

**PREDIKSI JANGKA PANJANG *SOLAR IRRADIANCE* PADA
PERMUKAAN PULAU JAWA MENGGUNAKAN
*ENSEMBLE LEARNING***

LAPORAN TUGAS AKHIR

Laporan ini disusun untuk memenuhi salah satu syarat memperoleh
Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika
Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



Disusun Oleh :

MUHAMMAD ZULFIKAR

32602100092

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG
2025**

***LONG-TERM PREDICTION OF SOLAR IRRADIANCE ON THE
SURFACE OF JAVA ISLAND USING
ENSEMBLE LEARNING***

FINAL PROJECT

*Proposed to complete the requirement to obtain a bachelor's degree (S1) at
Informatics Engineering Departement of Industrial Technology Faculty Sultan
Agung Islamic University*



ARRANGED BY :

MUHAMMAD ZULFIKAR

32602100092

***INFORMATICS ENGINEERING STUDY PROGRAM
FACULTY OF INDUSTRIAL TECHNOLOGY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG***

2025

**LEMBAR PENGESAHAN
TUGAS AKHIR**

**PREDIKSI JANGKA PANJANG *SOLAR IRRADIANCE* PADA
PERMUKAAN PULAU JAWA MENGGUNAKAN
*ENSEMBLE LEARNING***

**MUHAMMAD ZULFIKAR
NIM 32602100092**

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 12 Agustus 2025

TIM PENGUJI UJIAN SARJANA :

Bagus SWP, S.Kom., M.Cs

NIK. 210616051

(Penguji 1)

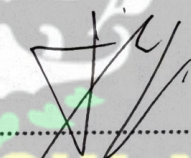


15-08-2025

Moch. Taufik, ST., MIT

NIK. 210604034

(Penguji 2)

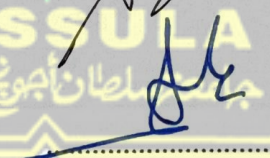


15-08-2025

Ir.Suryani Alifah, MT. Ph.D

NIK. 210601024

(Pembimbing)



22-08-2025

Semarang, 22 Agustus 2025

Mengetahui,

Kaprodi Teknik Informatika
Universitas Islam Sultan Agung



Moch. Taufik, ST., MIT

NIK. 210604034

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Muhammad Zulfikar

NIM : 32602100092

Judul Tugas Akhir : PREDIKSI JANGKA PANJANG *SOLAR IRRADIANCE*
PADA PERMUKAAN PULAU JAWA MENGGUNAKAN
ENSEMBLE LEARNING

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 27 Agustus 2025

Yang Menyatakan,



Muhamnad Zulfikar

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Zulfikar

NIM : 32602100092

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Alamat Asal : Demak

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : *Prediksi Jangka Panjang Solar Irradiance Pada Permukaan Pulau Jawa Menggunakan Ensemble Learning*.

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 27 Agustus 2025


Muhammad Zulfikar

KATA PENGANTAR

Puji syukur penulis panjatkan kepada ALLAH SWT, yang telah memberikan rahmat, taufik serta hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Prediksi Radiasi Surya Permukaan Pulau Jawa Menggunakan *Ensemble learning*.” ini dengan baik. Dengan penuh rasa hormat, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Hj. Novi Marlyana, S.T., M.T.
3. Dosen pembimbing penulis Ibu Ir.Suryani Alifah, MT., Ph, D yang telah memberikan arahan, bimbingan, dan saran yang berarti dalam penyelesaian tugas akhir ini.
4. Seluruh dosen Program Studi Teknik Informatika, Fakultas Teknologi Industri UNISSULA yang telah memberikan ilmunya kepada penulis.
5. Orang tua penulis, Ibu suyati yang selalu memberikan segala doa, dukungan, dan motivasi dengan penuh limpahan kasih sayang sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik
6. Untuk Diri Sendiri yang tidak pernah menyerah dan berhenti mencoba sesulit apapun rintangan kuliah dan selama proses penyusunan laporan ini.
7. Rekan seperjuangan, yang telah memberikan dukungan moral, motivasi, serta semangat dalam proses penyelesaian tugas akhir ini
8. Dan kepada seluruh pihak yang tidak dapat saya sebutkan satu persatu.

Semarang,

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR	iii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
ABSTRAK	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Pembatasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori.....	12
2.2.1 Radiasi Surya	12
2.2.2 <i>Ensemble Learning</i>	13
2.2.3 <i>Bagging (Bootstrap Aggregating)</i>	15
2.2.4 <i>Decision Tree</i>	16
2.2.5 <i>Boosting</i>	18
2.2.6 <i>Extreme Gradient Boosting (XGBoost)</i>	19
BAB III METODOLOGI PENELITIAN	21
3.1 Pengumpulan Data	22
3.2 <i>Preprocessing Data</i>	22
3.2.1 <i>Data Cleaning</i>	22
3.2.2 Transformasi Dan Pembersihan Data.....	23
3.2.3 Pemisahan Data Berdasarkan Lokasi	23

3.2.4	Agregasi Dan Perhitungan Rata-Rata	23
3.3	<i>Feature Engineering</i>	24
3.4	<i>Split Data</i>	24
3.5	Pembuatan Model.....	24
3.5.1	Pembuatan <i>Bagging Base model Decision Tree</i>	25
3.5.2	Pembuatan Model <i>Boosting (XGBoost Regressor)</i>	26
3.5.3	Penggabungan Model.....	27
3.6	Evaluasi Model.....	28
3.7	<i>Software</i> yang digunakan	29
BAB IV HASIL DAN ANALISIS PENELITIAN		33
4.1	Hasil Peneltian	33
4.1.1	Pengumpulan Data	33
4.1.2	<i>Preprocessing Data</i>	34
4.1.3	<i>Feature Engineering</i>	36
4.1.4	<i>Split Data</i>	36
4.1.5	Evaluasi Model.....	37
4.2	Hasil Evaluasi.....	37
4.3	Hasil <i>Deployment</i>	45
BAB V KESIMPULAN DAN SARAN		53
5.1	Kesimpulan	53
5.2	Saran.....	53
DAFTAR PUSTAKA		54

DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka	10
Tabel 4. 1 Dataset <i>ALLKSY_SFC_SW_DWN</i>	33
Tabel 4. 2 Mengecek Duplicate Data	34
Tabel 4. 3 Hasil Missing Value.....	35
Tabel 4. 4 Data Spliting	37
Tabel 4. 5 Hasil Evaluasi Lokasi LAT -6.5 LON 105.5	37
Tabel 4. 6 Hasil Evaluasi Lokasi LAT -6.5 LON 106.5	38
Tabel 4. 7 Hasil Evaluasi Lokasi LAT -6.5 LON 107.5	38
Tabel 4. 8 Hasil Evaluasi Lokasi LAT -6.5 LON 108.5	38
Tabel 4. 9 Hasil Evaluasi Lokasi LAT -6.5 LON 109.5	38
Tabel 4. 10 Hasil Evaluasi Lokasi LAT -6.5 LON 110.5	39
Tabel 4. 11 Hasil Evaluasi Lokasi LAT -6.5 LON 111.5	39
Tabel 4. 12 Hasil Evaluasi Lokasi LAT -6.5 LON 112.5	39
Tabel 4. 13 Hasil Evaluasi Lokasi LAT -6.5 LON 113.5	39
Tabel 4. 14 Hasil Evaluasi Lokasi LAT -6.5 LON 114.5	40
Tabel 4. 15 Hasil Evaluasi Lokasi LAT -7.5 LON 105.5	40
Tabel 4. 16 Hasil Evaluasi Lokasi LAT -7.5 LON 106.5	40
Tabel 4. 17 Hasil Evaluasi Lokasi LAT -7.5 LON 107.5	41
Tabel 4. 18 Hasil Evaluasi Lokasi LAT -7.5 LON 108.5	41
Tabel 4. 19 Hasil Evaluasi Lokasi LAT -7.5 LON 109.5	41
Tabel 4. 20 Hasil Evaluasi Lokasi LAT -7.5 LON 110.5	41
Tabel 4. 21 Hasil Evaluasi Lokasi LAT -7.5 LON 111.5	42
Tabel 4. 22 Hasil Evaluasi Lokasi LAT -7.5 LON 112.5	42
Tabel 4. 23 Hasil Evaluasi Lokasi LAT -7.5 LON 113.5	42
Tabel 4. 24 Hasil Evaluasi Lokasi LAT -7.5 LON 114.5	42
Tabel 4. 25 Hasil Evaluasi Lokasi LAT -8.5 LON 105.5	43
Tabel 4. 26 Hasil Evaluasi Lokasi LAT -8.5 LON 106.5	43
Tabel 4. 27 Hasil Evaluasi Lokasi LAT -8.5 LON 107.5	43
Tabel 4. 28 Hasil Evaluasi Lokasi LAT -8.5 LON 108.5	43
Tabel 4. 29 Hasil Evaluasi Lokasi LAT -8.5 LON 109.5	44
Tabel 4. 30 Hasil Evaluasi Lokasi LAT -8.5 LON 110.5	44
Tabel 4. 31 Hasil Evaluasi Lokasi LAT -8.5 LON 111.5	44
Tabel 4. 32 Hasil Evaluasi Lokasi LAT -8.5 LON 112.5	44
Tabel 4. 33 Hasil Evaluasi Lokasi LAT -8.5 LON 113.5	45
Tabel 4. 34 Hasil Evaluasi Lokasi LAT -8.5 LON 114.5	45

DAFTAR GAMBAR

Gambar 2. 1 Radiasi Surya.....	12
Gambar 2. 2 Arsitektur Enemble Learning.....	14
Gambar 2. 3 Arsitektur Bagging (Bootstrap Aggregation).....	16
Gambar 2. 4 Arsitektur Decision Tree	17
Gambar 2. 5 Arsitektur Boosting	18
Gambar 2. 6 Arsitektur XGBoost	20
Gambar 3. 1 Flowchart Penelitian.....	21
Gambar 3. 2 Pengambilan Data Pulau Jawa di Nasa Power DAV	22
Gambar 3. 3 Data Historis GHI di Pulau Jawa ((1984-2024).....	22
Gambar 3. 4 Flowchart Model Bagging.....	25
Gambar 3. 5 Flowchar Model XGBoost	26
Gambar 3. 6 Flowchart Gabungan	27
Gambar 4. 1 Halaman Awal.....	46
Gambar 4. 2 Peta Prediksi Radiasi Surya Bagging Base model Decsion Tree....	47
Gambar 4. 3 Hasil Prediksi Radiasi Surya Bagging Base model Decsion Tree ..	48
Gambar 4. 4 Peta Prediksi Radiasi Surya XGBoost	49
Gambar 4. 5 Hasil Prediksi XGBoost	50
Gambar 4. 6 Grafik Hasil Prediksi Penggabungan Ensemble Learning	51
Gambar 4. 7 Hasil Prediksi Penggabungan Ensemble Learning.....	52



ABSTRAK

Energi surya menjadi solusi strategis untuk memenuhi permintaan energi di Pulau Jawa, namun pemanfaatannya terkendala oleh variabilitas radiasi matahari. Penelitian ini bertujuan mengembangkan model prediksi jangka panjang radiasi surya yang akurat menggunakan pendekatan *ensemble learning*. Data yang digunakan adalah data historis dari tahun 1984 hingga 2024 yang bersumber dari NASA POWER. Penelitian ini mengimplementasikan tiga pendekatan: *Bagging* dengan *base model Decision Tree*, *Boosting* menggunakan *XGBoost*, dan sebuah model gabungan dengan metode *Weighted Average*. Model dievaluasi menggunakan metrik *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *R² Score*. Hasil penelitian menunjukkan bahwa model gabungan, yang memberikan bobot 0.9 pada prediksi *XGBoost* dan 0.1 pada *Bagging*, secara konsisten menunjukkan performa paling unggul. Hal ini membuktikan bahwa kombinasi model dengan pembobotan yang tepat mampu meningkatkan akurasi prediksi secara signifikan. Hasil prediksi ini dapat mendukung perencanaan dan pengembangan infrastruktur energi surya yang lebih efisien dan berkelanjutan di Pulau Jawa.

