitas lingkungan. Deteksi anomali pada data iklim mikro dapat membantu mengidentifikasi perubahan lingkungan yang tidak biasa dan memantau kenyamanan di suatu lokasi [5].

Data iklim mikro hanya menyajikan data mentah berupa kondisi langsung pada suatu waktu di perkebunan kelapa sawit tersebut. Pengelolaan perkebunan kelapa sawit yang baik membutuhkan pengukuran dan prediksi jangka panjang untuk mengetahui kondisi lingkungan hidup pohon-pohon sawit secara *real-time*, sehingga dibutuhkan suatu perangkat ukur untuk mengakuisisi data, sekaligus metode analisis yang tepat yang dapat digunakan untuk memantau data iklim mikro tersebut sehingga perencanaan dan pengelolaan perkebunan kelapa sawit akan lebih baik lagi. Oleh karena itu, penulis melakukan penelitian untuk mendeteksi anomali pada data iklim mikro di area perkebunan kelapa sawit yang bertujuan untuk menemukan data yang tidak normal.

Deteksi anomali adalah proses yang memeriksa titik data tertentu dan mendeteksi kejadian langka yang tampak mencurigakan karena berbeda dengan pola perilaku yang ada. Dalam beberapa tahun terakhir, deteksi anomali telah bergantung lebih banyak pada *machine learning (ML)* untuk mengidentifikasi *outlier* yang sulit diidentifikasi dan melindungi sistem dari bahaya [6].

Penelitian ini bertujuan untuk mendeteksi anomali dari data iklim mikro yang diambil dari penelitian di perkebunan kelapa sawit dengan memanfaatkan teknologi *Machine Learning* menggunakan logaritma *Isolation Forest*.

1.2. Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan untuk mengembangkan sistem deteksi anomali yang dapat mendeteksi kejadian langka (anomali) pada data iklim mikro menggunakan algoritma *Isolation Forest*. Algoritma ini diharapkan dapat mengidentifikasi pola yang tidak biasa dengan tingkat akurasi tinggi dan waktu respons yang cepat.

1.3. Manfaat Penelitian

Adapun manfaat yang diharapkan dapat dihasilkan pada penelitian ini adalah dapat membantu memantau kualitas lingkungan dengan mengidentifikasi perubahan lingkungan yang tidak biasa, sehingga memungkinkan tindakan yang cepat dan tepat untuk menjaga kualitas lingkungan.

1.4. Rumusan Masalah

Adapun permasalahan yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana menggunakan algoritma *Isolation forest* untuk deteksi anomali dari dataset parameter iklim mikro?
2. Bagaimana melakukan evaluasi untuk menguji apakah model *Isolation Forest* dapat digunakan dan sesuai untuk mendeteksi anomali pada data iklim mikro?

1.5. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya berfokus pada deteksi anomali dengan dataset hasil pengukuran stasiun iklim mikro milik Peneliti Universitas Lampung yang ditempatkan di lahan perkebunan kelapa sawit PT. Persada Dinamika Lestari di Kabupaten Hulu Sungai Utara, Provinsi Kalimantan Selatan.
2. Dataset yang digunakan adalah data suhu dan kelembaban udara, curah hujan, kecepatan dan arah angin, dan intensitas sinar matahari yang diakuisis mulai tanggal 1 Juni – 31 Desember 2024.
3. Data yang dianalisis adalah dataset yang sudah di *resampling* dari interval 1 menit menjadi interval 30 menit.
4. Studi ini hanya membahas deteksi anomali dengan menggunakan algoritma *Isolation forest*.

5. Evaluasi model *Isolation forest* hanya menggunakan *Mean Absolute Error (MAE)*, *Mean Absolute Percentage Error (MAPE)* dan *Root Mean Square Error (RMSE)*

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini terdiri dari lima bab yaitu sebagai berikut:

BAB I. PENDAHULUAN

Bab I Pendahuluan berisi tentang latar belakang masalah, tujuan penelitian, rumusan masalah, batasan masalah, manfaat penelitian, hipotesis dan sistematika penulisan.

BAB II. TINJAUAN PUSTAKA

Bab II tinjauan pustaka berisi mengenai teori yang berkaitan dan yang digunakan pada pengerjaan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab III metodologi penelitian berisi mengenai informasi berkenaan tentang waktu dan tempat penelitian, peralatan yang digunakan, tahapan penelitian, skenario penelitian dan hasil yang diharapkan.

BAB IV. HASIL DAN PEMBAHASAN

Bab IV hasil dan pembahasan berisi tentang hasil dan pembahasan dari penelitian yang dilakukan.

BAB V. KESIMPULAN DAN SARAN

Bab V kesimpulan dan saran berisi mengenai kesimpulan dan saran setelah penulis selesai melakukan penelitian ini yang berdasarkan dari hasil dan pembahasan yang telah didapat.

DAFTAR PUSTAKA

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Pustaka pada Penelitian yang Berkaitan

Penelitian ini mengacu pada penelitian-penelitian yang telah dilakukan sebelumnya, yaitu melakukan studi literatur tentang deteksi anomali dengan menggunakan model algoritma *Isolation forest*.

Penelitian [7] yang berjudul “*Anomaly Detection in Palm Oil Yield Using Machine Learning*” merupakan suatu penelitian yang membahas tentang penerapan teknik deteksi anomali yang memprediksi hasil panen minyak sawit menggunakan metode pembelajaran mesin. Penelitian ini memiliki fokus utama berupa mengidentifikasi pola yang tidak biasa dalam data hasil panen yang dapat disebabkan oleh beberapa faktor, termasuk dari perubahan iklim, pola cuaca dan kondisi tanah. Penelitian ini menggunakan metode pembelajaran mesin dengan dua model yaitu *Extra Tree (Extremely Randomized Tree) Regressor* and *AdaBoost (Adaptive Boosting) Regressor*. Kesimpulan yang didapat dari penelitian ini adalah model pembelajaran mesin dapat secara efektif mendeteksi anomali dalam hasil panen minyak sawit. Faktor-faktor seperti frekuensi curah hujan, kelembaban tanah, dan suhu memiliki pengaruh signifikan terhadap hasil panen. Model yang digunakan menunjukkan kemampuan untuk memberikan prediksi yang lebih baik dibandingkan dengan metode konvensional.

Penelitian [8] selanjutnya yang berjudul “Penerapan Metode *K-Means Clustering* untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri” merupakan suatu penelitian yang membahas penerapan metode *K-Means Clustering* dalam pengelompokan lahan sawit produktif di PT Kasih Agro Mandiri. Dengan menggunakan metode *K-Means* data lahan sawit akan dibagi menjadi beberapa kelompok berdasarkan karakteristik tertentu. Variabel yang digunakan

dalam penelitian ini adalah luas lahan, persentase pencapaian hasil produksi berdasarkan target, dan rata rata produksi per hektar. Hasil dari penelitian ini adalah metode *K-Means* dapat digunakan untuk membantu mengelompokkan lahan sawit berdasarkan tingkat produktivitasnya. Penerapan metode *K-Means Clustering* memberikan manfaat signifikan bagi PT Kasih Agro Mandiri dalam pengelolaan lahan sawit. Dengan aplikasi ini, perusahaan dapat lebih efisien dalam mengambil keputusan terkait manajemen lahan dan meningkatkan hasil produksi kelapa sawit. Penelitian ini menunjukkan potensi penggunaan teknik data *mining* dalam sektor pertanian untuk optimasi hasil produksi. Hasil pengelompokkan dengan metode tersebut menunjukkan adanya tiga *cluster* yaitu produktivitas tinggi 26 blok lahan, produktivitas sedang 25 blok lahan, dan produktivitas rendah ada 39 blok lahan. Pengaplikasian metode *K-Means* menunjukkan tingkat akurasi 100% dalam pengelompokkan data.

Pada penelitian [9] yang berjudul “*Anomaly detection in Network Traffic Using Unsupervised Machine learning Approach*” merupakan penelitian tentang penggunaan metode pembelajaran mesin yang tidak terkendali (*unsupervised*) untuk mendeteksi anomali dalam jaringan. Penelitian ini bertujuan untuk mengembangkan mekanisme inspeksi jaringan yang bersih dan menyeluruh untuk mendeteksi anomali dalam jaringan yang dapat digunakan untuk mengidentifikasi ancaman atau serangan jaringan lebih awal. Penelitian ini menggunakan tiga metode pembelajaran mesin yang tidak terkendali untuk mengidentifikasi anomali dalam jaringan: *Principal Component Analysis (PCA)*, *Autoencoder*, dan *Isolation forest*. Hasil penelitian ini menunjukkan bahwa ketiga metode tersebut dapat efektif dalam mendeteksi anomali dalam jaringan. Meskipun *PCA* dan *Autoencoder* kurang mampu untuk mendeteksi kelompok anomali, metode *Isolation forest* berhasil mengidentifikasi paket anomali dengan memanfaatkan struktur pohon untuk mengisolasi data yang menyimpang secara efisien. Penelitian ini menunjukkan bahwa pendekatan pembelajaran mesin yang tidak terkendali dapat digunakan untuk mendeteksi anomali dalam jaringan. Metode-metode ini memiliki potensi untuk mengidentifikasi ancaman atau serangan jaringan lebih awal, yang dapat membantu dalam mengelola jaringan dan melindungi sistem dari ancaman lebih awal.

Penelitian [10] dengan judul “*Detecting Anomalous Packets In Network Transfers: Investigations Using PCA, Autoencoder And Isolation Forest In TCP*” memiliki bahasan tentang penyelidikan mengenai metode ekstraksi fitur tanpa pengawasan untuk mengidentifikasi anomali jaringan dalam transfer TCP. Penelitian ini mencoba untuk mengidentifikasi paket-paket yang aneh dengan memecah jejak TCP menjadi tiga kelas awal. Metode-metode lain yang digunakan untuk mendeteksi anomali jaringan juga disebutkan, seperti metode berbasis formula dan metode berbasis sejarah. Dalam artikel ini, penulis mengeksplorasi pendekatan-pendekatan untuk menganalisis transfer jaringan yang aneh dengan menggunakan dataset yang dikumpulkan menggunakan Tstat, sebuah alat untuk mengumpulkan jejak TCP untuk transfer. Selain itu, artikel ini juga membahas tentang pendekatan statistik dan pembelajaran mesin untuk mengekstraksi anomali yang tidak diketahui atau tanpa pengawasan dengan tujuan untuk mengidentifikasi anomali seperti kehilangan paket, duplikasi, dan urutan pengiriman ulang yang mempengaruhi kinerja transfer file. Dengan menggunakan PCA dapat menemukan kelompok transfer normal dan abnormal yang jelas, tetapi hal ini tidak terlihat pada eksperimen alur kerja 1000 Genome. Alternatifnya, *autoencoder* gagal total dalam mengekstraksi *cluster* pada transfer normal atau abnormal di kedua eksperimen. Metode *Isolation Forest*, yang dilatih dengan transfer normal, mampu memberikan hasil yang lebih baik, namun terkadang tidak berhasil.

Pada penelitian [11] dengan judul “*Anomaly Detection And Critical Attributes Identification For Products With Multiple Operating Conditions Based On Isolation Forest*” adalah penelitian tentang penggunaan model *Isolation forest* untuk mendeteksi anomali dan mengidentifikasi atribut kritis yang terkait dengan kinerja produk yang memiliki berbagai kondisi operasi. Hasil penelitian ini menunjukkan bahwa metode *Isolation forest* efektif dalam mendeteksi anomali dan mengidentifikasi atribut kritis yang terkait dengan kinerja produk yang memiliki berbagai kondisi operasi. Penelitian ini menunjukkan bahwa metode *Isolation forest* dapat digunakan untuk mendeteksi anomali dan mengidentifikasi atribut kritis (*Active Power*, *Generator Stator Temperature*, *Generator Speed*, dan *generator Bearing 2 Temperature*) yang terkait dengan kinerja produk yang memiliki berbagai kondisi operasi yaitu *start-up phase*, *maximum wind energy*

tracking phase, dan *nominal power output phase*. Metode ini memiliki potensi untuk mendukung pengambilan keputusan dan pengelolaan produk dalam lingkungan industri.

Penelitian [12] dengan judul “*Anomaly Detection Of Power Grid Dispatching Platform Based On Isolation Forest And K-Means Fusion Algorithm*” membahas mengenai penggunaan algoritma *Isolation forest* dan algoritma *K-Means fusion* untuk mendeteksi anomali pada platform pengiriman listrik. Metode ini diusulkan untuk mengatasi keterbatasan *Isolation forest* yang hanya mampu mendeteksi anomali dalam data biner. Penelitian ini menekankan pentingnya mendeteksi dengan akurat dan cepat IP akses yang tidak normal dari platform pengiriman listrik guna memastikan keamanan produksi listrik. Fokus utama penelitian ini adalah pada deteksi anomali dalam transmisi dan akses data pada sistem pengendalian distribusi listrik. Metode ini diharapkan dapat memberikan kontribusi dalam meningkatkan keandalan dan keamanan sistem tenaga listrik. Dalam makalah ini, berdasarkan platform pengiriman *China Southern Power Grid*, diusulkan algoritma yang lebih baik untuk mendeteksi akses abnormal daripada *Isolation Forest*. Pada saat yang sama, sebagai titik awal, strategi pengaturan ambang batas sampel anomali dioptimalkan sehingga tidak bergantung pada pengalaman manusia, sehingga meningkatkan akurasi dan stabilitas algoritma deteksi anomali.