Kata Kunci : Prediksi Radiasi Surya, *Ensemble Learning*, *Bagging*, *XGBoost*, Pulau Jawa

ABSTRACT

Solar energy presents a strategic solution to meet the growing energy demand on Java Island, yet its utilization is hindered by the variability of solar radiation. This research aims to develop an accurate long-term solar radiation prediction model using an *Ensemble Learning approach*. The data utilized is historical data from 1984 to 2024, sourced from NASA POWER. This study implements three approaches: *Bagging* with a *Decision Tree base model*, *Boosting* using *XGBoost*, and a combined model using the *Weighted Average method*. The models were evaluated using *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, and *R² Score* metrics. The results indicate that the combined model, which assigns a weight of 0.9 to the *XGBoost* prediction and 0.1 to the *Bagging* prediction, consistently demonstrates the most superior performance. This proves that combining models with appropriate weighting can significantly enhance prediction accuracy. These prediction results can support the planning and development of more efficient and sustainable solar energy infrastructure on Java Island.

Keywords: Solar Radiation Prediction, *Ensemble Learning*, *Bagging*, *XGBoost*, Java Island

BAB I

PENDAHULUAN

1.1 Latar Belakang

Energi surya merupakan salah satu sumber energi terbarukan yang paling potensial dan menjanjikan dalam menjawab tantangan kebutuhan energi di masa depan. Dalam skala global, peralihan menuju energi bersih dan rendah emisi karbon menjadi prioritas utama untuk mengatasi krisis iklim serta mengurangi ketergantungan terhadap bahan bakar fosil yang kian menipis dan menimbulkan dampak negatif terhadap lingkungan. Indonesia, sebagai negara yang berada di wilayah tropis dan dilintasi garis khatulistiwa, memiliki potensi energi surya yang melimpah dan stabil sepanjang tahun. Kondisi geografis ini memberikan keuntungan tersendiri dalam pengembangan energi surya, yang dapat dimanfaatkan untuk mendukung transisi energi nasional menuju sistem yang lebih berkelanjutan dan ramah lingkungan (Dwisari dkk., 2023).

Pulau Jawa merupakan pusat aktivitas ekonomi, industri, pemerintahan di Indonesia, serta menjadi daerah dengan kepadatan penduduk tertinggi sekitar 56,10% dari seluruh penduduk Indonesia tinggal di pulau ini, padahal luas wilayahnya hanya sekitar 7% dari total wilayah Indonesia (Arif & Nurwati, 2022). Permintaan energi di wilayah ini sangat tinggi dan terus meningkat seiring dengan urbanisasi dan pertumbuhan ekonomi yang pesat. Namun, ketersediaan sumber energi fosil di Pulau Jawa semakin terbatas dan tidak seimbang dengan lonjakan kebutuhan energi.

Tingginya konsumsi energi berbasis fosil di Pulau Jawa juga berkontribusi besar terhadap emisi gas rumah kaca yang berdampak negatif terhadap lingkungan (Annisa Nur Oktaviani, 2025). Hal ini memperkuat untuk mengembangkan sumber energi alternatif yang bersih, efisien, dan dapat diperbarui. Energi surya menjadi salah satu solusi yang strategis, namun pemanfaatannya masih terkendala oleh berbagai tantangan teknis, terutama dalam hal perencanaan dan prediksi pasokan energi.

Radiasi matahari yang menjadi sumber utama energi surya sangat dipengaruhi oleh faktor-faktor dinamis seperti awan, kelembapan, polusi udara, curah hujan, dan

kondisi atmosfer lainnya. Variabilitas ini membuat prediksi radiasi matahari menjadi tugas yang kompleks, terutama jika hanya mengandalkan metode konvensional yang kurang adaptif terhadap perubahan kondisi lingkungan (Ismayanti & Baihaqi, 2024). Oleh karena itu, pendekatan berbasis kecerdasan buatan seperti pembelajaran mesin mulai dilirik untuk mengatasi tantangan ini.

Salah satu pendekatan dalam pembelajaran mesin yang terbukti efektif dalam meningkatkan akurasi prediksi adalah *ensemble learning*, yaitu metode yang menggabungkan beberapa model prediktif untuk menghasilkan hasil yang lebih kuat, stabil, dan andal. Dua teknik yang umum digunakan dalam *ensemble learning* adalah *bagging* dengan *base model decision tree*, serta *boosting* menggunakan algoritma (*Extreme Gradient Boosting*) XGBoost. Kedua teknik ini memiliki keunggulan dalam mengatasi *overfitting*, meningkatkan akurasi, dan memodelkan hubungan *non-linear* dalam data (Rizkiani & Brahma, 2024).

Berdasarkan uraian di atas, penelitian ini bertujuan untuk memprediksi radiasi surya di wilayah Pulau Jawa dengan menerapkan metode *ensemble learning*, khususnya menggunakan *bagging* dengan *base model decision tree*, serta teknik *boosting* menggunakan XGBoost. Pendekatan ini diharapkan mampu menghasilkan prediksi yang lebih akurat dan andal guna mendukung perencanaan dan pengembangan energi terbarukan di wilayah strategis Pulau Jawa.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana implementasi model *bagging base model decision tree* dalam memprediksi radiasi energi surya permukaan di wilayah Pulau Jawa?
2. Bagaimana implementasi model *boosting* menggunakan XGBoost dalam memprediksi radiasi energi surya permukaan di wilayah Pulau Jawa?
3. Bagaimana penerapan metode *weighted average* dalam menggabungkan prediksi dari model *bagging base model decision tree* dan XGBoost?
4. Bagaimana hasil evaluasi model *bagging base model decision tree*, XGBoost, dan gabungan keduanya melalui metode *weighted average ensemble* terhadap akurasi prediksi radiasi energi surya permukaan di wilayah Pulau Jawa?

1.3 Pembatasan Masalah

Beberapa batasan yang diambil adalah :

1. Data yang digunakan adalah data historis radiasi surya di wilayah Pulau Jawa dari tahun 1984 hingga 2024, yang diperoleh dari Nasa Power atau bisa diakses di [POWER | DAV](#).
2. Penelitian dibatasi pada pemodelan prediksi radiasi surya menggunakan dua pendekatan *ensemble learning*, yaitu *bagging base model decision tree* dan *boosting* menggunakan XGBoost.
3. Penelitian juga mencakup penggabungan prediksi menggunakan metode *weighted average ensemble*.
4. Evaluasi model menggunakan metrik *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *R² Score*.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi pendekatan *ensemble learning* dalam memprediksi radiasi energi surya permukaan di wilayah Pulau Jawa :

1. Mengimplementasikan dan mengevaluasi *bagging* dengan *base model decision tree* dalam memprediksi radiasi energi surya permukaan di wilayah Pulau Jawa.
2. Mengimplementasikan dan mengevaluasi XGBoost dalam memprediksi radiasi energi surya permukaan di wilayah Pulau Jawa.
3. Menggabungkan hasil prediksi dari model *bagging* dan XGBoost menggunakan metode *weighted average ensemble* untuk memperoleh prediksi yang lebih akurat.

1.5 Manfaat Penelitian

1. Penelitian ini memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang pemodelan data energi terbarukan dengan pendekatan *ensemble learning*.
2. Hasil prediksi akurat ini dapat membantu instansi pemerintah, peneliti, dan pengembang energi surya di Pulau Jawa untuk merencanakan pemanfaatan energi terbarukan yang efisien dan mendukung pembangunan berkelanjutan.

1.6 Sistematika Penulisan

BAB I	:	PENDAHULUAN
		Pada BAB I menjelaskan tentang latar belakang, pemilihan judul, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian dan sistematika penelitian
BAB II		TINJAUAN PUSTAKA DAN DASAR TEORI
		Pada Bab II ini memuat penelitian terdahulu serta landasan teori terkait membantu memahami konsep model <i>bagging</i> dengan <i>base model decision tree</i> dan <i>boosting</i> menggunakan XGBoost dalam memprediksi radiasi surya permukaan Pulau Jawa untuk melengkapi penelitian ini
BAB III		METODE PENELITIAN
		Pada BAB III Bab ini menjelaskan tahapan penelitian, mulai dari pengumpulan dan pemrosesan data, pemisahan data latih dan uji, hingga evaluasi model dari hasil prediksi untuk mendapatkan hasil akurasi yang baik.
BAB IV		HASIL DAN ANALISIS PENELITIAN
		Pada BAB IV berisi tentang pemaparan hasil penelitian yang dimulai dari pembuatan sistem sampai dengan proses deployment.
BAB V		KESIMPULAN DAN SARAN
		Pada BAB V merangkum keseluruhan proses penelitian dari awal sampai akhir.

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Salah satu penelitian yang mengimplementasikan pendekatan *ensemble learning* menunjukkan bahwa metode ini mampu memberikan performa prediksi yang lebih unggul dibandingkan dengan model tunggal. Hasil evaluasi menggunakan *Mean Squared Error* (MSE) menunjukkan bahwa model *bagging* memiliki tingkat kesalahan terendah, yaitu sebesar 145.448. Di sisi lain, ketika dievaluasi menggunakan metrik *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE), model *boosting* justru menunjukkan performa terbaik, dengan masing-masing nilai kesalahan sebesar 11,56%, 43,80, dan 38,760. Temuan ini mengindikasikan bahwa pendekatan *ensemble*, khususnya model *boosting*, dapat dimanfaatkan untuk menghasilkan prediksi yang lebih akurat dalam menentukan waktu penggantian baterai aki kendaraan (Rayadin dkk., 2024).