Pada penelitian [13] yang berjudul “*Sensors Anomaly Detection of Industrial Internet of Things Based on Isolated Forest Algorithm and Data Compression*” membahas mengenai penggunaan algoritma *Isolation forest* dan kompresi data untuk mendeteksi anomali pada sensor dalam lingkungan *Internet of Things (IoT)* industri. Penelitian ini bertujuan untuk mengembangkan metode yang efektif untuk mendeteksi anomali pada sensor dalam lingkungan IoT industri. Algoritma *Isolation forest* digunakan untuk mengidentifikasi anomali pada data sensor, sedangkan kompresi data digunakan untuk mengurangi ukuran data sensor dan meningkatkan efisiensi pemrosesan. Hasil penelitian ini menunjukkan bahwa metode yang diusulkan dapat efektif dalam mendeteksi anomali pada sensor dalam lingkungan IoT industri. Metode ini dapat menghasilkan hasil yang akurat dan efisien dalam mendeteksi anomali pada data sensor. Metode ini memiliki potensi untuk meningkatkan efisiensi dan keandalan sistem IoT industri dan dapat

membantu dalam mengidentifikasi masalah pada sensor sebelum terjadi kerusakan yang lebih serius.

Penelitian [14] dengan judul “*Perbandingan Metode One Class SVM dan Isolation forest Dalam Mendeteksi Anomali Dalam Activity Recognition Pada Rumah Dengan PIR Sensor*” membahas mengenai pendeteksian *activity recognition* pada rumah cukup penting untuk mengetahui adanya gerakan anomali. Algoritma One Class SVM & *Isolation forest* adalah algoritma yang dapat diterapkan untuk pendeteksian anomali, namun kedua algoritma tersebut memiliki kekurangan dan kelebihan masing masing. Tujuan dari penelitian ini adalah membandingkan algoritma *One Class SVM* dan *Isolation forest* untuk mencari hasil yang paling baik dari segi akurasi, *sensitivity*, dan *specificity* untuk membedakan gerakan anomali atau bukan dari penghuni rumah. Sensor PIR digunakan pada proses pengambilan dataset, dataset yang sudah didapat akan diolah secara manual terlebih dahulu, setelah dataset diolah maka akan diproses oleh algoritma *One Class SVM* dan *Isolation forest*. Hasil dari pengujian penelitian ini algoritma *One Class SVM* memiliki nilai Akurasi sebesar 96%, *Sensitivity* 100%, dan *Specificity* 98,2%, sedangkan algoritma *Isolation forest* memiliki nilai Akurasi sebesar 91%, *Sensitivity* 40%, dan *Specificity* 91,4%.

Penelitian [15] dengan judul “*Deteksi Anomali Menggunakan Extended Isolation Forest (EIF)*” membahas tentang penggunaan algoritma *Extended Isolation Forest (EIF)* untuk mendeteksi anomali pada data. EIF merupakan perpanjangan dari algoritma *Isolation forest* (IF) yang menghilangkan bias yang berasal dari *branching*. Penelitian ini mengimplementasikan EIF ke dalam platform sumber terbuka ML H2O-3 dan menunjukkan hasil pengujian yang menunjukkan bahwa EIF dapat mendeteksi anomali dengan baik. Hasil pengujian juga menunjukkan bahwa model EIF perlu disesuaikan untuk mendeteksi struktur data yang lebih kompleks. Penelitian ini memberikan wawasan tentang penggunaan EIF dalam deteksi anomali dan dapat menjadi referensi bagi para peneliti dan praktisi yang tertarik dalam deteksi anomali data. Hasil pengujian menunjukkan bahwa *Extended Isolation Model* perlu disesuaikan. Tes kinerja deteksi anomali mengungkapkan sedikit ketidaksempurnaan dalam deteksi struktur data jika dibandingkan dengan satu-satunya implementasi algoritma Python yang tersedia. Masalah ini dapat

diperbaiki dengan *subsampling* yang lebih baik atau di percabangan *iTree*, di mana algoritma bisa menghindari titik perpecahan dengan data kosong.

Penelitian [16] dengan judul “*Real-Time Malicious Traffic Detection With Online Isolation forest Over SD-WAN*” membahas mengenai penggunaan algoritma *Isolation forest* dalam mendeteksi lalu lintas yang bermasalah (*malicious*) di atas jaringan SD-WAN (*Software-Defined Wide Area Network*). Metode yang digunakan pada penelitian ini adalah algoritma *Isolation forest* untuk mendeteksi lalu lintas yang bermasalah secara *real-time*. Algoritma ini memiliki kemampuan untuk menghasilkan hasil yang akurat dan cepat dalam mendeteksi anomali. Algoritma yang diusulkan, yang disebut OADSD (*Online Anomali Detection in SD-WAN*), dapat secara akurat mendeteksi berbagai jenis serangan dengan kinerja tinggi. Eksperimen penelitian ini pada dataset publik menunjukkan bahwa OADSD dapat secara akurat mendeteksi berbagai jenis serangan dengan kinerja tinggi. Penelitian ini menunjukkan bahwa algoritma *Isolation forest* dapat digunakan untuk mendeteksi lalu lintas yang bermasalah secara *real-time* di atas jaringan SD-WAN. Metode ini memiliki potensi untuk membantu dalam mengelola jaringan dan melindungi sistem dari serangan berbahaya.

Penelitian terdahulu yang berkaitan pada penelitian yang dilakukan oleh penulis dirangkum dalam Tabel 2.1 dan Gambar 2.1 berikut.

Tabel 2. 1. Penelitian Terkait

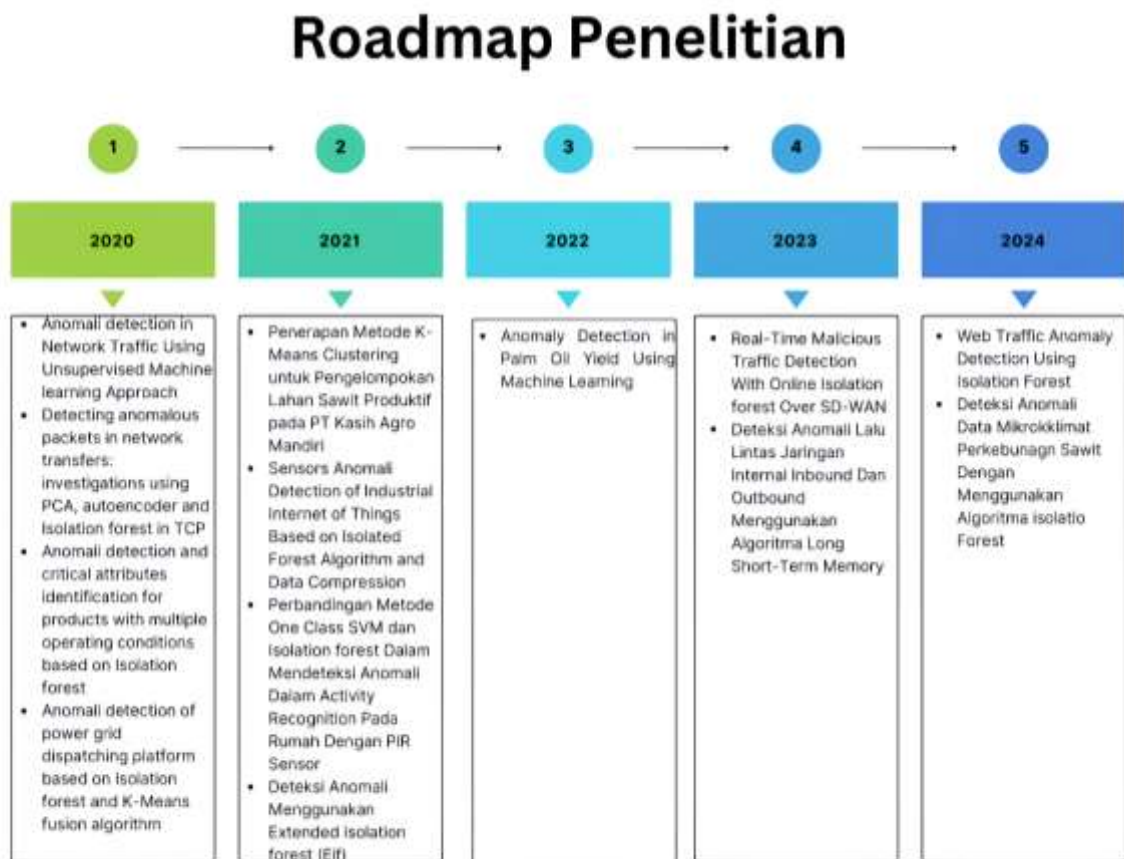
No	Judul Jurnal	Penulis & Tahun	Metode	Hasil
1	<i>Anomaly detection in Network Traffic Using Unsupervised Machine learning Approach</i>	Aditya Vikram, Mohana (2020)	Analisis Komponen <i>Principal (PCA)</i> , <i>Autoencoder</i> , dan <i>Isolation forest</i>	Ketiga metode dapat efektif dalam mendeteksi anomali dalam jaringan. Meskipun PCA dan <i>Autoencoder</i> sulit menemukan sulit mendeteksi kelompok, <i>Isolation forest</i> dapat mengidentifikasi paket anomali dengan memecah jaringan TCP menjadi kelas kayu dengan benar.

No	Judul Jurnal	Penulis & Tahun	Metode	Hasil
2	<i>Detecting anomalous packets in network transfers: investigations using PCA, autoencoder and Isolation forest in TCP</i>	Mariam Kiran, Cong Wang, George Papadimitriou, Anirban Mandal, Ewa Deelman (2020)	Analisis Komponen Principal (PCA), <i>Autoencoder</i> , dan <i>Isolation forest</i> .	Dengan menggunakan PCA dapat menemukan kelompok transfer normal dan abnormal yang jelas, tetapi hal ini tidak terlihat pada eksperimen alur kerja 1000 Genome. Alternatifnya, <i>autoencoder</i> gagal total dalam mengekstraksi cluster pada transfer normal atau abnormal di kedua eksperimen. Hutan isolasi, yang dilatih dengan transfer normal, mampu memberikan hasil yang lebih baik, namun terkadang tidak berhasil
3	<i>Anomaly Detection And Critical Attributes Identification For Products With Multiple Operating Conditions Based On Isolation Forest</i>	Hansi Chen, Hongzhan Ma, Xuening Chu, Deyi Xue (2020)	<i>Isolation forest</i>	Metode <i>Isolation forest</i> dapat digunakan untuk mendeteksi anomali dan mengidentifikasi atribut kritis yang terkait dengan kinerja produk yang memiliki berbagai kondisi operasi. Metode ini memiliki potensi untuk mendukung pengambilan keputusan dan pengelolaan produk dalam lingkungan industri.
4	<i>Anomaly Detection Of Power Grid Dispatching Platform Based On Isolation Forest And K-Means Fusion Algorithm</i>	Liang Shouyu, Zhang Kun, Fang Wenchong, Zhou Zhifeng, Hu Rong, Zhu Wen, Li Yingchen, Wang Yichang, Hou Jian (2020)	<i>Isolation forest and K-Means fusion algorithm</i>	Berdasarkan platform pengiriman <i>China Southern Power Grid</i> , berdasarkan <i>Isolation Forest</i> , diusulkan algoritma yang lebih baik untuk mendeteksi akses abnormal ke hutan yang terisolasi. Pada saat yang sama, sebagai titik awal, strategi pengaturan ambang batas sampel anomali dioptimalkan sehingga tidak bergantung pada

No	Judul Jurnal	Penulis & Tahun	Metode	Hasil
				pengalaman manusia, sehingga meningkatkan akurasi dan stabilitas algoritma deteksi anomali
5	Penerapan Metode <i>K-Means Clustering</i> untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri	Hendra Effendi, Ahmad Syahrial, Sefran Prayoga, Wahyu Dwi Hidayat (2021)	<i>K-Means Clustering</i>	Penelitian menghasilkan aplikasi yang mampu melakukan pengelompokan lahan sawit produktif dengan menerapkan metode <i>K-Means Clustering</i>. Hasil pengelompokan menunjukkan tiga <i>cluster</i> dan menunjukkan tingkat akurasi 100% dalam mengelompokkan data. Sehingga Penerapan metode <i>K-Means Clustering</i> memberikan manfaat signifikan bagi PT Kasih Agro Mandiri dalam pengelolaan lahan sawit
6	<i>Sensors Anomali Detection of Industrial Internet of Things Based on Isolated Forest Algorithm and Data Compression</i>	Desheng Liu, Hang Zhen, Dequan Kong, Xiaowei Chen, Lei Zhang, Mingrun Yuan, and Hui Wang (2021)	algoritma <i>Isolation forest</i> dan kompresi data untuk mendeteksi anomali pada sensor	penggunaan algoritma <i>Isolation forest</i> dan kompresi data dapat digunakan untuk mendeteksi anomali pada sensor dalam lingkungan IoT industri. Metode ini memiliki potensi untuk meningkatkan efisiensi dan keandalan sistem IoT industri dan dapat membantu dalam mengidentifikasi masalah pada sensor sebelum terjadi kerusakan yang lebih serius
7	Perbandingan Metode <i>One Class SVM</i> dan <i>Isolation forest</i> Dalam Mendeteksi Anomali Dalam	Hatta Chaidir, Aji Gautama Putrada, Maman Abdurrohman (2021)	Algoritma <i>One Class SVM</i> dan <i>Isolation forest</i>	Algoritma <i>One Class SVM</i> memiliki nilai Akurasi sebesar 96%, <i>Sensitivity</i> 100%, dan <i>Specificity</i> 98,2%, sedangkan algoritma <i>Isolation forest</i> memiliki

No	Judul Jurnal	Penulis & Tahun	Metode	Hasil
	<i>Activity Recognition Pada Rumah Dengan PIR Sensor</i>			nilai Akurasi sebesar 91%, <i>Sensitivity</i> 40%, dan <i>Specificity</i> 91,4%.
	Deteksi Anomali Menggunakan <i>Extended Isolation forest (Eif)</i>	Milka Wijayanti Sunarto, Dendy Kurniawan, Edy Siswanto, Haris Ihsanil Huda (2021)	algoritma <i>Extended Isolation forest</i> (EIF)	bahwa <i>Extended Isolation</i> Model perlu disesuaikan. Tes kinerja deteksi anomali mengungkapkan ketidaksempurnaan dalam deteksi struktur data jika dibandingkan dengan satu-satunya implementasi algoritma Python yang tersedia. Masalah ini dapat diperbaiki dengan <i>sub-sampling</i> yang lebih baik atau di percabangan <i>iTree</i> , di mana algoritma bisa menghindari titik perpecahan.
8	<i>Anomaly Detection in Palm Oil Yield Using Machine Learning</i>	Nuzhat Khan , Mohamad Anuar Kamaruddin, Usman Ullah Sheikh, Mohd Hafiz Zawawi, Yusri Yusup, Muhammed Paend Bakht and Norazian Mohamed Noor (2022)	<i>Extra Tree (Extremely Randomized Tree) Regressor dan AdaBoost (Adaptive Boosting) Regressor</i>	Penelitian menunjukkan bahwa model pembelajaran mesin dapat secara efektif mendeteksi anomali dalam hasil panen minyak sawit. Faktor-faktor seperti frekuensi curah hujan, kelembaban tanah, dan suhu memiliki pengaruh signifikan terhadap hasil panen. Model yang digunakan menunjukkan kemampuan untuk memberikan prediksi yang lebih baik dibandingkan dengan metode konvensional.
9	<i>Real-Time Malicious Traffic Detection With Online Isolation forest Over SD-WAN</i>	Pei Zhang , Fangzhou He, Han Zhang, Xiaohong Huang , Jilong	<i>Isolation forest</i>	Eksperimen penelitian ini pada dataset publik dunia menunjukkan bahwa OADSD dapat secara akurat mendeteksi berbagai jenis serangan

No	Judul Jurnal	Penulis & Tahun	Metode	Hasil
		Wang, Xia Yin (2023)		dengan kinerja tinggi. menunjukkan bahwa algoritma <i>Isolation forest</i> dapat digunakan untuk mendeteksi lalu lintas yang bermalah secara <i>real-time</i> di atas jaringan SD-WAN. Metode ini memiliki potensi untuk membantu dalam mengelola jaringan dan melindungi sistem dari serangan malicious
10	Deteksi Anomali Lalu Lintas Jaringan Internal Inbound Dan Outbound Menggunakan Algoritma <i>Long Short-Term Memory</i>	Khairunnisa Salsabila Riswanti, Faqih Hamami, Tien Fabrianti Kusumasari (2023)	Algoritma <i>Long Short-Term Memory</i>	mencakup penelitian yang dilakukan untuk mendeteksi anomali dalam lalu lintas jaringan internal <i>inbound</i> dan <i>outbound</i> . Penelitian ini menggunakan algoritma <i>Long Short-Term Memory (LSTM)</i> untuk memproses data sekuen dan data skala besar, yang dapat membantu dalam menjaga aktivitas jaringan tetap normal dan aman.
11	<i>Web Traffic Anomaly Detection Using Isolation Forest</i>	Wilson Chua, Arsenn Lorette Diamond Pajas, Crizelle Shane Castro, Sean Patrick Panganiban, April Joy Pasuquin, Merwin Jan Purganan, Rica Malupeng, Divine Jessa Pingad, John Paul Orolfo, Haron Hakeen Lua, dan Lemuel Clark Velasco (2024)	<i>Isolation Forest</i>	Penelitian ini mengeksplorasi penggunaan algoritma <i>Isolation Forest</i> untuk mendeteksi anomali dalam lalu lintas web. Dengan menerapkan model ini menggunakan pustaka <i>Scikit-learn</i> di Python, studi ini berhasil mencapai: Akurasi: 93%, Presisi: 95%, <i>Recall</i> : 90%, <i>F1-Score</i> : 92% Hasil ini menunjukkan bahwa <i>Isolation Forest</i> efektif dalam mengidentifikasi lalu lintas web yang tidak normal, menjadikannya solusi yang layak untuk deteksi anomali dalam data web.



Gambar 2. 1 Roadmap Penelitian Terkait

2.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah suatu metode yang digunakan pada tahap awal proses analisis data. EDA memiliki tujuan mencari pola data dengan menemukan data yang tidak normal dan memahami hubungan antar variabel. EDA dapat dilakukan dengan menggunakan berbagai teknik, seperti visualisasi data, statistik deskriptif dan uji statistik [17]. EDA berperan dalam deteksi anomali untuk membantu mengidentifikasi pola umum yang bisa membedakan data normal dan anomali, mengungkap fitur mana yang paling rawan *missing* atau *noise* dan juga untuk menunjukkan distribusi yang tidak wajar. EDA bukan hanya tentang menemukan apa yang sudah diketahui akan tetapi juga tentang kemungkinan peneliti untuk menemukan hal-hal yang tidak terduga dalam data. Peran visualisasi dalam pengungkapan informasi yang tersembunyi atau pola tidak terlihat melalui analisis numerik adalah penting [18]. Sehingga dengan melakukan EDA, peneliti

dapat mengidentifikasi perilaku data seperti *outlier*, *missing values*, atau tren musiman yang penting untuk diperhitungkan sebelum memasuki tahap *modeling*.

2.3 *Data Cleaning dan Penanganan Missing value*

Data Cleaning (pembersihan data) merupakan sebuah proses untuk mendeteksi dan memperbaiki kesalahan atau inkonsistensi dalam dataset untuk meningkatkan kualitas data sebelum analisis atau pemodelan dilakukan. Langkah ini penting agar hasil analisis yang didapatkan menjadi lebih akurat dan tidak berantakan. Dalam proses pengolahan data, tahap pembersihan data (*data cleaning*) merupakan langkah krusial untuk memastikan kualitas data yang akan digunakan dalam analisis dan pemodelan. Data yang belum dibersihkan seringkali mengandung kesalahan seperti nilai hilang (*missing values*), duplikasi, inkonsistensi format, maupun pencilan yang dapat mempengaruhi hasil akhir dari model analisis data. *data cleaning* didefinisikan sebagai proses untuk mendeteksi dan memperbaiki kesalahan dalam data guna meningkatkan kualitasnya. Tahapan ini mencakup identifikasi kesalahan, normalisasi format, dan penanganan data yang hilang atau tidak valid. Pembersihan data bukan sekadar proses satu kali, tetapi harus menjadi siklus berkelanjutan dalam *pipeline data science*. Mereka juga menyoroti bahwa automasi dalam *data cleaning* sangat membantu, tetapi tetap membutuhkan supervisi manusia untuk menjamin hasil yang berkualitas [19].

Dalam praktiknya, salah satu tantangan utama dalam *data cleaning* adalah menangani *missing values*. Kehilangan data dapat terjadi akibat kesalahan sensor, kerusakan sistem penyimpanan, atau kegagalan dalam transmisi data. Untuk mengatasi hal ini, terdapat berbagai metode imputasi yang dapat digunakan, salah satunya adalah interpolasi. Interpolasi linier merupakan metode matematis yang digunakan untuk memperkirakan nilai yang hilang atau tidak tersedia dalam suatu rangkaian data, dengan cara membentuk garis lurus antara dua titik data yang diketahui. Nilai yang diinterpolasi diasumsikan berada di antara kedua titik tersebut dan mengikuti trend linier [20].

2.4 Interpolasi Data

Interpolasi merupakan salah satu teknik dalam *preprocessing* data yang bertujuan untuk mengestimasi nilai-nilai yang hilang (*missing value*) pada suatu rangkaian data berdasarkan pola yang ada. Salah satu metode yang paling umum digunakan adalah interpolasi linier, di mana nilai yang hilang dihitung sebagai garis lurus antara dua titik data yang tersedia.

Interpolasi linier merupakan suatu metode yang digunakan untuk menentukan nilai antara dua rentan nilai yang dihasilkan berdasarkan fungsi persamaan. Interpolasi linier menghubungkan dua buah titik nilai yang memiliki hubungan linier sehingga setiap titik yang berada dalam diantara dua buah titik linier dapat ditentukan nilainya [21].

$$y = \frac{y_1 - y_0}{x_1 - x_0} (x - x_0) + y_0 \dots \dots \dots (2.1)$$

dengan:

- Titik pertama: (x_0, y_0)
- Titik kedua : (x_1, y_1)

Metode interpolasi linier bersifat sederhana dan cepat dalam perhitungan, serta cocok diterapkan pada data yang memiliki perubahan nilai secara konstan atau tidak terlalu fluktuatif. Misalnya, dalam konteks perhitungan volume permukaan datar seperti lapangan sepak bola, interpolasi linier menghasilkan nilai galat RMS yang lebih kecil dibandingkan dengan metode lain seperti interpolasi persamaan bola [22].

Namun demikian, beberapa studi juga menunjukkan bahwa interpolasi linier bukanlah metode yang paling optimal untuk semua jenis data. Fajri et al. [23] melakukan perbandingan antara metode interpolasi *Inverse Distance Weighting (IDW)*, *Kriging*, dan *Spline* pada data suhu permukaan laut. Hasilnya menunjukkan bahwa IDW memiliki standar deviasi terkecil dan dianggap sebagai metode paling akurat dalam konteks tersebut. Hal ini menandakan bahwa pemilihan metode interpolasi sangat bergantung pada jenis dan pola data.

Selain IDW, interpolasi *Newton* juga memberikan performa yang lebih baik dibandingkan metode *Lagrange* dalam estimasi nilai data saham, sebagaimana ditunjukkan oleh Astuti et al. [24]. Mereka menemukan bahwa metode *Newton* menghasilkan galat estimasi yang lebih rendah dibandingkan *Lagrange*, yang menunjukkan keunggulan metode tersebut dalam konteks data finansial.

Sementara itu, Firanto dan Idayani [25] membandingkan interpolasi *Newton-Gregory* maju dan mundur dalam estimasi jumlah penduduk. Mereka menyimpulkan bahwa interpolasi maju menghasilkan galat yang lebih kecil (0,0142) dibandingkan dengan metode mundur (0,0472), sehingga lebih direkomendasikan untuk digunakan dalam estimasi tren populasi yang terus meningkat.

Dalam konteks data klimatologi, interpolasi IDW juga menunjukkan keunggulan atas metode *Spline*. Studi oleh Kurniadi et al. [26] menunjukkan bahwa pada interpolasi curah hujan di Jawa Timur, metode IDW memiliki nilai RMSE yang jauh lebih rendah dibandingkan *Spline*, sehingga memberikan estimasi yang lebih akurat.

Berdasarkan studi-studi tersebut, metode interpolasi linier tetap menjadi pilihan yang efisien untuk data dengan tren sederhana. Namun, untuk data yang kompleks atau memiliki pola spasial yang signifikan, metode seperti IDW, *Newton*, atau *Kriging* dapat memberikan akurasi yang lebih tinggi. Oleh karena itu, pemilihan metode interpolasi perlu disesuaikan dengan karakteristik data serta kebutuhan analisis yang dilakukan.

2.5 *Anomaly Detection (AD)*

Anomali dalam KBBI memiliki arti ketidaknormalan, penyimpangan dari normal dan kelainan. Data anomali dapat diartikan sebagai data yang tidak normal atau data yang menyimpang atau berbeda daripada data pada umumnya [15]. Harusnya anomali untuk diidentifikasi pada studi lebih lanjut karena ilmu pengetahuan ini dapat digunakan untuk membuat model yang lebih kuat yang dapat menghitung adanya anomali tanpa adanya ketertarikan pada anomali yang lebih spesifik.