Penelitian lain menunjukkan bahwa penerapan teknik *ensemble learning* dapat secara signifikan meningkatkan performa model prediktif. Metode *AdaBoost* tercatat mampu meningkatkan akurasi algoritma *Naive Bayes* sebesar 7,42%, membuktikan bahwa model sederhana sekalipun dapat ditingkatkan secara efektif melalui pendekatan *boosting*. Selain itu, kombinasi antara *Decision Tree* dan *Bagging* menghasilkan akurasi tertinggi dalam klasifikasi penyakit stroke, yakni sebesar 96,91%. Kombinasi lainnya seperti *Decision Tree* dengan *AdaBoost* serta *Naive Bayes* dengan *AdaBoost* juga menunjukkan performa yang kompetitif. Temuan ini menggarisbawahi peran penting *ensemble learning*, baik *bagging* maupun *boosting*, dalam meningkatkan akurasi, generalisasi, dan stabilitas model prediksi, khususnya di bidang kesehatan yang menuntut ketepatan tinggi dalam diagnosis (Mukaromah dkk., 2024).

Penelitian lain yang mengimplementasikan pendekatan *ensemble learning* juga menunjukkan hasil yang menjanjikan dalam konteks deteksi anomali kesehatan janin pada ibu hamil. Dalam studi tersebut, penggunaan algoritma *decision tree* tanpa pendekatan *ensemble* hanya mampu mencapai tingkat akurasi sebesar

89,80%. Namun, setelah diterapkan teknik *ensemble* berbasis *decision tree*, akurasi model meningkat secara signifikan menjadi 92,66%. Peningkatan akurasi sebesar 2,86% ini mengindikasikan bahwa metode *ensemble learning* mampu meningkatkan kemampuan model dalam melakukan klasifikasi secara lebih akurat dan andal. Hasil ini juga menegaskan bahwa pendekatan *ensemble* dapat mengurangi risiko kesalahan prediksi yang dapat berdampak serius, khususnya dalam konteks medis yang berkaitan dengan keselamatan ibu dan janin. Oleh karena itu, penerapan teknik *ensemble learning* tidak hanya memberikan keuntungan dari segi performa numerik, tetapi juga berkontribusi dalam peningkatan kualitas pengambilan keputusan di bidang kesehatan berbasis data (Aziz, 2025).

Dalam penelitian lainnya, penerapan algoritma *Extreme Gradient Boosting* (XGBoost) sebagai bagian dari pendekatan *ensemble learning* terbukti memberikan dampak signifikan dalam proses kategorisasi produk pada platform *e-commerce*. Teknik ini dinilai mampu meningkatkan akurasi klasifikasi serta efisiensi dalam proses otomatisasi pengelompokan produk yang sangat krusial dalam skala bisnis digital. XGBoost, yang dikenal dengan kemampuannya dalam mengatasi *overfitting* dan mempercepat proses pelatihan, efektif digunakan untuk menangani kompleksitas fitur produk yang beragam dalam sistem *e-commerce*. Hasil evaluasi menunjukkan bahwa model yang dikembangkan berhasil mencapai tingkat akurasi sebesar 97,17%, yang mencerminkan performa tinggi dari algoritma tersebut. Capaian ini tidak hanya menandakan ketepatan model dalam mengklasifikasikan data, tetapi juga menunjukkan potensi besar dalam meningkatkan efisiensi operasional, memperbaiki struktur navigasi katalog produk, serta secara langsung berkontribusi terhadap peningkatan profitabilitas bisnis dan kualitas pengalaman pengguna dalam transaksi digital. Dengan demikian, XGBoost terbukti sebagai solusi *ensemble* yang unggul untuk permasalahan klasifikasi produk di ranah perdagangan daring (Sepbriant dkk, 2024).

Penelitian ini memanfaatkan pendekatan *ensemble learning*, khususnya jenis *boosting* yang mencakup algoritma *Extreme Gradient Boosting* (XGBoost), *AdaBoost*, *Gradient Boosting*, dan *LightGBM*. Tujuan utama dari penggunaan

pendekatan ini adalah untuk meningkatkan performa prediktif dari masing-masing algoritma dengan memanfaatkan kekuatan kolektif dari beberapa model. Dalam proses pengujian, digunakan teknik *majority voting* untuk menggabungkan hasil prediksi dari keempat model, sehingga diperoleh prediksi akhir yang lebih andal dan akurat. Berdasarkan hasil evaluasi yang dilakukan, algoritma XGBoost menunjukkan performa paling unggul dengan tingkat akurasi mencapai 92,30%, mengungguli teknik *boosting* lainnya yang digunakan dalam penelitian ini. Temuan ini memperkuat reputasi XGBoost sebagai algoritma *ensemble* yang tidak hanya efisien secara komputasi, tetapi juga konsisten dalam menghasilkan akurasi tinggi pada berbagai jenis data dan skenario prediksi (Rama dkk., 2024).

Penelitian terkait klasifikasi penyakit ginjal kronis (PGK) menunjukkan bahwa penggunaan algoritma *Extreme Gradient Boosting* (XGBoost), sebagai salah satu metode *ensemble learning* berbasis *boosting*, mampu memberikan performa klasifikasi yang sangat tinggi. Studi ini menggunakan 400 data pasien yang dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian. Untuk mengoptimalkan kinerja model, dilakukan proses *hyperparameter tuning* menggunakan teknik *grid search* terhadap lima parameter utama, yaitu *n_estimators*, *max_depth*, *learning_rate*, *subsample*, dan *colsample_bytree*. Hasil evaluasi menggunakan *confusion matrix* menunjukkan bahwa model XGBoost yang telah dioptimalkan berhasil mencapai akurasi sebesar 99,16%, precision 98,17%, recall 99,16%, dan f1-score 99,16%. Capaian metrik yang hampir sempurna ini membuktikan bahwa XGBoost, ketika dikonfigurasi dengan tepat, sangat efektif dalam mengklasifikasikan pasien penderita PGK dan non-PGK. Selain itu, hasil tersebut juga menegaskan potensi besar dari metode *ensemble* berbasis *boosting* dalam mendukung sistem pengambilan keputusan di bidang kesehatan yang menuntut tingkat akurasi dan sensitivitas yang tinggi (Soelistijadi dkk., 2024).

Salah satu penelitian membahas penerapan algoritma klasifikasi dalam diagnosis kanker payudara, yang merupakan salah satu penyakit mematikan dengan prevalensi tinggi pada wanita di seluruh dunia. Studi ini mengevaluasi performa beberapa algoritma klasifikasi seperti *Decision Tree*, *Naïve Bayes*, dan *K-Nearest Neighbor* (KNN), yang kemudian dikombinasikan dengan pendekatan *ensemble*

learning seperti *AdaBoost* dan *Bagging* guna meningkatkan akurasi prediksi. Hasil penelitian menunjukkan bahwa pendekatan *ensemble* mampu meningkatkan performa klasifikasi secara signifikan dibandingkan dengan penggunaan algoritma tunggal. Kombinasi terbaik diperoleh dari *Decision Tree* yang dikombinasikan dengan metode *ensemble*, menghasilkan akurasi sebesar 82,76%. Sementara itu, nilai *Area Under Curve* (AUC) tertinggi, yaitu sebesar 0,950 yang masuk dalam kategori sangat baik dicapai oleh kombinasi *KNN* dan *Bagging*. Temuan ini menegaskan bahwa teknik *ensemble learning* tidak hanya meningkatkan ketepatan prediksi, tetapi juga memperkuat keandalan model dalam konteks diagnosis penyakit kritis seperti kanker payudara, sehingga sangat potensial untuk diterapkan dalam sistem pendukung keputusan di bidang medis (Majid & Nawangsih, 2024).

Penelitian ini mengevaluasi efektivitas algoritma *Extreme Gradient Boosting* (XGBoost), sebagai salah satu metode *ensemble learning*, dalam mendeteksi penipuan pada transaksi kartu kredit. Fokus utama penelitian adalah membandingkan performa XGBoost dengan algoritma *machine learning* tradisional dalam konteks data yang tidak seimbang, yang umum dijumpai dalam kasus deteksi penipuan. Menggunakan dataset publik yang representatif, hasil evaluasi menunjukkan bahwa XGBoost secara konsisten mengungguli algoritma lainnya dalam berbagai metrik performa, termasuk akurasi, presisi, *recall*, dan *Area Under Curve* (AUC). Secara khusus, XGBoost berhasil mencapai akurasi hingga 96%, yang menunjukkan kemampuannya dalam mengenali pola-pola penipuan secara efektif meskipun dihadapkan pada distribusi data yang tidak seimbang. Temuan ini menjadi sangat relevan dalam konteks pertumbuhan pesat transaksi daring pasca pandemi Covid-19, yang secara bersamaan meningkatkan risiko penipuan digital. Oleh karena itu, penggunaan model prediktif yang efisien dan akurat seperti XGBoost dinilai sangat krusial bagi institusi keuangan dalam upaya mitigasi risiko dan perlindungan aset digital. Selain itu, hasil studi ini membuka peluang pengembangan sistem deteksi penipuan berbasis model *hibrida*, adaptif, dan *real-time* yang lebih responsif terhadap dinamika data transaksi di masa mendatang (Mohbey dkk., 2022).

Penelitian ini mengembangkan model deteksi kecurangan laporan keuangan dengan memanfaatkan data laporan keuangan perusahaan yang tersedia secara publik di wilayah Timur Tengah dan Afrika Utara (MENA). Model prediktif tersebut dirancang menggunakan pendekatan *ensemble learning*, khususnya algoritma *Extreme Gradient Boosting* (XGBoost), yang dikenal memiliki kemampuan tinggi dalam menangani data yang kompleks dan bersifat non-linear. Selain XGBoost, penelitian ini juga membandingkan performa beberapa algoritma *machine learning* lainnya seperti *Logistic Regression*, *Decision Tree*, *Support Vector Machine* (SVM), *AdaBoost*, dan *Random Forest*. Hasil evaluasi menunjukkan bahwa XGBoost menghasilkan performa terbaik di antara seluruh model yang diuji. Setelah dilakukan proses optimasi terhadap parameter-parameter utama, model XGBoost mampu mencapai tingkat akurasi akhir sebesar 96,05% dalam mengidentifikasi praktik kecurangan dalam laporan keuangan perusahaan. Temuan ini menegaskan bahwa XGBoost tidak hanya unggul secara teknis, tetapi juga sangat aplikatif untuk diterapkan dalam konteks audit berbasis data dan sistem pendeteksian anomali di sektor keuangan, yang memerlukan tingkat presisi tinggi dan keandalan dalam klasifikasi (Ali dkk., 2023).

Dalam penelitian ini, pendekatan *ensemble learning* diterapkan untuk meningkatkan akurasi klasifikasi dalam mendeteksi keterlambatan perkembangan motorik pada bayi. Teknik ini menggabungkan beberapa model pembelajaran mesin, antara lain *Bagging Tree* dengan akurasi sebesar 90%, *Boosting Tree* dengan akurasi 92%, serta metode *ensemble* lainnya yang mencapai akurasi 91%. Tujuan utama dari penggabungan model-model ini adalah untuk mengurangi bias dan variansi model tunggal, serta meningkatkan konsistensi dalam mengenali pola-pola pergerakan yang mencerminkan perkembangan motorik bayi. Dengan memanfaatkan kekuatan kolektif dari beberapa algoritma klasifikasi, metode *ensemble learning* terbukti mampu menghasilkan prediksi yang lebih stabil, andal, dan efektif. Hasil penelitian ini menunjukkan bahwa pendekatan *ensemble* tidak hanya relevan dalam bidang keuangan dan kesehatan, tetapi juga sangat potensial untuk diterapkan dalam bidang tumbuh kembang anak, khususnya untuk deteksi dini gangguan motorik. Deteksi yang lebih akurat sejak dini dapat mendukung

intervensi lebih cepat dan tepat, yang pada akhirnya berkontribusi pada peningkatan kualitas hidup anak (Sindar dkk., 2024).

Tabel 2. 1 Tinjauan Pustaka

Object	Metode	Dataset	Hasil
Prediksi waktu pengantian baterai aki kendaraan	<i>Bagging</i> dan <i>Boosting</i>	Data historis penggantian beterei aki mobil Toyota di satu perusahaan otomotif di Kota Makassar.	<i>Bagging</i> MSE = 145,448. <i>Boosting</i> : MAPE = 11,56%, MAE = 43,80, RMSE = 38,760.
Prediksi penyakit stroke	<i>Bagging</i> + <i>Decision Tree</i> , <i>AdaBoost</i> + <i>Decision Tree</i> , <i>AdaBoost</i> + <i>Naive Bayes</i> .	Dataset penyakit stroke	Akurasi tertinggi: <i>Decision Tree</i> + <i>Bagging</i> = 96,91%.
Deteksi anomali kesehatan janin	<i>Decision Tree</i> + <i>Ensemble</i>	Data kesehatan pada janin,	<i>Decision Tree</i> tunggal : 89,80%. Dengan <i>Ensemble</i> : 92,66% kenaikan 2,86%.
Kategorisasi produk <i>e-commerce</i>	XGBoost	Data produk <i>e-commerce</i>	XGBoost Akurasi: 97,17%.
Ulasan Pelanggan <i>e-commerce</i>	<i>Boosting</i> : <i>XGBoost</i> , <i>AdaBoost</i> , <i>Gradient Boosting</i> ,	Kumpulan ulasan pengguna yang fokus pada produk <i>headphone</i>	Akurasi tertinggi pada XGBoost: 92,30%.

	<i>LightGBM + Majority Voting</i>		
Klasifikasi Penyakit Ginjal Kronis (PGK)	XGBoost (<i>Ensemble</i>)	Data pasien PGK	Akurasi: 99,16%; Precision: 98,17%; Recall: 99,16%; F1-score: 99,16%
Deteksi kanker payudara	<i>Decision Tree, Naïve Bayes, KNN + Bagging & AdaBoost</i>	<i>Data breast cancer</i>	Akurasi tertinggi: 82,76%; AUC tertinggi: 0,950 (KNN + <i>Bagging</i>)
Deteksi penipuan transaksi kartu kredit	XGBoost (<i>Ensemble</i>)	Data transaksi kartu kredit (publik)	Akurasi: 96%; performa unggul pada precision, recall, dan AUC
Deteksi kecurangan laporan keuangan	XGBoost, <i>Logistic Regression, SVM, AdaBoost, Random Forest, Decision Tree</i>	Laporan keuangan perusahaan di wilayah MENA	XGBoost akurasi akhir: 96,05%
Deteksi keterlambatan motorik pada bayi	<i>Bagging Tree, Boosting Tree, Ensemble lainnya</i>	Data dari Satudata.com yang memuat empat fase perkembangan motorik bayi usia 0 –	Akurasi: <i>Bagging Tree</i> (90%), <i>Boosting Tree</i> (92%), <i>Ensemble lainnya</i> (91%).

		12 bulan.	
--	--	-----------	--

Kesimpulan dari tinjauan pustaka ini menunjukkan bahwa *ensemble learning* secara konsisten mampu meningkatkan performa model prediksi dibandingkan model tunggal di berbagai bidang, seperti kesehatan, *e-commerce*, keuangan, hingga otomotif. Teknik seperti *bagging* dan *boosting* seperti XGBoost, memberikan akurasi yang lebih tinggi, menurunkan tingkat kesalahan, serta meningkatkan stabilitas prediksi. Temuan ini mendukung penerapan *ensemble learning* dalam prediksi *Global Horizontal Irradiance* (GHI) di wilayah Pulau Jawa.

2.2 Dasar Teori

2.2.1 Radiasi Surya



Gambar 2. 1 Radiasi Surya

Radiasi surya merupakan bentuk energi elektromagnetik yang berasal dari proses fusi nuklir di inti matahari. Energi ini dipancarkan dalam bentuk gelombang elektromagnetik yang terdiri dari komponen medan listrik dan medan magnet yang saling tegak lurus, serta merambat melalui ruang hampa hingga mencapai permukaan bumi. Radiasi surya menjadi sumber energi utama bagi kehidupan di bumi dan memainkan peran penting dalam berbagai proses atmosfer dan lingkungan.

Dalam konteks energi terbarukan, pemahaman terhadap intensitas radiasi surya sangatlah penting. Salah satu parameter utama yang digunakan adalah *Global*

Horizontal Irradiance (GHI), yaitu jumlah energi matahari yang diterima per satuan luas pada permukaan horizontal bumi, yang dinyatakan dalam satuan kilowatt-jam per meter persegi (kWh/m^2). GHI terdiri atas dua komponen utama, yaitu *Direct Normal Irradiance* (DNI) yang merupakan radiasi langsung dari matahari tanpa hambatan atmosfer, serta *Diffuse Horizontal Irradiance* (DHI) yang merupakan radiasi yang telah mengalami hamburan oleh partikel atmosfer seperti debu, aerosol, dan uap air keduanya juga diukur dalam kWh/m^2 . Ketiga komponen ini membentuk dasar penting dalam perancangan sistem pembangkit tenaga surya, pengukuran potensi energi matahari, serta kajian klimatologi (Deqita, 2022).

Namun demikian, prediksi nilai GHI merupakan tantangan yang tidak sederhana. Variasi posisi matahari, tutupan awan, musim, kelembapan, dan tingkat polusi menyebabkan nilai GHI sangat dinamis dan bersifat nonlinier. Oleh karena itu, metode konvensional berbasis fisika sering kali kurang akurat dalam memodelkan fluktuasi tersebut. Dalam hal ini, pendekatan berbasis pembelajaran mesin seperti *ensemble learning* menjadi solusi yang menjanjikan. Algoritma ini mampu menggabungkan keunggulan beberapa model sekaligus untuk menangkap pola-pola kompleks dalam data historis GHI, sehingga meningkatkan akurasi prediksi dan mendukung pemanfaatan energi surya secara optimal.

2.2.2 Ensemble Learning

Ensemble learning adalah teknik dalam pembelajaran mesin yang menggabungkan beberapa model prediktif untuk memperoleh hasil yang lebih akurat dan stabil dibandingkan dengan penggunaan satu model tunggal. Pendekatan ini bertujuan untuk mengurangi kelemahan dari masing-masing model individu dengan memanfaatkan kelebihan yang dimiliki oleh setiap model. Konsep dasar *ensemble* merujuk pada penyatuan beberapa metode pembelajaran yang saling melengkapi, sehingga mampu meningkatkan performa dalam menyelesaikan masalah klasifikasi maupun regresi yang kompleks (Rayadin dkk., 2024).

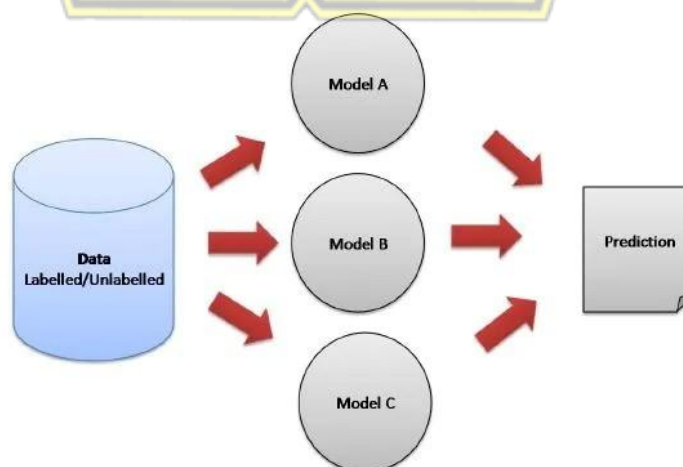
Terdapat tiga pendekatan utama dalam *ensemble learning*, yaitu *bagging*, *boosting*, dan *stacking* :

- *Bagging* bekerja dengan cara membangun banyak model yang dilatih secara paralel pada subset data yang diambil secara acak dengan pengembalian

(*bootstrap*). Hasil dari semua model tersebut kemudian digabungkan melalui *voting* (klasifikasi) atau rata-rata (regresi) untuk menghasilkan prediksi akhir. Pendekatan ini efektif dalam mengurangi *variance* dan meningkatkan stabilitas model.

- *Boosting* adalah metode *ensemble* yang membangun model secara berurutan. Setiap model baru berfokus untuk memperbaiki kesalahan prediksi dari model sebelumnya. Teknik ini mampu mengurangi bias dan meningkatkan akurasi, terutama pada data yang sulit dipelajari.
- *Stacking* berbeda dari dua metode sebelumnya, menggabungkan beberapa model dengan jenis yang berbeda dan menggunakan model *meta-learner* untuk belajar dari output model-model sebelumnya, sehingga meningkatkan kemampuan generalisasi.

Meskipun *ensemble learning* pada dasarnya terdiri dari sekumpulan model alternatif yang jumlahnya terbatas, pendekatan ini memungkinkan terbentuknya struktur prediksi yang lebih fleksibel dan adaptif dibandingkan dengan model tunggal. Evaluasi prediksi dari model *ensemble* biasanya memerlukan sumber daya komputasi yang lebih tinggi. Namun, peningkatan kinerja yang dihasilkan sering kali mampu mengimbangi keterbatasan dari model pembelajaran yang lemah. Dengan demikian, penggunaan beberapa model dalam *ensemble* tidak hanya memperkuat akurasi, tetapi juga menjadikan sistem prediksi lebih tangguh dalam menghadapi data berdimensi tinggi dan kompleks (Rifai dkk., 2024).



Gambar 2. 2 Arsitektur *Ensemble Learning*

Gambar 2.2 menampilkan arsitektur dasar dari *ensemble learning*, proses dimulai dari data masukan, baik yang telah berlabel maupun belum, yang kemudian dialirkan ke beberapa model pembelajaran (Model A, B, dan C). Masing-masing model dilatih secara independen menggunakan data tersebut dan menghasilkan prediksi sendiri-sendiri. Selanjutnya, hasil prediksi dari seluruh model dikombinasikan melalui mekanisme agregasi, seperti *voting* pada klasifikasi atau rata-rata pada regresi, untuk menghasilkan prediksi akhir yang lebih akurat dan stabil. Pendekatan ini dirancang untuk memaksimalkan kelemahan yang mungkin dimiliki oleh model individual. Secara keseluruhan *ensemble learning* tidak hanya meningkatkan kinerja model, tetapi juga memberikan solusi yang lebih efektif (Hanif & Utomo, 2025).