Deteksi Anomali dari sudut pandang algoritmanya, memiliki karakteristik yang sama antara kebaruan dan anomalnya, akan tetapi kebaruan dimasukkan dalam distribusi normal jika model dibangun kembali [27].

Data yang dapat ditangani oleh deteksi anomali mencakup semua jenis data, dari data gambar, deret waktu, atau data tabular yang sangat umum [15].

Terdapat beberapa jenis algoritma yang dapat digunakan untuk mendeteksi anomali antara lain:

1. *K-Means* : Merupakan Algoritma yang berbasis cluster. *K-Means* akan melakukan pembentukan klaster dari titik titik data yang serupa dengan jumlah klaster yang ditentukan sebelumnya. *Output* dari algoritma ini adalah klasifikasi titik sebagai normal atau sebagai anomali [15].
2. DBSCAN : Algoritma ini cocok untuk memisahkan klaster dengan kepadatan tinggi dari klaster dengan kepadatan rendah di sisi lain, algoritma akan berjuang dengan kelompok yang memiliki tingkat kepadatan yang berbeda beda. *Output* dari algoritma ini dapat bervariasi bergantung pada apakah data dimulai dengan pilihan acak dari titik pertama seperti *K-Means*, tetapi jumlah anomali tidak berpengaruh pada perhitungan titik klaster. Tujuan DBSCAN adalah untuk memperkirakan secara tepat kelompok titik normal sembari mempertimbangkan akan terjadinya titik anomali. Namun dari beberapa sumber, DBSCAN tidak dioptimalkan untuk menemukan anomali [15].
3. *Autoencoder* : *Autoencoder* atau juga *Replicator Neural Networks (RNN)* mewakili data di dalam beberapa lapisan tersembunyi dengan rekonstruksi iteratif dari data *input*. *Autoencoder* yang terlatih dengan baik dapat dengan tepat merekonstruksi contoh normal apa pun, tetapi untuk anomali, mereka menghasilkan kesalahan yang signifikan

selama rekonstruksi. Poin yang menghasilkan jumlah kesalahan yang tinggi dianggap sebagai anomali [15]

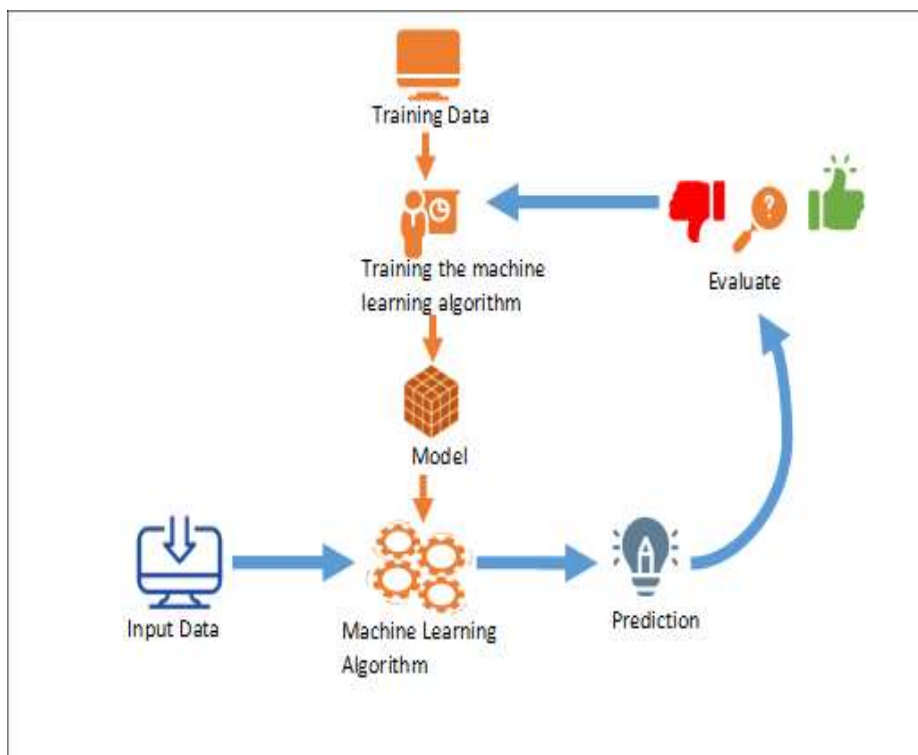
- 4 *Isolation forest (IF)* : Algoritma IF mengusulkan teknik baru pendeteksian anomali yang berbeda dari yang sebelumnya. Algoritma dibangun berdasarkan asumsi bahwa anomali adalah "sedikit dan berbeda". Alih-alih membuat profil distribusi normal dan mengukur jarak setiap titik ke titik normal, algoritma malah mengisolasi anomali dari titik lainnya dan mengukur perbedaannya menggunakan Pohon Isolasi (*iTree*). Pendekatan ini, bersama dengan ansambel Isolation Trees dan metode *subsampling*, menghasilkan model dengan kebutuhan memori yang rendah serta kompleksitas logaritmik yang rendah dari tahapan pelatihan dan evaluasi [28].

2.6 *Machine learning*

Machine learning adalah kumpulan dari algoritma pemrograman yang dapat digunakan untuk mengoptimalkan kerja komputer atau sistem berdasarkan data sampel yang ada sebelumnya. Prinsip dasar *machine learning* yaitu memanfaatkan data untuk membuat model statistik, model tersebut pada umumnya digunakan sistem untuk melakukan prediksi masa depan berdasarkan data masa lalu yang diinputkan atau mempelajari pola yang terdapat di dalam data. Salah satu keunggulan utama dari *machine learning* adalah kemampuan modifikasi serta adaptasi dalam merespon perubahan data [29]. *Machine learning* dapat dimanfaatkan dalam banyak hal, sebagai contoh beberapa peneliti yang telah mengembangkan produk dengan memanfaatkan *machine learning*, seperti pada penelitian dengan tema prediksi pengambil mata kuliah. Tujuan penelitian tersebut adalah memprediksi apakah mata kuliah yang ada akan diambil oleh mahasiswa atau tidak. Hasil dari penelitian adalah prediksi jawaban ‘Ya’ atau ‘Tidak’ untuk mahasiswa dalam mengambil mata kuliah [30].

Data akan dibagi menjadi data pembelajaran (*train dataset*) dan data pengujian (*test dataset*) untuk memastikan efisiensi model yang terbentuk. Pembagian data yang

digunakan bervariasi tergantung algoritma yang digunakan. Pada umumnya train dataset lebih banyak dari test dataset, misalnya dengan rasio 3:1. Test dataset digunakan untuk menghitung seberapa efisien model yang dihasilkan untuk melakukan klasifikasi atau prediksi kedepan yang disebut *test score*. Semakin banyak data yang digunakan, *test score* yang dihasilkan semakin baik. Nilai *test score* berada dalam rentang 0-1. Cara kerja *Machine learning* dapat dilihat pada Gambar 2.2.

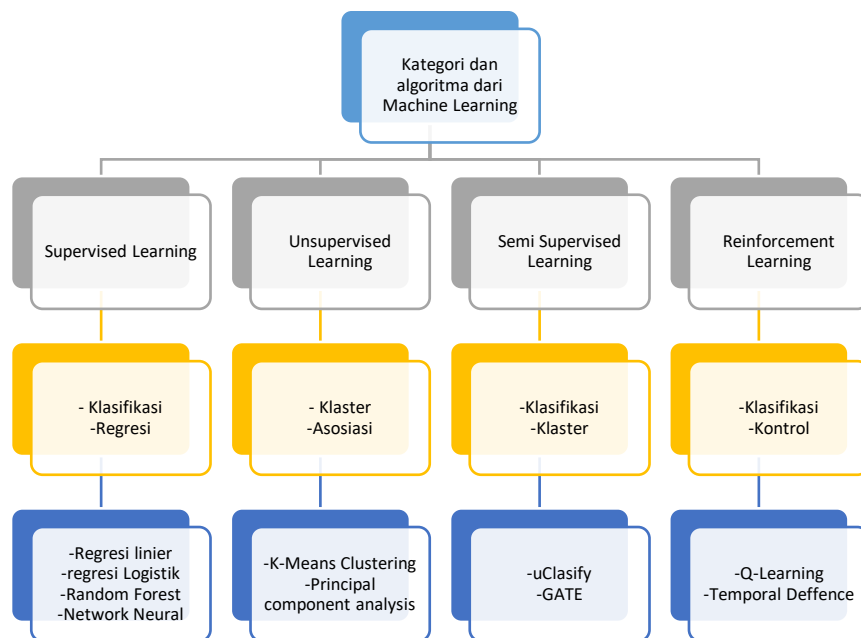


Gambar 2. 2 Cara Kerja *Machine learning*

Machine learning juga dapat diartikan sebagai aplikasi dari disiplin ilmu kecerdasan buatan (*Artificial Intelligence*) yang menggunakan teknik statistika untuk menghasilkan suatu model otomatis dari sekumpulan data, dengan tujuan memberikan komputer kemampuan untuk "*Learning*". Berdasarkan dataset tersebut, komputer akan melakukan proses belajar (*training*) untuk menghasilkan suatu model. Proses belajar ini menggunakan algoritma *machine learning* sebagai penerapan teknik statistika. Model inilah yang menghasilkan informasi, kemudian dapat dijadikan pengetahuan untuk memecahkan suatu permasalahan sebagai

proses *input - output*. Model yang dihasilkan dapat melakukan klasifikasi ataupun prediksi ke depan.

Machine learning memiliki banyak jenis dan tipenya berdasarkan hasil yang akan didapatkan dengan melihat kondisi data yang diperoleh. Jenis dari *Machine learning* dapat dilihat dari Gambar 2.3.



Gambar 2. 3 Kategori *Machine learning* dan Algoritma [31]

1. *Supervised Learning*

Supervised Learning (pembelajaran terarah) adalah salah satu metode pembelajaran mesin dimana hasil yang diharapkan pengguna, sudah diketahui atau dimiliki informasinya oleh sistem. Hal ini berarti bahwa metode pembelajaran ini bekerja dengan memanfaatkan kembali data-data dan hasil *output* yang pernah dimasukkan oleh pengguna atau dikerjakan oleh sistem sebelumnya. Pada metode ini, pola *input* dan pola *output* dibutuhkan untuk mengenali suatu informasi dalam bank memori. Ketika suatu pola *input* dibentuk, sistem akan meneruskan rangsangan data hingga ke bank memori dan sistem *output*. Sistem *output* yang menerima rangsangan data akan menampilkan pola *output* dan mencocokkan polanya dengan pola *input*. Jika

pola cocok, data akan ditampilkan dari bank memori dalam bentuk *output*. Apabila pola *input* dan pola *output* tidak ada yang cocok, maka *output* akan error. Dan jika nilai *error* cukup besar, pembelajaran lebih lanjut perlu dilakukan [32].

2. *Unsupervised Learning*

Unsupervised Learning (pembelajaran tidak terarah) adalah metode lain dalam materi pembelajaran mesin. Konsep yang metode ini gunakan jauh berbeda dengan metode *Supervised Learning* dimana pada metode ini hasil yang diharapkan tidak dapat diketahui oleh siapapun. Dengan kata lain, hasil yang akan ditampilkan hanya bergantung kepada nilai bobot yang disusun pada awal pembangunan sistem dan tentu masih dalam ruang lingkup tertentu. Tujuan utama dari metode pembelajaran ini adalah agar para penggunaanya dapat mengelompokkan objek-objek yang dinilai sejenis dalam ruang atau area tertentu. Metode pembelajaran ini sangat cocok digunakan untuk mencari atau mengklasifikasi suatu pola dari banyak objek sejenis yang tidak sepenuhnya sama [32].

3. *Semi-Supervised Learning*

Semi-Supervised Learning merupakan gabungan dari *supervised learning* dan *unsupervised learning*. Di sini dataset untuk pelatihan sebagian memiliki label dan sebagian tidak. *Google Photos* adalah contoh implementasi yang sering kita gunakan. Pada *Google Photos* kita bisa memberi *tag* atau label untuk setiap orang yang ada dalam sebuah foto. Alhasil, ketika kita mengunggah foto baru dengan wajah orang yang sebelumnya sudah kita beri label, *Google Photos* akan secara otomatis mengenali orang tersebut [32].

4. *Reinforcement Learning*

Reinforcement Learning dikenal sebagai model yang belajar menggunakan sistem *reward* dan *penalty*. Menurut Winder [33], *reinforcement learning* adalah teknik yang mempelajari bagaimana membuat keputusan terbaik, secara berurutan, untuk memaksimalkan ukuran sukses kehidupan nyata. Pembelajaran yang memiliki tujuan untuk pembuat keputusan belajar melalui proses *trial* dan *error*.