2.2.3 Bagging (*Bootstrap Aggregating*)

Bagging (*Bootstrap Aggregating*) merupakan salah satu metode dalam *ensemble learning* yang digunakan untuk meningkatkan akurasi dan kestabilan prediksi dengan cara menggabungkan beberapa model dasar (*base learners*) yang dilatih secara paralel. Teknik ini menggunakan pendekatan *bootstrap*, yaitu mengambil sampel data secara acak dengan pengembalian untuk membentuk beberapa subset data pelatihan. Masing-masing subset digunakan untuk melatih model yang sama (biasanya *decision tree*), dan hasil prediksi akhir diperoleh melalui proses agregasi, seperti menggunakan *voting* untuk klasifikasi atau rata-rata untuk regresi (Algiffary & Sutabri, 2023).

Rumus prediksi pada metode *bagging* dituliskan sebagai berikut :

$$\hat{y} = \frac{1}{M} \sum_{m=1}^M h_m(x) \dots \dots \dots (1)$$

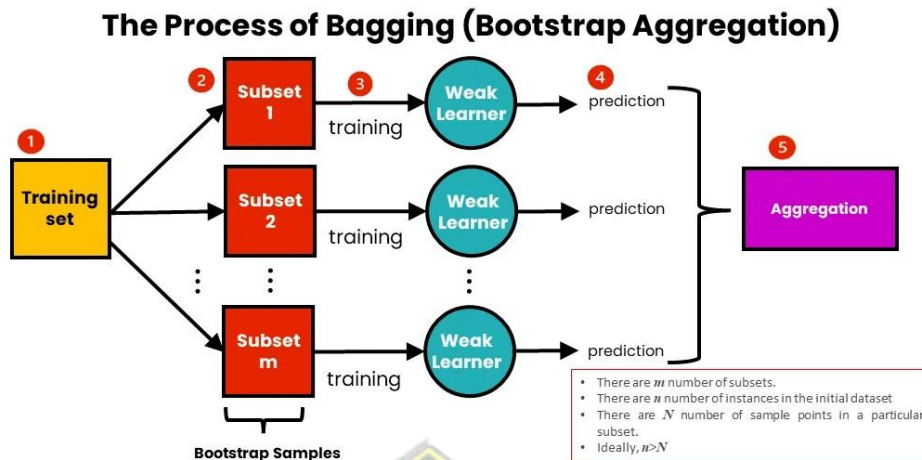
Keterangan:

$\hat{y}$: prediksi akhir

M : jumlah model (*weak learners*)

$h_m(x)$: prediksi dari model ke M untuk input $\hat{y}$

Kelebihan bagging adalah kemampuannya menurunkan varians model dan menangani *overfitting*. Namun, metode ini kurang efektif dalam mengurangi bias dari model dasar.



Gambar 2. 3 Arsitektur Bagging (*Bootstrap Aggregation*)

Gambar 2.3 memperlihatkan arsitektur dasar dari metode *bagging*, yang terdiri dari dua tahap utama, yaitu *bootstrap* dan *aggregating*. Pada tahap *bootstrap*, beberapa model dasar dilatih secara independen menggunakan subset data yang berbeda, dihasilkan melalui proses sampling ulang secara acak dari data asli. Variasi data ini memungkinkan tiap model menangkap pola yang berbeda, sehingga menciptakan keanekaragaman model. Selanjutnya, pada tahap *aggregating*, semua hasil prediksi dari model tersebut digabungkan untuk menghasilkan prediksi akhir. Penggabungan ini dilakukan oleh model akhir yang bertugas menyatukan output dari seluruh *base model*, sehingga menghasilkan prediksi yang lebih stabil dan akurat (Helmut & Triantoro Murdiansyah, 2020).

2.2.4 Decision Tree

Algoritma *decision tree* termasuk dalam kelompok algoritma pembelajaran terawasi. Berbeda dengan beberapa algoritma *supervised learning* lainnya, metode ini mampu menangani baik tugas klasifikasi maupun regresi. Tujuan utama dari pohon keputusan adalah membangun model yang mampu memprediksi label atau nilai target dengan cara mempelajari serangkaian aturan keputusan yang diperoleh dari data pelatihan sebelumnya dalam proses prediksi, algoritma pohon keputusan dimulai dari simpul akar. Nilai atribut pada simpul akar dibandingkan dengan nilai atribut pada data yang akan diprediksi. Berdasarkan hasil perbandingan tersebut, algoritma akan mengikuti cabang yang sesuai dan berpindah ke simpul selanjutnya hingga mencapai prediksi akhir (Siregar, 2020).

Secara matematis, hasil prediksi akhir dari model *bagging* (yang seringkali menggunakan *decision tree* sebagai *base estimator*) dapat dirumuskan sebagai berikut :

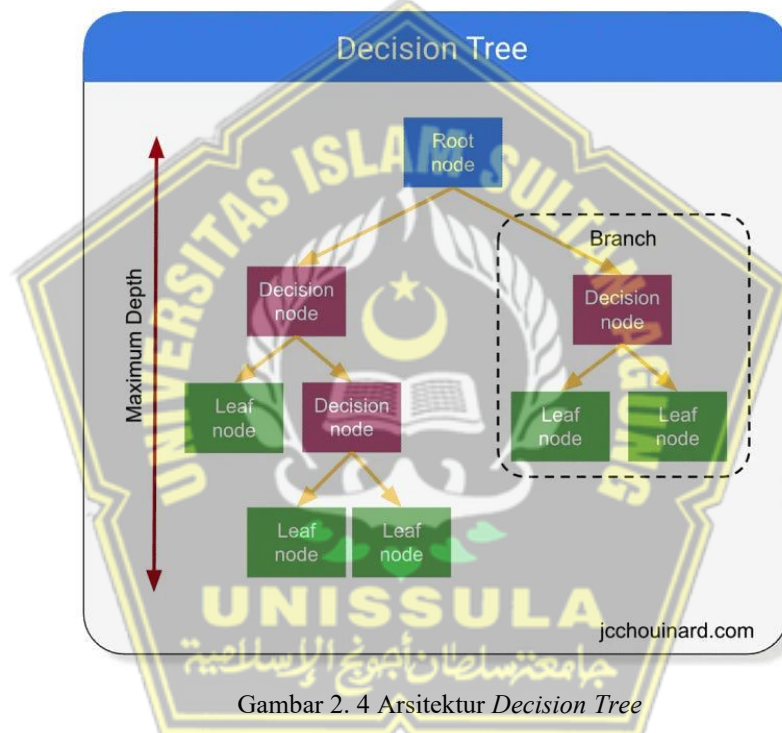
$$f(x) = \frac{1}{B} \sum_{b=1}^B f_b(x) \dots \dots \dots (2)$$

Keterangan :

$f(x)$: hasil prediksi akhir dari model *bagging*

B : jumlah pohon (*estimators*)

$f_b(x)$: hasil prediksi dari *decision tree* ke B



Gambar 2. 4 Arsitektur *Decision Tree*

Gambar 2.4 menggambarkan arsitektur dasar algoritma *decision tree*, yang terdiri dari *root node* sebagai titik awal pembagian data, *decision node* sebagai percabangan berdasarkan fitur tertentu, dan *leaf node* sebagai *output* akhir. Struktur ini membentuk cabang (*branch*) hingga mencapai kedalaman maksimum (*maximum depth*) yang memengaruhi kompleksitas dan generalisasi model. Meskipun fleksibel dan mudah dipahami, *decision tree* rentan *overfitting* jika tidak dikendalikan dengan parameter seperti kedalaman pohon atau *pruning* (Mia dkk., 2024).

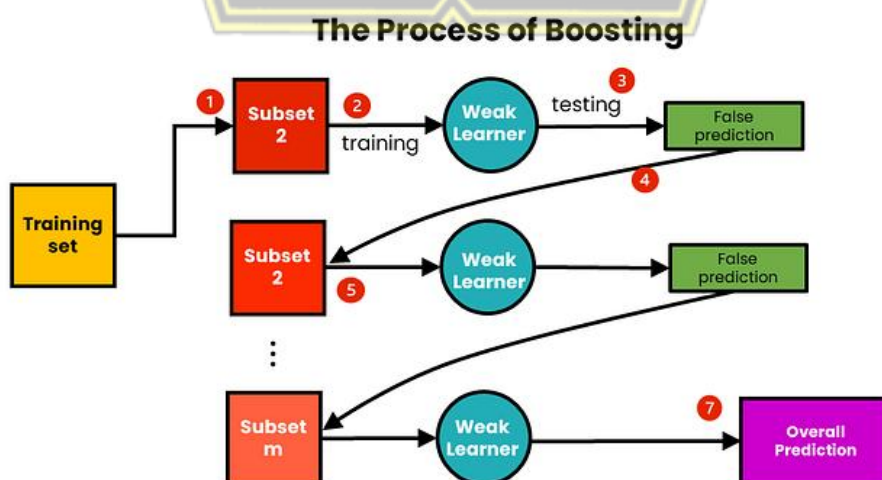
2.2.5 Boosting

Boosting merupakan salah satu metode dalam *ensemble learning* yang bertujuan untuk meningkatkan akurasi model prediktif dengan menggabungkan sejumlah *weak learners* secara berurutan, berbeda dengan *bagging* yang melatih model secara paralel, *boosting* membangun model secara bertahap. Setiap model baru difokuskan untuk memperbaiki kesalahan prediksi yang dilakukan oleh model sebelumnya, sehingga kinerja keseluruhan meningkat secara progresif (Sindar dkk., 2024).

Boosting terutama berfokus pada pengurangan bias dalam model, ketimbang pada variasi, dengan cara secara bertahap meningkatkan *weak learners* awal yang memiliki bias tinggi. Secara keseluruhan, pendekatan *boosting* adalah metode yang efektif untuk mengurangi kesalahan dalam analisis data prediktif dan meningkatkan kinerja model, khususnya pada data yang kompleks atau sulit diklasifikasikan, serta dapat bekerja baik dengan berbagai algoritma (Sepbriant & Utomo, 2024).

Beberapa algoritma yang populer dalam implementasi *boosting* antara lain:

- AdaBoost* : Menggabungkan model-model lemah dengan bobot yang disesuaikan berdasarkan kesalahan prediksi sebelumnya.
- Gradient Boosting* : Membangun model secara berurutan dengan pendekatan fungsi gradien untuk meminimalkan fungsi loss.
- XGBoost : Merupakan pengembangan dari *gradient boosting* dengan efisiensi komputasi tinggi, regularisasi, dan kontrol *overfitting* yang lebih baik.