Reinforcement learning memiliki empat komponen, yaitu *action*, *agent*, *environment*, dan *reward*. *Action* adalah setiap keputusan yang diambil. Misal, saat

kita berkendara, *action* yang kita lakukan adalah mengendalikan kemudi, menginjak gas, dan mengerem. *Agent* adalah entitas yang membuat keputusan, contohnya adalah perangkat lunak, atau robot, atau bahkan manusia. *Environment* adalah sarana untuk berinteraksi, yang dapat menerima *action* dan memberikan respon berupa hasil maupun data berupa satu set observasi baru. *Reward* diberikan saat *agent* berhasil menyelesaikan tantangan. Mekanisme *feedback* ini membuat *agent* belajar tentang tindakan mana yang menyebabkan kesuksesan (menghasilkan *reward*), atau kegagalan (menghasilkan penalti). Keempat komponen ini merepresentasikan *Markov Decision Process (MDP)* [32].

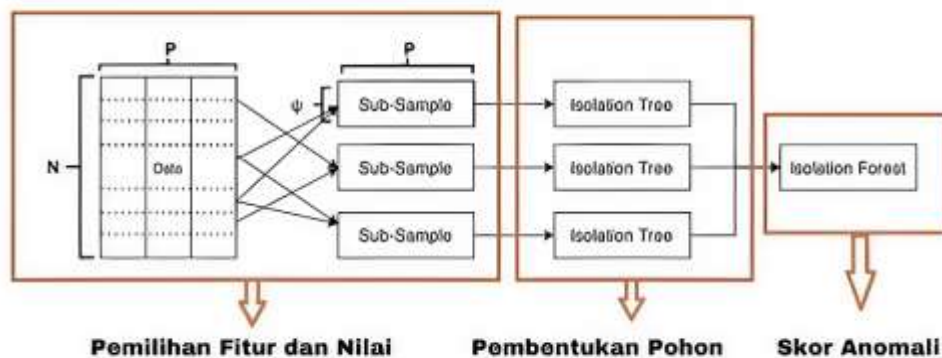
2.7 *Isolation forest*

Isolation forest merupakan salah satu dari *Unsupervised Machine learning* yang dapat digunakan untuk mendeteksi anomali. *Isolation forest* unik karena hanya berfokus pada identifikasi *outlier*, bukan data normal. Sehingga *Isolation forest* dapat dijadikan sebuah alat yang ampuh untuk mendeteksi anomali pada kumpulan data multidimensi atau multivariat. *Isolation forest* atau *i-Forest* berguna untuk membangun *iTrees* untuk kumpulan data tertentu, lalu anomali adalah kejadian yang memiliki panjang jalur rata-rata pendek di *iTrees*. Hanya ada dua variabel dalam metode ini: jumlah pohon yang akan dibangun dan ukuran *sub-sampling*. Kinerja deteksi *iForest* menyatu dengan cepat dengan jumlah pohon yang sangat sedikit, dan hanya memerlukan ukuran *sub-sampling* yang kecil untuk mencapai kinerja deteksi tinggi dengan efisiensi tinggi. Terlepas dari perbedaan utama antara isolasi dan pembuatan profil, *iForest* dibedakan dari metode berbasis model, berbasis jarak, dan berbasis kepadatan dalam beberapa hal berikut:

- Karakteristik isolasi *iTrees* memungkinkan mereka membangun model parsial dan mengeksploitasi *sub-sampling* hingga tingkat yang tidak dilakukan dalam metode yang ada. Karena sebagian besar *iTree* yang mengisolasi titik normal tidak diperlukan untuk deteksi anomali; itu tidak perlu dibangun. Ukuran sampel yang kecil menghasilkan *iTrees* yang lebih baik karena efek sampling berkurang.

- *iForest* tidak menggunakan pengukuran jarak atau kepadatan untuk mendeteksi anomali. Hal ini menghilangkan biaya komputasi yang besar dalam penghitungan jarak di semua metode berbasis jarak dan metode berbasis kepadatan.
- *iForest* memiliki kompleksitas waktu linier dengan konstanta rendah dan kebutuhan memori rendah. Metode yang ada dengan kinerja terbaik hanya mencapai perkiraan kompleksitas waktu linier dengan penggunaan memori yang tinggi.
- *iForest* mempunyai kapasitas untuk meningkatkan skala untuk menangani ukuran data yang sangat besar dan masalah berdimensi tinggi dengan sejumlah besar atribut yang tidak relevan.

Tahapan proses deteksi anomali dengan menggunakan algoritma *Isolation forest* adalah diawali dengan *input* data. Kemudian data yang sudah diinputkan akan membentuk *sub-sample* berdasarkan jenis datanya masing masing, ketika *sub-sample* sudah terbentuk maka *sub-sample* akan membentuk *isolation tree* yang akan menjadi *isolation forest* [28]. Gambar 2.4 memberikan gambaran mengenai cara kerja *Isolation forest*.



Gambar 2. 4 Cara Kerja *Isolation Forest*

Gambar 2.4 menjelaskan mengenai cara kerja *Isolation forest* yang dibagi menjadi 3 proses yaitu Pemilihan Fitur dan Nilai, Pembentukan Pohon, dan Skor Anomali.

1. Pemilihan Fitur dan Nilai

Pada proses ini dimulai dengan *input* data, dataset dengan N sebagai jumlah sampel (baris) dan P sebagai jumlah fitur (kolom). Kemudian dari data asli, algoritma mengambil subset data secara acak untuk membuat beberapa pohon keputusan (*decision trees*) khusus yang disebut *Isolation Trees* dimana setiap *sub-sample* digunakan untuk melatih satu *isolation tree*.

2. Pembentukan Pohon

Pada proses pembentukan pohon, *Isolation Tree* adalah pohon biner yang dibuat dengan cara memisahkan data menggunakan aturan pembagian acak pada fitur dan nilai (*split random*).

Proses ini diulang hingga setiap sampel data berhasil diisolasi dalam satu *node*.

3. Skor Anomali

Setelah beberapa *Isolation Trees* dibuat, semua pohon digabung menjadi satu model yang disebut *Isolation Forest*.

Skor anomali diperlukan untuk setiap metode deteksi anomali. Kesulitan dalam mendapatkan skor seperti itu dari $h(x)$ adalah bahwa meskipun tinggi maksimum *iTree* tumbuh di urutan n , tinggi rata-rata tumbuh di urutan $\log n$. Normalisasi $h(x)$ dengan salah satu ketentuan di atas tidak dibatasi atau tidak dapat dibandingkan secara langsung. Dalam konteks Algoritma *Isolation forest (IF)* dan *Extended Isolation forest (EIF)*, istilah "*iTree*" sebenarnya merujuk pada struktur data yang mirip dengan *Binary Search Tree (BST)*, tetapi dengan beberapa perbedaan yang signifikan dalam implementasi dan tujuan.

BST adalah struktur data yang merupakan subkelas dari *Binary Tree*. Ciri-ciri utama dari BST adalah:

1. Aturan Kiri-Kanan : Setiap *node* kecil kiri harus memiliki nilai yang lebih kecil dari akarnya, sedangkan setiap *node* kecil kanan harus memiliki nilai yang lebih besar dari akarnya.
2. Efisiensi Pencarian: Struktur ini memungkinkan pencarian data dengan efisiensi logaritma karena setiap *node* dapat membatasi ruang pencarian dengan cepat.

iTree pada konteks Algoritma *Isolation forest* (IF) dan *Extended Isolation forest* (EIF) tidak secara eksplisit dijelaskan sebagai BST dalam sumber-sumber yang diberikan. Namun, dapat disimpulkan bahwa *iTree* memiliki struktur yang mirip dengan BST dalam beberapa aspek:

1. Struktur Pohon: *iTree* memiliki struktur pohon yang mirip dengan BST, dengan *node-node* yang memiliki hubungan hierarki dan percabangan maksimal dua.
2. Pemisahan Data: pada IF dan EIF, *iTree* digunakan untuk memisahkan data anomali dari data normal. Algoritma ini membangun ensemble dari *iTrees* yang memiliki struktur seperti pohon pencarian biner untuk mendeteksi anomali

Perbedaan antara *iTree* dan BST adalah sebagai berikut:

1. Tujuan utama dari *iTree* pada IF dan EIF adalah untuk mendeteksi anomali dalam data, bukan untuk mencari atau menyimpan data secara efisien seperti pada BST.
2. *iTree* pada IF dan EIF menggunakan percabangan yang lebih acak dan tidak hanya terfokus pada garis yang sejajar dengan sumbu, sehingga menghilangkan bias yang terjadi pada percabangan BST

Dalam keseluruhan, *iTree* pada konteks Algoritma *Isolation forest* dan *Extended Isolation forest* memiliki struktur yang mirip dengan *Binary Search Tree*, tetapi digunakan untuk tujuan yang berbeda dan dengan implementasi yang lebih acak untuk mendeteksi anomali dalam data.

Karena *iTrees* memiliki struktur yang setara dengan *Binary Search Tree* atau BST, estimasi rata-rata $h(x)$ untuk penghentian *node* eksternal sama dengan pencarian yang gagal di BST. Mengingat kumpulan data n contoh, panjang jalur rata-rata pencarian yang gagal di BST sebagai [28]:

$$c(n) = 2H_{(n-1)} - \left(2_{(n-1)/n}\right) \dots\dots\dots(2.2)$$

dimana $H(i)$ adalah bilangan harmonik dan dapat diperkirakan dengan $\ln(i) + 0,5772156649$ (Konstanta Euler). Karena $c(n)$ adalah rata-rata dari $h(x)$ jika

diberikan n , kita menggunakannya untuk menormalkan $h(x)$. Skor anomali s dari sebuah *instance* x didefinisikan sebagai [28]:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \dots\dots\dots(2.3)$$

dimana $E(h(x))$ adalah rata-rata $h(x)$ dari kumpulan pohon isolasi. Dalam Persamaan (2.3):

- ketika $E(h(x)) \rightarrow c(n)$, $s \rightarrow 0,5$;
- ketika $E(h(x)) \rightarrow 0$, $s \rightarrow 1$;
- dan kapan $E(h(x)) \rightarrow n - 1$, $s \rightarrow 0$.

Dengan s monotonik terhadap $h(x)$.

2.8 Evaluation Metrics

Dalam penelitian ini, evaluasi kinerja model dilakukan untuk mengukur seberapa baik model mendeteksi nilai anomali dibandingkan dengan nilai acuan (*baseline*). Beberapa matrik evaluasi yang umum digunakan dalam konteks regresi dan deteksi anomali antara lain *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Ketiga matrik ini memberikan gambaran yang komprehensif terhadap akurasi model dari berbagai perspektif kesalahan prediksi. Tidak ada satu metrik yang selalu unggul dalam semua konteks. Oleh karena itu, pemilihan metrik evaluasi harus disesuaikan dengan karakteristik data dan tujuan model [33]. Dalam konteks data iklim mikro dengan kemungkinan keberadaan *noise* atau nilai ekstrim, kombinasi ketiga matrik ini dapat memberikan evaluasi performa model yang lebih komprehensif. Berikut ini penjelasan mengenai ketiga metrik

2.8.1 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual. MAE memberikan indikasi kesalahan prediksi tanpa

mempertimbangkan arah kesalahan, sehingga bersifat lebih stabil terhadap *outlier* dibandingkan RMSE [33].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \dots\dots\dots (2.4)$$

dengan:

- n = jumlah data
- y = nilai aktual
- $\hat{y}$ = nilai prediksi

Semakin kecil MAE, semakin baik performa model. MAE memberikan gambaran rata-rata kesalahan tanpa memperhitungkan arah (positif atau negatif). MAE memiliki interpretasi yang intuitif karena memiliki satuan yang sama dengan data asli, sehingga cocok digunakan dalam analisis awal performa model. Kelebihan MAE adalah tidak sensitif terhadap *outlier* dan mudah diinterpretasikan karena berada dalam skala yang sama dengan data. MAE juga memiliki keterbatasan yaitu tidak memperhitungkan variabilitas atau besarnya kesalahan ekstrim.

2.8.2 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) menghitung akar kuadrat dari rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual, memberikan penalti lebih besar pada kesalahan yang lebih besar. RMSE memberikan bobot lebih besar pada kesalahan yang lebih besar, sehingga lebih sensitif terhadap *outlier* dibandingkan MAE [34].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \dots\dots\dots (2.5)$$

RMSE lebih sensitif terhadap *outlier* karena memberikan penalti lebih besar pada kesalahan yang lebih besar, sehingga lebih cocok ketika kesalahan besar perlu diperhatikan. Kelebihan RMSE adalah sensitif terhadap *outlier* dan menekankan pada kesalahan besar. Keterbatasan RMSE adalah bisa memberikan gambaran yang terlalu negatif jika data memiliki *noise* tinggi dan juga tidak seimbang apabila dibandingkan dengan MAE dalam distribusi yang tidak normal.

2.8.3 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) mengukur kesalahan prediksi dalam bentuk persentase, memungkinkan perbandingan antar dataset dengan skala berbeda. berguna untuk memahami kesalahan dalam konteks persentase, namun dapat bermasalah ketika nilai aktual mendekati nol [34].

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \dots\dots\dots (2.6)$$

MAPE berguna untuk membandingkan performa model dalam skala yang berbeda, karena hasilnya dalam persentase (%). Kelebihan MAPE adalah memiliki skala bebas (*unitless*), akan tetapi memiliki keterbatasan berupa tidak dapat didefinisikan ketika nilai aktual 0 dan dapat memberikan nilai ekstrim jika nilai aktual sangat kecil.

BAB III

METODOLOGI PENELITIAN

3.1. Waktu dan Tempat Penelitian

Pengerjaan penelitian dilaksanakan pada waktu dan tempat sebagai berikut:

Waktu : Juli 2024 – Mei 2025

Tempat : Laboratorium Teknik Telekomunikasi, Jurusan Teknik elektro,
Fakultas Teknik, Universitas Lampung

3.2. Alat dan Bahan

Adapun peralatan dan bahan-bahan yang digunakan pada penelitian ini adalah sebagai berikut:

1. Sebuah personal komputer,
2. *Software* Phyton (Google Collab)

3.3. Metode Penelitian

Metode penelitian yang digunakan untuk penyelesaian penelitian ini adalah dengan menggunakan pemodelan dan simulasi dengan tahapan-tahapan yang dilakukan sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan sebagai sarana pencarian informasi terkait dengan topik penelitian yang sedang dilakukan, baik buku, jurnal, bahan dari internet serta sumber sumber lain yang berkaitan dengan penelitian ini, yaitu:

- a. Deteksi anomali,
- b. *Machine Learning*,
- c. Metode *Isolation forest*,
- d. *Exploratory Data Analysis (EDA)*
- e. *Data Cleaning*
- f. Interpolasi Data

2. Pemodelan Sistem

Pemodelan sistem yang dilakukan dalam penelitian ini adalah untuk menggambarkan alur kerja deteksi anomali pada data iklim mikro yang dikumpulkan dari perkebunan kelapa sawit PT. Persada Dinamika Lestari, yang terletak di Kabupaten Hulu Sungai Utara, Kalimantan Selatan. Secara umum, tahapan yang dilakukan meliputi akuisisi data, analisis data, pembersihan data (*data cleaning*), transformasi data, deteksi anomali, dan evaluasi performa model. Berikut ini adalah tahapan pemodelan sistem dalam penelitian ini:

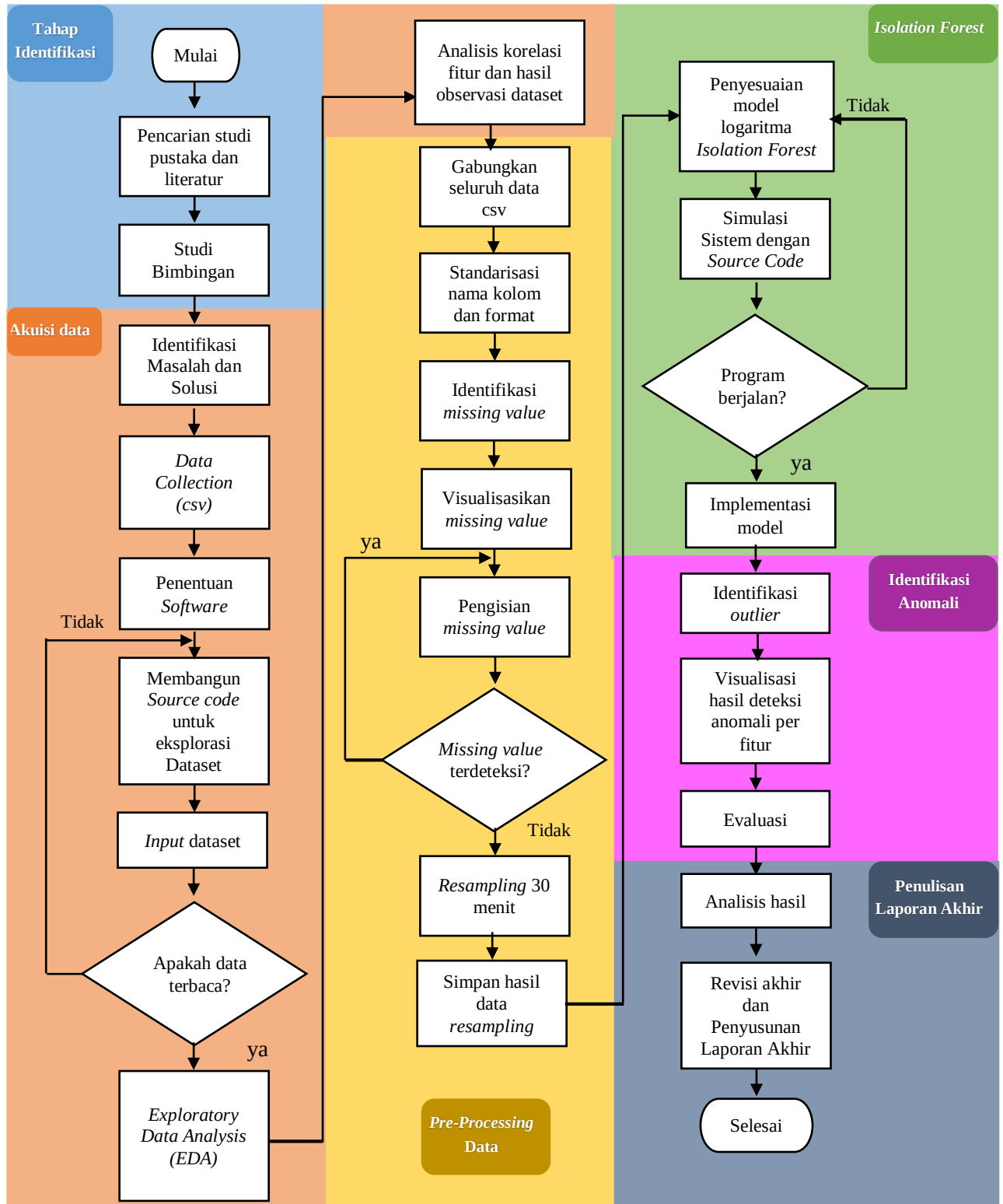
- Akuisisi data dimana dataset yang digunakan merupakan data iklim mikro dengan frekuensi pencatatan per menit, diperoleh dari sensor yang terpasang di lingkungan perkebunan kelapa sawit. Dataset terdiri dari beberapa variabel, seperti *Temperature*, *Humidity*, *Rainfall*, *wind_speed*, *pyrano*, dan lainnya, serta dikumpulkan dari bulan Juni hingga Desember 2024.
- Setelah proses akuisisi data selesai, *Exploratory Data Analysis (EDA)* diterapkan untuk menganalisis data mentah sebelum masuk ke proses pembersihan data. Pada penelitian ini EDA akan menampilkan analisis awal terhadap data seperti statistik deskriptif (*mean*, *median*, *std dev*, *min*, *max*, dll), distribusi masing-masing fitur (*histogram/boxplot*), korelasi antar fitur (*heatmap*), identifikasi awal terhadap *outlier* atau pola musiman dan juga deteksi awal *missing values*. Tujuannya adalah untuk memahami karakteristik data sebelum dilakukan pemodelan dan pembersihan.
- *Pre-Processing Data* mencakup pembersihan data dari nilai hilang dan data tidak valid. Proses ini dilakukan melalui metode interpolasi linier yang mampu memperkirakan nilai-nilai yang hilang di antara dua titik data berdasarkan garis

lurus yang menghubungkannya. Metode ini dipilih karena sederhana, efisien, dan sesuai untuk data lingkungan yang memiliki kecenderungan fluktuasi yang tidak terlalu tajam.

- Oleh karena data dikumpulkan setiap menit, proses *resampling* dilakukan untuk mengurangi *noise* dan mendapatkan pemahaman yang lebih makro terhadap tren data. Data diubah ke dalam interval 30 menit. Teknik agregasi yang digunakan dalam proses ini melibatkan pengambilan nilai rata-rata (*mean*) pada setiap interval waktu yang baru.
- Metode *Isolation forest* digunakan sebagai algoritma utama dalam proses deteksi anomali. Algoritma ini bekerja dengan prinsip isolasi, yaitu membangun pohon secara acak dan memisahkan sampel-sampel berdasarkan nilai fitur. Anomali lebih mudah terisolasi karena biasanya berada jauh dari mayoritas data. Model ini cocok untuk dataset berdimensi tinggi dan tidak memerlukan asumsi distribusi data.
- Hasil deteksi anomali dievaluasi menggunakan metrik statistik yaitu *Mean Absolute Error (MAE)*, *Root Mean Square Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Metrik-metrik ini digunakan untuk mengukur seberapa jauh prediksi model menyimpang dari nilai yang dianggap sebagai baseline, yaitu hasil *moving average*. Evaluasi dilakukan pada masing-masing fitur numerik dalam dataset.
- Visualisasi dilakukan untuk mendapatkan pemahaman yang lebih baik terhadap performa model, hasil deteksi anomali juga divisualisasikan. Visualisasi ini mencakup kurva perbandingan antara nilai aktual, nilai hasil interpolasi, dan hasil deteksi anomali. Dengan demikian, anomali yang terdeteksi dapat dianalisis secara lebih intuitif.

3.4. Diagram Alir Pengerjaan Penelitian

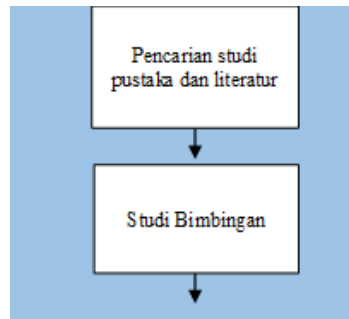
Adapun pengerjaan penelitian ini ditunjukkan pada gambar berikut:



Gambar 3. 1 Diagram Alir Pengerjaan Penelitian

1. Tahap Identifikasi

Tahap ini mempunyai peranan penting dalam deteksi anomali agar tujuan dapat ditentukan dengan tepat sasaran. Pada tahap ini penulis melakukan beberapa pendekatan terhadap objek yang dikaji.

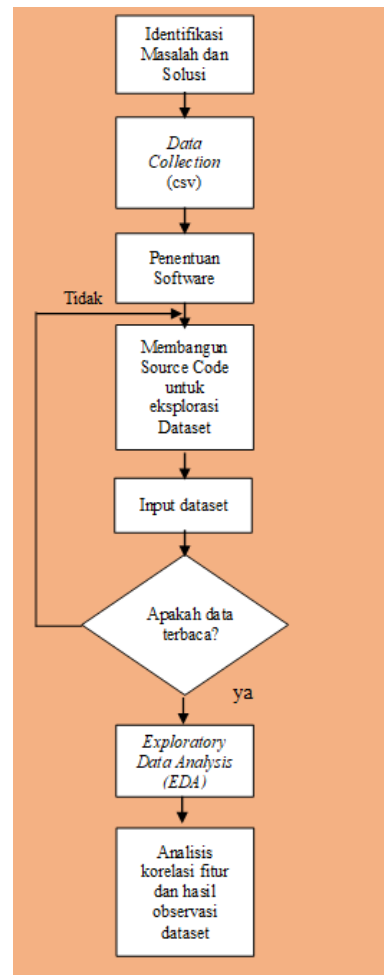


Gambar 3. 2 Tahap Identifikasi

- a. Studi literatur bertujuan untuk meningkatkan pemahaman tentang para ahli dan para praktisi yang memiliki peranan aktif dalam mengembangkan teknologi. Kegiatan ini dilakukan dengan mempelajari dan memahami topik yang berkaitan dari berbagai literatur yang berasal dari beberapa macam referensi maupun sumber ilmiah yang resmi seperti penelitian terdahulu, jurnal, *e-book*, dan artikel resmi
- b. Studi Bimbingan dilakukan melalui kegiatan diskusi dengan dosen pembimbing untuk mengulas materi dan permasalahan yang dapat ditemukan dari penelitian terkait. Kegiatan ini membantu penulis dalam mendapatkan wawasan tambahan dan pengetahuan yang lebih dalam setelah melakukan studi literatur.

2. Akuisisi Data

Akuisisi Data merupakan proses untuk mengumpulkan dataset berupa data mentah untuk diolah pada penelitian ini. Dataset yang digunakan merupakan data iklim mikro dengan frekuensi pencatatan per menit, diperoleh dari sensor yang terpasang di lingkungan perkebunan kelapa sawit. Dataset terdiri dari beberapa variabel, seperti *Temperature*, *Humidity*, *Rainfall*, *wind_speed*, *pyrano*, dan lainnya, serta dikumpulkan dari bulan Juni hingga Desember 2024.