Gambar 2. 5 Arsitektur *Boosting*

Gambar 2.5 proses *boosting*, di mana beberapa model lemah (*weak learners*) dibentuk secara berurutan. Setiap model dilatih menggunakan subset data yang disesuaikan berdasarkan kesalahan prediksi dari model sebelumnya. Kesalahan ini kemudian diperhatikan dan digunakan untuk memfokuskan pelatihan model berikutnya agar lebih baik dalam menangani data yang sulit. Setelah seluruh model terbentuk, prediksi akhir dihasilkan dengan menggabungkan kontribusi dari semua model, sehingga menghasilkan prediksi yang lebih akurat dan andal.

2.2.6 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost), adalah algoritma *supervised learning* yang mengimplementasikan pendekatan *ensemble learning* dan dikembangkan berdasarkan metode *Gradient Boosting*. Dalam *Gradient Boosting*, sejumlah *weak learner* biasanya berupa pohon keputusan (*decision tree*) digabungkan untuk membentuk model prediktif yang lebih kuat. XGBoost sendiri merupakan sistem *tree boosting* yang telah banyak digunakan dalam berbagai penelitian untuk menyelesaikan beragam permasalahan, baik dalam bentuk klasifikasi maupun regresi. Beberapa keunggulan utama dari XGBoost meliputi fitur *regularization*, pemangkasan pohon (*tree pruning*), kemampuan untuk melakukan pemrosesan paralel, serta dukungan terhadap data numerik maupun kategorikal (Murdiansyah, 2024).

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \dots \dots \dots (3)$$

Keterangan:

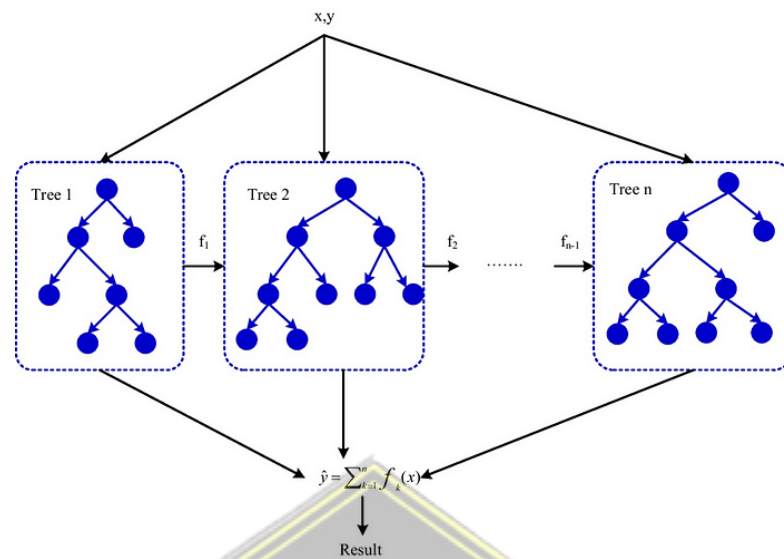
$\hat{y}_i$ = prediksi akhir untuk data ke i

K = jumlah total *tree*

f_k = model (fungsi) ke K

x_i = input data ke i

$\mathcal{F}$ = ruang fungsi semua pohon keputusan regresi



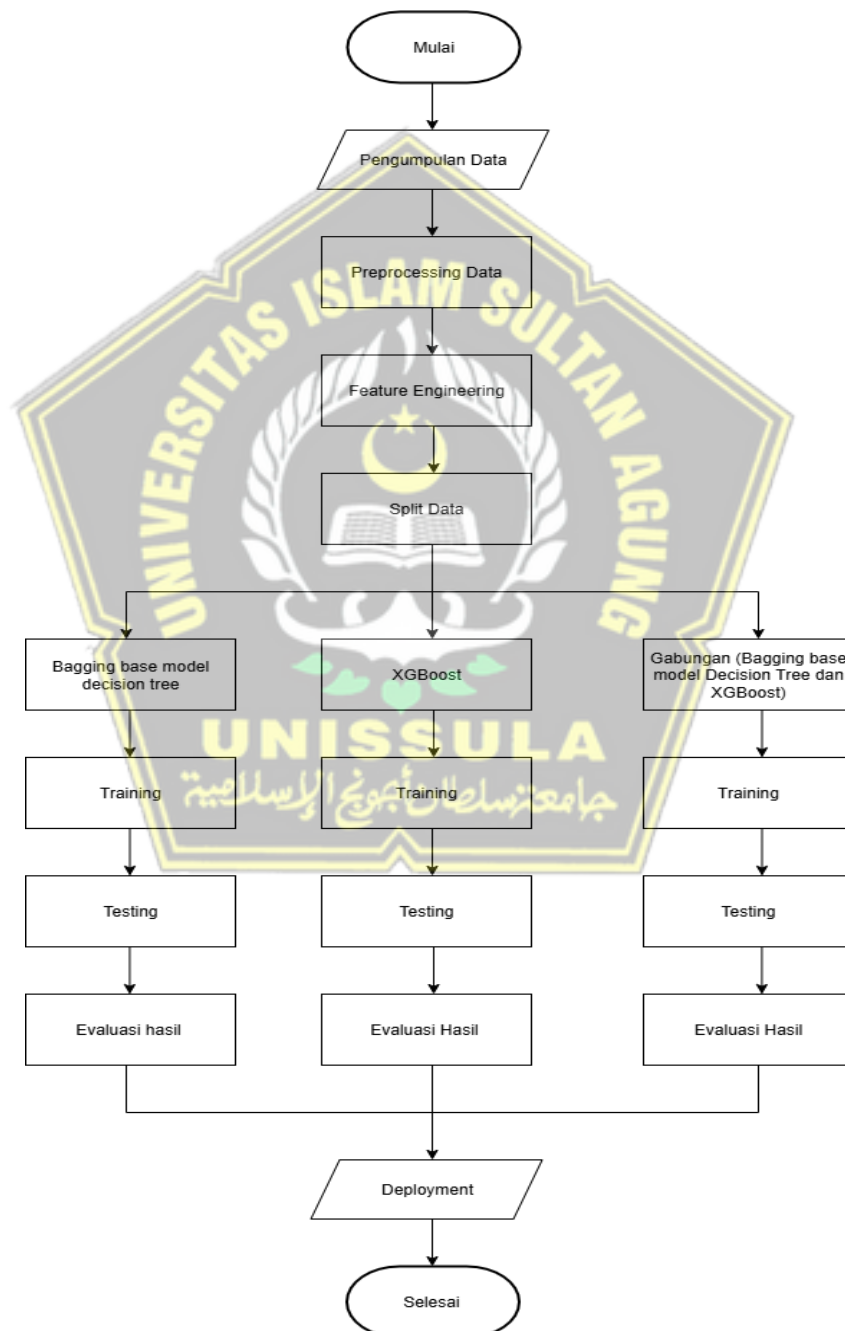
Gambar 2. 6 Arsitektur XGBoost

Pada Gambar 2.6 adalah arsitektur XGBoost, Proses kerja dimulai dengan menetapkan model awal yang memberikan prediksi konstan (biasanya berupa rata-rata dari variabel target) untuk seluruh data. Selanjutnya, model menghitung *residual*, yaitu selisih antara nilai prediksi awal dan nilai aktual target pada setiap data. XGBoost kemudian membangun pohon keputusan (*decision tree*) secara bertahap, di mana setiap pohon yang baru ditujukan untuk mengoreksi kesalahan yang dihasilkan oleh pohon sebelumnya. Proses ini berlangsung secara iteratif hingga memenuhi kriteria penghentian, seperti tercapainya jumlah maksimum pohon atau tidak lagi terjadi peningkatan signifikan dalam perbaikan *residual* (Sudarman & Budi, 2023).

BAB III

METODOLOGI PENELITIAN

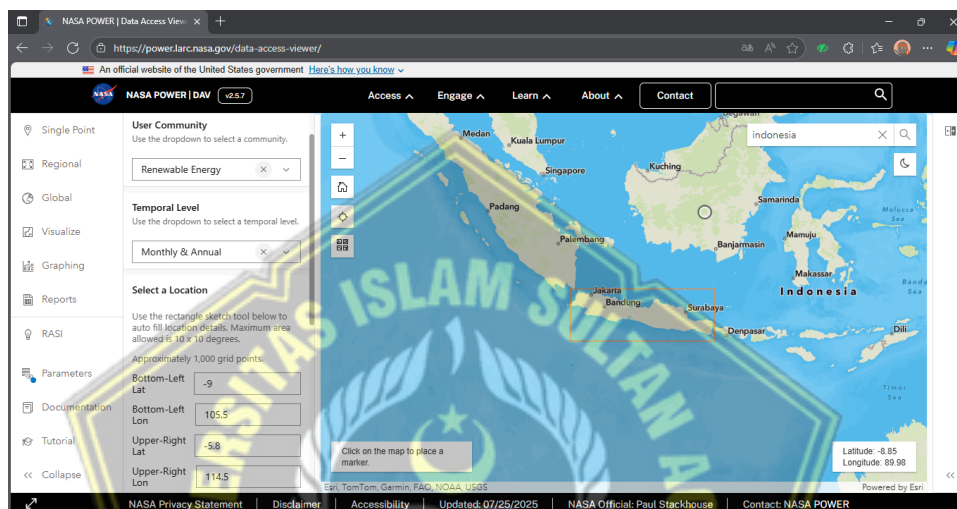
Bagian ini akan membahas mengenai langkah-langkah yang akan digunakan dalam penelitian. Langkah pertama yaitu pengumpulan data, *preprocessing data*, *Feature Engineering*, *split data*, dilanjutkan pemodelan, pengujian diakhiri dengan penarikan kesimpulan sebagai hasil akhir.



Gambar 3. 1 *Flowchart* Penelitian

3.1 Pengumpulan Data

Pada proses ini adalah pengumpulan data yang diperlukan oleh peneliti. Sumber data pada penelitian ini menggunakan data historis radiasi surya permukaan di Pulau Jawa *All_Sky_Sfc_Sw_Dwn*, data yang diperoleh dari sumber Nasa Power atau bisa diakses di [POWER | DAV](https://power.larc.nasa.gov/data-access-viewer/). Data historis yang digunakan berisi informasi berupa data *paramater, year, lat, lon*, bulan Januari hingga Desember dan Ann.



Gambar 3. 2 Pengambilan Data Pulau Jawa di Nasa Power | DAV

PARAMETER	YEAR	LAT	LON	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANN
ALLSKY_SFC_SW	1984	-6.5	105.5	4.5	4.35	4.81	5.07	4.78	4.94	4.21	5.15	5.2	5.16	5.56	4.14	4.86
ALLSKY_SFC_SW	1984	-6.5	106.5	4.49	4.46	4.82	4.89	4.84	4.68	3.86	4.8	4.54	4.99	5.42	4.46	4.69
ALLSKY_SFC_SW	1984	-6.5	107.5	4.39	4.68	5.13	4.94	4.61	4.79	4.08	5.1	4.78	5.2	5.6	4.67	4.83
ALLSKY_SFC_SW	1984	-6.5	108.5	4.06	4.88	5.45	5.0	4.49	4.8	4.3	5.22	5.01	5.45	5.41	4.67	4.89
ALLSKY_SFC_SW	1984	-6.5	109.5	4.14	4.93	5.51	5.09	4.86	4.93	4.63	5.57	5.36	5.72	5.37	4.22	5.03
ALLSKY_SFC_SW	1984	-6.5	110.5	3.93	4.49	5.24	4.91	4.84	5.05	4.49	5.72	5.59	6.08	5.64	4.05	5.0
ALLSKY_SFC_SW	1984	-6.5	111.5	4.41	4.47	5.28	4.78	4.66	4.86	4.59	5.84	5.56	5.92	5.42	4.43	5.02
ALLSKY_SFC_SW	1984	-6.5	112.5	4.32	4.87	5.6	4.77	4.94	5.09	4.81	5.98	5.88	6.13	5.75	4.33	5.21
ALLSKY_SFC_SW	1984	-6.5	113.5	4.13	4.7	5.49	4.64	4.91	5.07	4.8	6.05	5.71	6.34	5.84	4.17	5.16
ALLSKY_SFC_SW	1984	-6.5	114.5	4.16	4.82	5.54	4.74	4.79	5.07	4.44	5.99	5.59	6.7	5.93	4.11	5.16
ALLSKY_SFC_SW	1984	-7.5	105.5	4.05	4.74	5.26	5.03	4.78	4.86	4.14	5.1	4.66	5.18	5.75	4.59	4.84
ALLSKY_SFC_SW	1984	-7.5	106.5	4.32	5.04	5.36	5.16	4.88	5.02	4.41	5.12	4.68	5.22	5.84	4.9	5.04
ALLSKY_SFC_SW	2024	-8.5	107.5	5.28	5.79	4.88	5.18	5.24	4.3	5.21	5.7	6.2	6.2	5.02	4.41	5.28
ALLSKY_SFC_SW	2024	-8.5	108.5	5.24	5.74	5.02	5.23	5.25	4.43	5.15	5.64	6.12	6.36	5.12	4.72	5.33
ALLSKY_SFC_SW	2024	-8.5	109.5	5.24	5.71	5.24	5.37	5.35	4.74	5.2	5.76	6.25	6.45	5.3	4.82	5.45
ALLSKY_SFC_SW	2024	-8.5	110.5	5.48	5.88	5.37	5.36	5.41	4.86	5.27	5.94	6.36	6.48	5.43	4.73	5.55
ALLSKY_SFC_SW	2024	-8.5	111.5	5.7	5.93	5.38	5.59	5.33	4.72	5.24	5.85	6.28	6.25	5.5	4.77	5.54
ALLSKY_SFC_SW	2024	-8.5	112.5	5.69	5.96	5.36	5.45	5.35	4.73	5.21	5.91	6.25	6.44	5.64	4.67	5.55
ALLSKY_SFC_SW	2024	-8.5	113.5	5.74	5.85	5.42	5.21	5.41	4.76	5.2	5.89	6.29	6.62	5.98	4.58	5.58
ALLSKY_SFC_SW	2024	-8.5	114.5	5.59	5.8	5.3	5.02	5.19	4.58	5.02	5.68	6.14	6.7	5.98	4.18	5.43

Gambar 3. 3 Data Historis GHI di Pulau Jawa ((1984-2024)

3.2 Preprocessing Data

Tahapan ini adalah proses mengolah data mentah menjadi data yang siap digunakan untuk pemodelan dengan tujuan untuk menghasilkan data yang akurat. Data *preprocessing* meliputi tahapan sebagai berikut :

3.2.1 Data Cleaning

Data cleaning adalah proses yang dilakukan untuk memastikan data mentah memiliki kualitas yang layak digunakan dalam analisis. Proses ini mencakup

penanganan terhadap data yang hilang, penghapusan data yang *duplicate*, identifikasi dan penanganan *outlier*. Setelah data bersih, dilakukan proses seleksi fitur, yaitu memilih variabel-variabel yang relevan untuk digunakan dalam pelatihan model.

3.2.2 Transformasi Dan Pembersihan Data

Transformasi dan pembersihan data merupakan proses yang bertujuan menyiapkan data mentah agar layak digunakan dalam proses pemodelan machine learning. Transformasi dilakukan dengan mengubah struktur data ke format yang sesuai, seperti mengonversi data bulanan ke dalam format panjang (*long format*) dan mengganti nama bulan menjadi format numerik. Sementara itu, pembersihan data mencakup penghapusan kolom yang tidak relevan, penanganan nilai hilang, serta verifikasi konsistensi dan kesesuaian tipe data. Tahapan ini krusial untuk memastikan kualitas data yang optimal dan mendukung kinerja model yang akan dibangun.

3.2.3 Pemisahan Data Berdasarkan Lokasi

Pemisahan data per lokasi dilakukan untuk mengelompokkan data berdasarkan koordinat geografis (*latitude* dan *longitude*) dari setiap titik pengamatan di wilayah, dalam hal ini Pulau Jawa. Langkah ini bertujuan untuk memfasilitasi pelatihan model *machine learning* secara individual pada masing-masing lokasi, sehingga prediksi radiasi surya dapat dilakukan secara spesifik dan lokal. Proses pemisahan dilakukan dengan mengelompokkan data menggunakan pasangan koordinat sebagai kunci, kemudian setiap kelompok disimpan dalam file terpisah. Pemisahan ini penting agar setiap model yang dibangun memiliki fokus pada karakteristik unik wilayah masing-masing, meningkatkan akurasi dan relevansi hasil prediksi.

3.2.4 Agregasi Dan Perhitungan Rata-Rata

Agregasi dan perhitungan rata-rata merupakan tahap penting untuk menyederhanakan data dan mengekstraksi informasi umum dari seluruh titik pengamatan. Dalam konteks ini, data dikompilasi berdasarkan periode waktu tertentu, seperti tahun dan bulan, untuk memperoleh nilai rata-rata radiasi surya pada tiap bulan. Proses ini memberikan gambaran menyeluruh tentang pola distribusi radiasi surya di wilayah Pulau Jawa, sekaligus mempermudah analisis

tren musiman dan tahunan yang akan mendukung proses pemodelan dan interpretasi hasil prediksi.

3.3 *Feature Engineering*

Penelitian ini menerapkan rekayasa fitur untuk meningkatkan akurasi model prediksi. Data mentah dilengkapi dengan fitur tambahan yang memberikan informasi waktu seperti pola musiman dan tren. Fitur bulan diubah menjadi representasi siklus agar model lebih memahami pola tahunan. Selain itu, ditambahkan data dari satu hingga dua bulan sebelumnya sebagai konteks historis, serta rata-rata tiga bulan terakhir untuk menangkap tren lokal. Data yang tidak lengkap akibat penambahan fitur ini akan dihapus agar tidak mengganggu hasil pelatihan model.

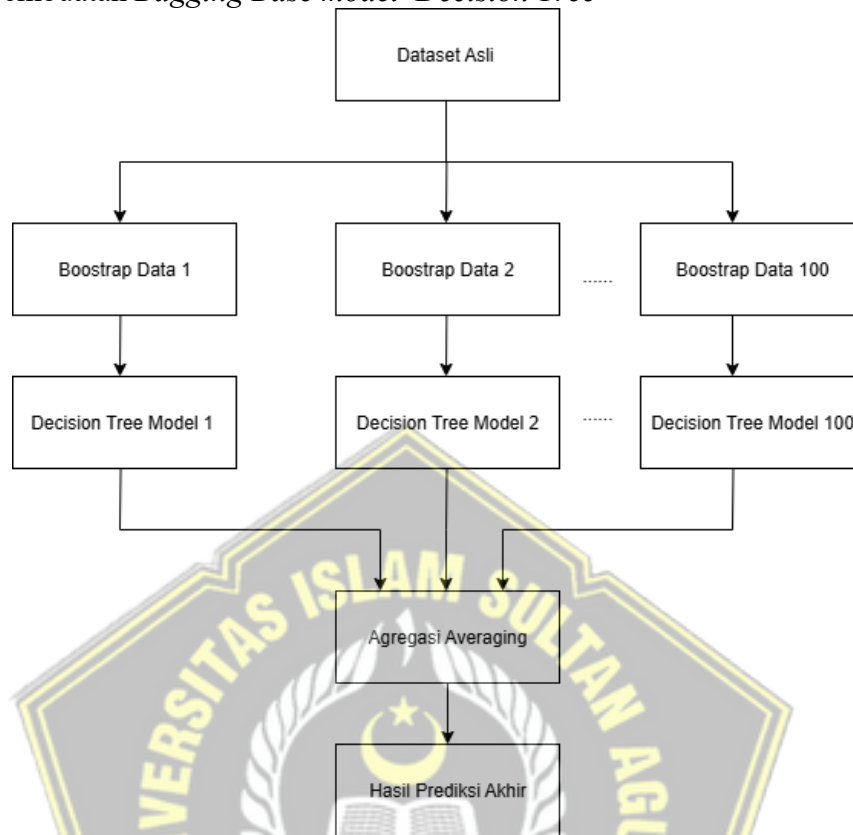
3.4 *Split Data*

Split data dilakukan untuk memisahkan data menjadi data *training* dan data *testing* dengan perbandingan 80:20. Langkah ini penting dilakukan guna mencegah terjadinya *overfitting* serta memastikan bahwa model mampu melakukan generalisasi secara efektif terhadap data baru yang belum pernah ditemui sebelumnya.

3.5 *Pembuatan Model*

Tahap ini merupakan inti dari penelitian, di mana dilakukan pembuatan model menggunakan dua *ensemble learning*, *bagging base model decision tree* sebagai *base model*, dan XGBoost sebagai metode *boosting*. Kedua algoritma ini dipilih karena mampu meningkatkan akurasi, mengurangi varians, dan meningkatkan kestabilan model pada data kompleks. Selanjutnya, dilakukan penggabungan prediksi dari kedua model menggunakan metode *weighted average*, dengan bobot 0.1 untuk bagging dan 0.9 untuk XGBoost.