Gambar 3. 3 Tahap Akuisisi Data

Gambar 3.3 menunjukkan Tahapan Akuisisi data, berikut adalah penjelasan untuk setiap langkah langkahnya:

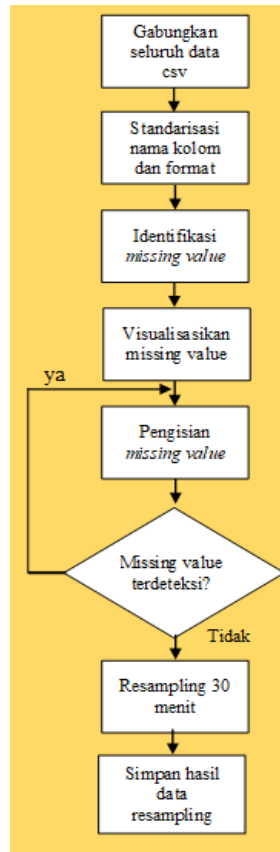
- a. Identifikasi Masalah dan Solusi merupakan tahap mendapatkan gambaran permasalahan yang akan diangkat setelah melakukan pengamatan teori dan merancang langkah solutif yang digunakan untuk menyelesaikan permasalahan terkait penelitian yang dilakukan.
- b. *Data Collection (csv)* adalah proses pengumpulan data yang akan di olah dengan menggunakan algoritma *Isolation Forest*. Pada penelitian ini, dataset yang digunakan merupakan data iklim mikro dengan frekuensi pencatatan per menit, diperoleh dari sensor yang terpasang di lingkungan perkebunan kelapa sawit. Dataset terdiri dari beberapa variabel, seperti *Temperature*, *Humidity*, *Rainfall*, *wind_speed*, *pyrano*, dan lainnya, serta

dikumpulkan dari bulan Juni hingga Desember 2024. Data iklim mikro yang digunakan memiliki rentang waktu 1 menit tiap barisnya.

- c. *Software* yang digunakan pada penelitian ini adalah *Phyton* yang akan dieksekusi pada aplikasi *Google Collab*.
- d. *Source code* untuk Eksplorasi dataset adalah *source code* yang memiliki tujuan melakukan *Exploratory Data Analysis (EDA)* sebelum masuk ke proses *Pre-Processing* data. Setelah *source code* dibangun maka data akan dimasukkan terlebih dahulu dan melakukan pengecekan apakah dataset yang dimasukkan terbaca atau tidak.
- e. *Exploratory Data Analysis (EDA)* adalah proses awal yang digunakan oleh data scientist atau analis data untuk memahami, menganalisis, dan mengeksplorasi dataset yang ada sebelum melangkah lebih jauh ke dalam model analisis atau pembuatan prediksi. Tujuan utama EDA adalah untuk mendapatkan wawasan awal tentang data, mencari pola, mengidentifikasi anomali atau *outlier*, memeriksa hubungan antara variabel, serta memvalidasi asumsi yang ada. Data yang digunakan oleh penulis adalah Data iklim mikro Kebun Kelapa Sawit PT. Persada Dinamika Lestari, yang terletak di Kabupaten Hulu Sungai Utara, Kalimantan Selatan. Data yang akan diolah adalah Data iklim mikro dari bulan Juni hingga Desember 2024. EDA pada data iklim mikro dilakukan untuk memeriksa jumlah data yang sesuai, jumlah *missing value* dan lainnya sebelum dilakukan proses yang lebih lanjut.

3. *Pre-Processing Data*

Tahap ini mencakup pembersihan data dari nilai hilang dan data tidak valid. Proses ini dilakukan melalui metode interpolasi linier yang mampu memperkirakan nilai-nilai yang hilang di antara dua titik data berdasarkan garis lurus yang menghubungkannya. Metode ini dipilih karena sederhana, efisien, dan sesuai untuk data lingkungan yang memiliki kecenderungan fluktuasi yang tidak terlalu tajam.



Gambar 3. 4 Tahap *Pre-Processing Data*

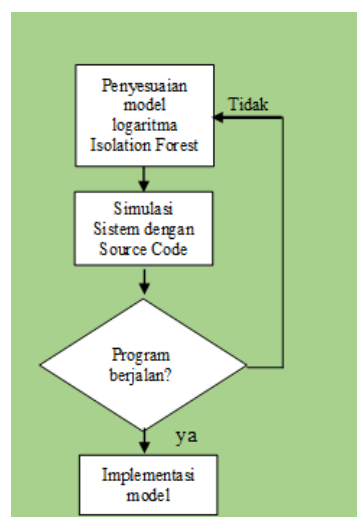
- a. Penggabungan data CSV dilakukan setelah dataset selesai dilakukan EDA untuk setiap file, setelah digabungkan data akan di standarisasi sesuai dengan kolom dan format untuk setiap fitur yang ada.
- b. Identifikasi *missing value* apabila seluruh file sudah selesai digabungkan dan di standarisasikan, kemudian visualisasikan banyaknya *missing value* dalam sebuah grafik.
- c. Apabila *missing value* sudah teridentifikasi untuk seluruh data yang ada, maka selanjutnya adalah proses pengisian *missing value*. Pengisian *missing value* pada penelitian ini akan dilakukan dengan menggunakan interpolasi linier karena sederhana, efisien, dan sesuai untuk data lingkungan yang memiliki kecenderungan fluktuasi yang tidak terlalu tajam. Kemudian akan diperiksa kembali apakah *missing value* sudah sepenuhnya terisi.
- d. Proses *resampling* data akan dilakukan ketika pengisian *missing value* sudah selesai dan tidak ada baris dan kolom yang terdeteksi sebagai

missing value. Proses *resampling* yang akan dilakukan pada penelitian ini adalah *resampling* 30 menit, dimana data yang tadinya adalah per menit dan akan disederhanakan menjadi per 30 menit dengan menggunakan teknik agregasi pengambilan nilai rata-rata (*mean*) pada setiap interval waktu yang baru.

4. Penyesuaian Algoritma *Isolation Forest*

Tahap algoritma *isolation forest* disini merupakan tahapan untuk pembentukan dan pelatihan *source code isolation forest* untuk mendeteksi anomali dengan data yang sudah di *resampling*. Gambar 3.5 menunjukkan proses untuk penyesuaian Algoritma *Isolation Forest*.

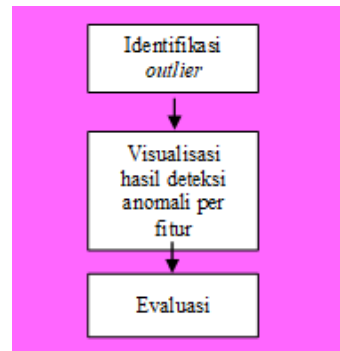
- Penyesuaian model *isolation forest* merupakan pembangunan *source code* untuk mengolah dataset yang sudah di cleaning agar bisa menemukan hasil yang diinginkan pada penelitian ini.
- Setelah model disesuaikan, *source code* akan mulai mensimulasikan salah satu data mentah terlebih dahulu agar bisa dilihat apakah model bisa digunakan atau masih ada error atau gangguan.
- Apabila simulasi sudah berhasil dan tidak mengalami *error* lagi maka langsung implementasikan model pada dataset yang sudah melewati *pre-processing data*.



Gambar 3. 5 Tahap Penyesuaian Algoritma *Isolation Forest*

5. Identifikasi Anomali

Tahap identifikasi anomali merupakan tahap yang penting dalam penelitian ini. Pada tahap ini, anomali didefinisikan sebagai *outlier* yang menyimpang daripada data lainnya.



Gambar 3. 6 Tahap Identifikasi Anomali

Gambar 3.6 menunjukkan Tahap Identifikasi Anomali. Berikut penjelasan untuk tahap ini:

- a. Pada tahapan identifikasi anomali dengan mengidentifikasi *outlier*. Setiap fitur numerik dalam dataset, seperti kelembaban (*Humidity*), curah hujan (*Rainfall*), kecepatan angin (*wind speed*), dan radiasi matahari (*pyrano*), dianalisis secara independen. Sebelum model diterapkan, data yang mengandung nilai hilang (*missing values*) dihapus agar tidak mengganggu proses pelatihan model. Proses deteksi dilakukan dengan langkah-langkah sebagai berikut:
 - Data dari setiap fitur numerik diambil dan dibersihkan dari nilai NaN.
 - Model *Isolation Forest* diinisialisasi dengan parameter *contamination* = 0.01, yang mengasumsikan bahwa 1% data merupakan anomali.
 - Model kemudian dilatih (fit) menggunakan data bersih dari masing-masing fitur.
 - Setelah pelatihan, model menghasilkan prediksi terhadap setiap data, dengan nilai 1 menandakan data normal dan -1 menandakan anomali.

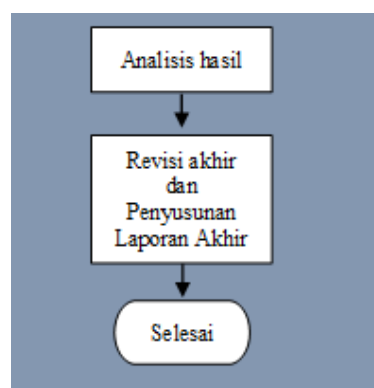
- Hasil prediksi kemudian dikonversi ke dalam bentuk biner, yaitu 0 untuk data normal dan 1 untuk data anomali.

Setelah proses tersebut, hasil deteksi anomali disimpan dalam sebuah struktur data terpisah yang disebut *anomaly_flags*, yang memiliki index waktu yang sama dengan data hasil *resampling*. Dengan adanya informasi ini, dapat dilakukan visualisasi dan evaluasi lanjutan terhadap distribusi dan pola anomali dalam data iklim mikro.

- Setelah hasil deteksi anomali divisualisasikan, maka langkah selanjutnya adalah evaluasi. Pada penelitian ini evaluasi dari hasil deteksi anomali akan dilakukan dengan menggunakan metrik evaluasi *Mean Absolute Error (MAE)*, *Root Mean Square Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*.

6. Penulisan Laporan Akhir

Tahap Laporan merupakan tahap akhir yang dilakukan setelah semua indikator keberhasilan penelitian sudah tercapai. Pembuatan laporan disusun sesuai ketentuan yang dicantumkan dalam pedoman penulisan skripsi Universitas Lampung. Gambar 3.7 menunjukkan Tahap Penulisan Laporan Akhir.



Gambar 3. 7 Tahap Penulisan Laporan Akhir

- Analisis hasil dilakukan dalam tahap laporan pada bab pembahasan untuk mengetahui hasil akhir yang telah didapatkan.

- b. Revisi akhir dan penyusunan laporan dilakukan sesuai ketentuan yang dicantumkan dalam pedoman penulisan skripsi Universitas Lampung.

3.5. Indikator Keberhasilan

Penelitian ini membahas mengenai pendeteksian anomali yang ada pada data iklim mikro. Indikator keberhasilan pada penelitian ini ditentukan berdasarkan kemampuan model dalam mendeteksi anomali secara akurat menggunakan metrik evaluasi kuantitatif yang umum digunakan dalam domain data *science* dan pembelajaran mesin, yakni *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*.

Penggunaan MAE, RMSE, dan MAPE bertujuan untuk mengukur sejauh mana nilai hasil deteksi mendekati nilai aktual (*ground truth*) berdasarkan perbandingan dengan *baseline* yang telah ditentukan, seperti *moving average*. Ketiga metrik ini dipilih karena mewakili aspek berbeda dari kesalahan model: MAE mengukur rata-rata kesalahan absolut, RMSE lebih sensitif terhadap kesalahan besar karena sifat kuadratnya, sedangkan MAPE memberikan perspektif kesalahan relatif terhadap nilai sebenarnya.

Keberhasilan sistem deteksi anomali ini akan dinilai berdasarkan seberapa rendah nilai MAE, RMSE, dan MAPE yang dihasilkan. Dalam beberapa studi, nilai MAE dan RMSE yang rendah (semakin mendekati nol) mengindikasikan bahwa prediksi sistem sangat mendekati data aktual [35] [36]. Untuk MAPE, nilai di bawah 10% umumnya dikategorikan sebagai performa yang sangat baik, sedangkan nilai di bawah 20% masih dalam kategori baik [37].

Selain itu, sebagai bagian dari proses validasi, hasil deteksi anomali juga akan divisualisasikan untuk mengobservasi apakah titik-titik anomali terdeteksi pada waktu-waktu yang secara visual menunjukkan penyimpangan terhadap pola data normal. Hal ini penting untuk memperkuat interpretasi dari metrik evaluasi numerik.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan yang didapatkan dari pengerjaan penelitian ini adalah sebagai berikut:

1. Berdasarkan hasil analisis *Exploratory Data Analysis (EDA)* data iklim mikro perkebunan kelapa sawit PT. Persada Dinamika Lestari, yang terletak di Kabupaten Hulu Sungai Utara, Kalimantan Selatan memiliki total 307.712 baris dan 7 kolom atribut yaitu fitur cuaca dan kategorik dengan 0,0201% *missing values* dan 17,2% *outliers*.
2. Deteksi Anomali dengan menggunakan algoritma *Isolation Forest* menunjukkan performa yang sangat baik pada *temperature* ($MAE = 1,77$; $RMSE = 2,41$; $MAPE = 5,53\%$), cukup baik pada *humidity* ($MAE = 6,81$; $RMSE = 9,02$; $MAPE = 11,24\%$) dan *rainfall* ($MAE = 0,003$; $RMSE = 9,02$; $MAPE = 10,68\%$), serta kurang akurat pada fitur yang bersifat fluktuatif seperti *direction* ($MAE = 38,7$; $RMSE = 61,94$; $MAPE = 50,22\%$), *wind_speed* ($MAE = 0,101$; $RMSE = 0,188$; $MAPE = 24,88\%$), dan *pyrano* ($MAE = 225,205$; $RMSE = 264,06$; $MAPE = 37,45\%$).
3. Berdasarkan hasil evaluasi yang tidak konsisten dengan 3 metrik evaluasi (MAE , $RMSE$ dan $MAPE$) menunjukkan bahwa algoritma *Isolation Forest* kurang cocok digunakan untuk mendeteksi anomali untuk data iklim mikro dengan asumsi basis *moving average* yang digunakan tidak sesuai dengan jenis data kecepatan angin dan intensitas matahari; metode evaluasi $RMSE$ tidak kompatibel dengan model *Isolation Forest*; model *Isolation Forest* untuk deteksi anomali tidak kompatibel dengan jenis data kedua parameter tersebut;

dan kualitas data kecepatan angin dan intensitas sinar matahari tidak bagus akibat *noise* atau kesalahan pembacaan sensor. .

5.2. Saran

Saran yang diberikan penulis untuk menunjang pengembangan lebih lanjut adalah sebagai berikut:

1. Penyesuaian diperlukan untuk metrik evaluasi yang dilakukan dengan penggunaan *moving average* sebagai baseline dengan karakteristik parameter seperti menambahkan salah satu metric evaluasi *Earth Square*.
2. Sebelum melakukan pengujian data, akan lebih baik jika dilakukan *pre-processing* data lanjutan untuk melakukan pembersihan data dan memeriksa apakah data yang akan diuji terdistribusi normal atau tidak sehingga dapat meningkatkan kualitas data sebelum proses deteksi anomali dilakukan.
3. Eksplorasi dengan menggunakan metode deteksi anomali yang berbeda diperlukan seperti menggunakan metode *Local Outlier Factor (LOF)*, *AutoEncoder*, *Interquartil Range (IQR)* atau *Z-Score* untuk mengetahui pendekatan mana yang paling efektif untuk mendeteksi anomali dalam konteks data iklim mikro perkebunan kelapa sawit.

DAFTAR PUSTAKA

- [1] Direktorat Jenderal Perkebunan, “Statistik Perkebunan Indonesia: Kelapa Sawit 2016–2018,” Jakarta: Kementerian Pertanian Republik Indonesia, 2018. [Online]. Tersedia: <https://ditjenbun.pertanian.go.id/?publikasi=buku-publikasi-statistik-2016-2018>. [Diakses: 09 Juli 2024]
- [2] I. Bakhtiar, D. Suradiredja, H. Santoso, dan W. Saputra. *Hutan Kita Bersawit: Gagasan Penyelesaian untuk Perkebunan Kelapa Sawit dalam Kawasan Hutan*, Jakarta: Yayasan Kehati, 2019. [Online]. Tersedia: <https://sposindonesia.org/wp-content/uploads/2019/11/Buku-Hutan-Kita-Bersawit.pdf>
- [3] A. Syahza, D. R. Aslan, and E. Iskandar, “What’s wrong with palm oil, why is it accused of damaging the environment?”. J. Phys.: Conf. Ser., vol. 1655, no. 1, p. 012134, 2020, DOI: [10.1088/1742-6596/1655/1/012134](https://doi.org/10.1088/1742-6596/1655/1/012134)
- [4] S. Lestari, H. Nugroho, and S. W. Ananta, “Opportunities and challenges for land use-based peatland restoration in Kayu Labu Village, South Sumatra, Indonesia,” IOP Conf. Ser.: Earth Environ. Sci., vol. 1107, no. 1, p. 012037, 2023, DOI: [10.1088/1755-1315/1107/1/012037](https://doi.org/10.1088/1755-1315/1107/1/012037).
- [5] T. S. Budi, “Algoritma pembelajaran mesin untuk deteksi anomali dalam sistem komputer yang terhubung IoT,” *Jurnal Portal Data*, vol. 3, no. 5, 2023, DOI: [10.31645/JISRC.23.21.1.5](https://doi.org/10.31645/JISRC.23.21.1.5)
- [6] (2019, 17 Oktober), “Apa itu deteksi anomali?,” *Amazon Web Services*, [Online]. Tersedia: <https://aws.amazon.com/what-is/anomaly-detection/>. [Diakses: 14 April 2024].
- [7] N. Khan et al., “Prediction of oil palm yield using machine learning in the perspective of fluctuating weather and soil moisture conditions: Evaluation of a generic workflow,” *Plants*, vol. 11, no. 13, p. 1697, 2022, DOI: <https://doi.org/10.3390/plants11131697>

- [8] E. H. Effendi, "Penerapan Metode K-Means Clustering untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri," *TEKNOMATIKA*, vol. 11, no. 2, pp. 117–126, 2021. Tersedia: <https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/543>
- [9] M. A. Vikram, "Anomaly Detection in Network Traffic Using Unsupervised Machine Learning Approach," in *Proc. 5th Int. Conf. Commun. Electron. Syst. (ICCES)*, Coimbatore, India, 2020. Tersedia: <https://www.researchgate.net/publication/342852797>
- [10] M. Kiran et al., "Detecting anomalous packets in network transfers: Investigations using PCA, autoencoder and Isolation Forest in TCP," *Machine Learning*, vol. 109, pp. 1127–1143, 2020. Tersedia: <https://www.researchgate.net/publication/339895292>
- [11] H. Chen et al., "Anomaly detection and critical attributes identification for products with multiple operating conditions based on Isolation Forest," *Advanced Engineering Informatics*, vol. 46, pp. 1–10, 2020. Tersedia: <https://www.researchgate.net/publication/343124187>
- [12] S. Liang et al., "Anomaly detection of power grid dispatching platform based on Isolation Forest and K-means fusion algorithm," *J. Phys.: Conf. Ser.*, vol. 1601, no. 2, p. 022010, 2020. DOI: [10.1088/1742-6596/1601/2/022010](https://doi.org/10.1088/1742-6596/1601/2/022010)
- [13] D. Liu et al., "Sensors Anomaly Detection of Industrial Internet of Things Based on Isolated Forest Algorithm and Data Compression," *Scientific Programming*, vol. 2021, Art. no. 6699313, pp. 1–9, Jan. 2021. DOI: <https://doi.org/10.1155/2021/6699313>
- [14] H. Chaidir, "Perbandingan Metode One Class SVM dan Isolation Forest dalam Mendeteksi Anomali dalam Activity Recognition pada Rumah dengan PIR Sensor," *Skripsi, S1 Informatika, Universitas Telkom, Bandung*, ISSN: 2355-9365, 2021. Tersedia: <https://repository.telkomuniversity.ac.id/pustaka/172424/>
- [15] M. W. Sunarto, D. Kurniawan, E. Siswanto, dan H. I. Huda, "Deteksi Anomali Menggunakan Extended Isolation Forest (EIF)," *Teknik: Jurnal Ilmu Teknik dan Informatika*, vol. 1, no. 2, pp. 96–111, Juni 2023. DOI: <https://doi.org/10.51903/teknik.v1i2.324>

- [16] Z. Pei et al., "Real-Time Malicious Traffic Detection With Online Isolation Forest Over SD-WAN," *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, vol. 18, pp. 2076-2091, 2023. DOI: <https://doi.org/10.1109/TIFS.2023.3262121>
- [17] R. Ramadhani, R. Ramadhanu, dan T. Hidayat, "Exploratory Data Analysis (EDA) untuk Mengetahui Distribusi Data Kualitas Susu Sapi," *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 23, no. 1, pp. 68–76, Feb. 2024. DOI: <https://doi.org/10.53513/jis.v23i1.9500>
- [18] J. W. Tukey, *Exploratory Data Analysis*, Reading, USA: Addison-Wesley, 1977.
- [19] M. Abdelaal, C. Hammacher, dan H. Schöning, "REIN: A Comprehensive Benchmark Framework for Data Cleaning Methods in ML Pipelines," *arXiv preprint arXiv:2302.04702*, 2023. DOI: <https://doi.org/10.48550/arXiv.2302.04702>
- [20] M. R. Ismail, A. Zain, J. Jamaludin, F. Dewantoro, dan D. Pratiwi, "Perhitungan Data Curah Hujan yang Hilang dengan Menggunakan Metode Interpolasi Linier," *Jurnal Teknik Sipil*, vol. 4, no. 2, pp. 1–7, 2023. Tersedia: <https://jim.teknokrat.ac.id/index.php/tekniksipil/article/view/5023>
- [21] I. G. I. Sudipa, I. K. A. G. Wiguna, I. N. T. A. Putra, dan K. Hardiatama, "Implementasi Metode Analytical Hierarchy Process dan Interpolasi Linier dalam Penentuan Lokasi Wisata di Kabupaten Karangasem," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 866–878, 2021. Tersedia: <https://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/383>
- [22] S. N. Iman W., "Studi Komparasi Metode Interpolasi Linier dan Interpolasi Persamaan Bola dalam Perhitungan Volumetrik," *Skripsi, Program Studi Kartografi dan Penginderaan Jauh, Universitas Gadjah Mada, Yogyakarta*, 2014. Tersedia: <https://etd.repository.ugm.ac.id/penelitian/detail/71002>
- [23] I. Fajri, S. B. Susilo, dan R. E. Arhatin, "Perbandingan Metode Interpolasi IDW, Kriging, dan Spline pada Data Spasial Suhu Permukaan Laut," *Jurnal Ilmu dan Teknologi Kelautan Tropis*, vol. 8, no. 1, pp. 183–194, 2016. [Online]. Tersedia: <https://repository.ipb.ac.id/handle/123456789/87016>

- [24] L. W. Astuti, S. Sudarwanto, dan L. Ambarwati, "Perbandingan Metode Lagrange dan Metode Newton pada Interpolasi Polinomial dalam Mengestimasi Harga Saham," *Jurnal Matematika: Teori dan Terapan*, vol. 1, no. 2, pp. 25–32, 2015. Tersedia:
<https://journal.unj.ac.id/unj/index.php/jmt/article/download/7165/5149>
- [25] F. Agus, dan I. Darsih, "Perbandingan Performa Metode Interpolasi Polinomial Newton-Gregory Maju Dan Newton-Gregory Mundur Dalam Mengestimasi Jumlah Penduduk Di Provinsi Papua," *Jurnal Matematika Sains Dan Teknologi*, vol. 23, no. 2, pp. 106-113, 2023. Tersedia:
<http://dx.doi.org/10.33830/jmst.v23i2.5147.2022>
- [26] K. Hanif et al., "Perbandingan metode IDW dan Spline dalam interpolasi data curah hujan (Studi kasus: curah hujan bulanan di Jawa Timur periode 2012–2016)," *Prosiding Seminar Nasional Geotik 2018*, Surakarta, Indonesia, April. 2018, pp. 213-220. Tersedia:
<https://www.researchgate.net/publication/344169321>
- [27] Z. Arthur, dan F. Peter, "There and back again: Outlier detection between statistical reasoning and data mining algorithms," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 6, pp. 1-26. 2018. [Online]. DOI: <https://doi.org/10.1002/widm.1280>
- [28] F. T. Liu, K. M. Ting and Z. -H. Zhou, "Isolation Forest," 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413-422, DOI: [10.1109/ICDM.2008.17](https://doi.org/10.1109/ICDM.2008.17).
- [29] E. Alpaydin, *Introduction to Machine Learning*, 4th ed. Cambridge, MA: MIT Press, 2020.
- [30] F. A. Bachtiar, I. K. Syahputra, dan S. A. Wicaksono, "Perbandingan Algoritme Machine Learning Untuk Memprediksi Pengambil Mata Kuliah," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 6, no. 5, pp. 543-548. 2018. DOI: <http://dx.doi.org/10.25126/jtiik.2019651755>
- [31] M. T. Mohammad, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future *Directions*," *Computers* 2023, vol. 12, no. 6, pp. 91, 2023. DOI: <https://doi.org/10.3390/computers12050091>

- [32] J. S. Suroso, “Jenis-jenis Machine Learning,” *MMSI BINUS University*, Sep. 27, 2022. Tersedia: <https://mmsi.binus.ac.id/2022/09/27/jenis-jenis-machine-learning/>. [Diakses: 10 Oktober 2024]
- [33] W. Phil, *Reinforcement Learning: Industrial Applications of Intelligent Agents*, California, O’Reilly Media, 2020.
- [34] L. J. G. van Zyl, J. T. Marais, dan B. M. Herbst, “Comparison of Regression Error Metrics for Machine Learning Models in Model-Based Testing,” *IEEE Access*, vol. 8, pp. 207396–207407, 2020. DOI: <https://doi.org/10.1109/AC>