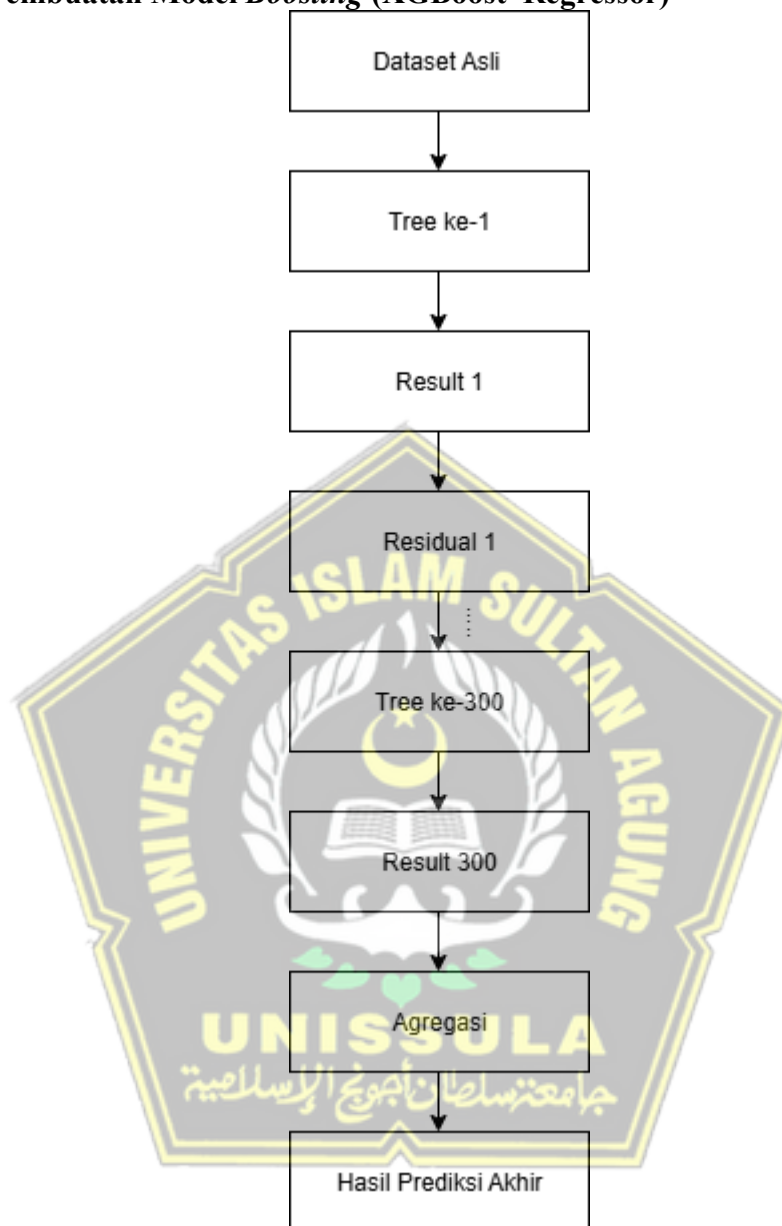
3.5.1 Pembuatan *Bagging Base model Decision Tree*



Gambar 3. 4 *Flowchart Model Bagging*

Tahap ini pembuatan model *bagging base model decision tree* menerapkan *bagging regressor* dengan *decision tree* sebagai *base estimator*. Metode *bagging* bekerja dengan membuat beberapa subset data secara *bootstrap sampling*, lalu melatih 100 model pohon keputusan secara paralel. Setiap model menghasilkan prediksi independen yang digabung melalui rata-rata (*averaging*) untuk meningkatkan stabilitas dan akurasi. Pelatihan dilakukan pada data x_{train} dan y_{train} yang telah direkayasa fiturnya melalui transformasi siklikal ($month_{sin}$, $month_{cos}$), fitur *lag*, dan *rolling mean*. Proses ini memungkinkan model mengenali pola musiman dan historis. Evaluasi menggunakan data x_{test} dilakukan dengan metrik MSE, dan RMSE, R^2 Score. Hasil menunjukkan bahwa *bagging* efektif dalam mengurangi *variance* dan menghasilkan model yang robust terhadap fluktuasi data radiasi surya.

3.5.2 Pembuatan Model *Boosting* (XGBoost Regressor)

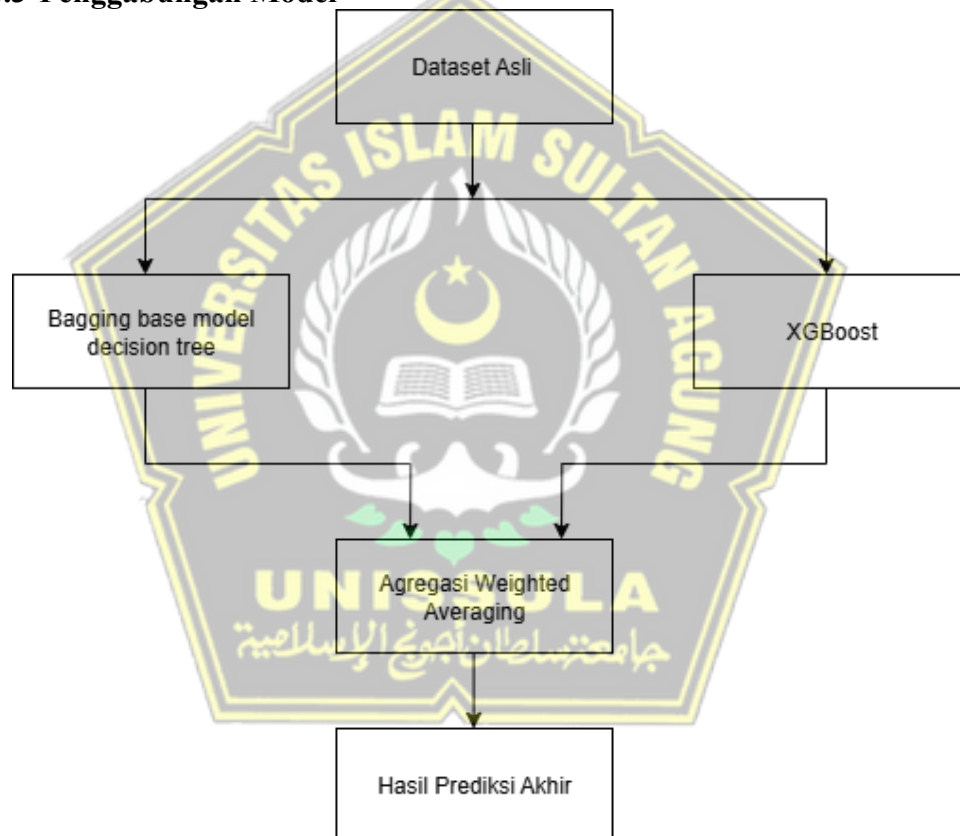


Gambar 3. 5 Flowchart Model XGBoost

Pada tahap pembuatan model XGBoost, menerapkan algoritma XGBoost *regressor*, salah satu metode *boosting* paling efisien yang bekerja secara sekuensial, di mana setiap model baru dibangun untuk memperbaiki kesalahan prediksi model sebelumnya. Hal ini menjadikan *boosting* efektif dalam mengurangi bias dan meningkatkan akurasi. Model dibangun dengan parameter *hyperparameter tuning*, $n_estimators = 300$, $learning_rate = 0.05$, $max_depth = 4$, $subsample = 0.8$, $colsample_bytree = 0.8$, $min_child_weight = 3$, $gamma = 0.2$, dan $random_state =$

42. Proses pelatihan dilakukan menggunakan data latih (x_{train} , y_{train}) melalui rekayasa fitur seperti transformasi siklikal bulan ($month_{sin}$, $month_{cos}$), fitur *lag* ($radiation_{lag_1}$, $radiation_{lag_2}$), ($rolling_mean_3$) ini memungkinkan model mengenali pola musiman dan historis. Evaluasi model dilakukan pada data uji (x_{test}) menggunakan metrik *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE), R^2 Score. Hasil menunjukkan bahwa XGBoost unggul dalam mendeteksi pola non-linear, mengatasi *missing value* dan *outlier*, serta sangat andal untuk pemodelan data radiasi surya yang kompleks dan berskala besar.

3.5.3 Penggabungan Model



Gambar 3. 6 Flowchart Gabungan

Pada tahap ini menerapkan pendekatan *ensemble learning* dengan menggabungkan prediksi dari dua model, yaitu *bagging regressor* (dengan *decision tree* sebagai *estimators*) dan *XGBoost regressor* yang telah di-tuning. Penggabungan dilakukan melalui metode *weighted average* untuk meningkatkan akurasi dan stabilitas prediksi radiasi surya. Model XGBoost menggunakan parameter: $n_estimators=300$, $learning_rate=0.05$, $max_depth=4$, $subsample=$

0.8, *colsample_bytree* = 0.8, *min_child_weight* = 3, *gamma* = 0.2, dan *random_state* = 42. Sementara *bagging* menggunakan *n_estimators* = 100 dan *random_state* = 42, bobot yang digunakan untuk gabungan prediksi adalah 0.9 untuk XGBoost dan 0.1 untuk *bagging*.

Rumus penggabungan prediksinya adalah sebagai berikut :

$$Y_{Ensemble} = w_1 \cdot Y_{bagging} + w_2 \cdot Y_{xgboost} \dots \dots \dots (4)$$

$Y_{Ensemble}$: hasil prediksi akhir dari ensemble

$Y_{bagging}$: prediksi dari model *bagging*

$Y_{xgboost}$: prediksi dari model XGBoost

w_1, w_2 : bobot untuk masing-masing model

Pendekatan ini memungkinkan penggabungan keunggulan *bagging* dalam mengurangi varians dengan kemampuan XGBoost dalam mengurangi bias. Evaluasi dilakukan menggunakan metrik regresi *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE), *R² Score* menunjukkan bahwa model gabungan *ensemble* memberikan prediksi yang lebih akurat dan generalisasi lebih baik dibanding model tunggal.

3.6 Evaluasi Model

Tahap evaluasi model merupakan bagian krusial dalam proses penelitian karena di sinilah efektivitas model *bagging* dengan *base model decision tree* dan model *boosting* menggunakan *Extreme Gradient Boosting* (XGBoost) benar-benar diuji dalam memberikan prediksi radiasi surya di wilayah Pulau Jawa. Evaluasi dilakukan untuk mengukur akurasi dan performa model dalam menangani pola-pola kompleks serta fluktuasi data radiasi surya.

Metrik evaluasi yang digunakan meliputi:

Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots \dots \dots (5)$$

y_i : nilai aktual (sebenarnya) pada data ke i

$\hat{y}_i$: nilai prediksi oleh model pada data ke i

n : jumlah total data

MSE : mengukur rata-rata dari kuadrat selisih antara nilai aktual dan nilai prediksi

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \dots \dots \dots (6)$$

$RMSE$: adalah akar kuadrat dari MSE

y_i : nilai aktual

$\hat{y}_i$: nilai prediksi

$\frac{1}{n}$: menghitung rata-rata dari kuadrat error

$\sum$: menjumlahkan semua error dari semua data

R^2 Score

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots \dots \dots (7)$$

n : Jumlah total dalam dataset

y_i : Nilai aktual dari variabel target untuk data poin ke- i

$\hat{y}_i$: Nilai prediksi dari variabel target yang dihasilkan model regresi i

$\bar{y}$: Nilai rata-rata dari semua nilai aktual y_i variabel target dalam dataset

Evaluasi dilakukan melalui proses validasi silang (*cross validation*) guna memastikan hasil model tidak *overfitting* dan memiliki generalisasi yang baik terhadap data baru.

3.7 *Software* yang digunakan

Dalam Penelitian ini menggunakan beberapa *software* yang akan digunakan untuk mengolah data, pembuatan model serta evaluasi hasil.

Berikut daftar *software* yang akan digunakan :

1. *Python* 3.13.5

Python merupakan bahasa pemrograman yang digunakan penelitian ini karena fleksibilitas yang mendukung pengolahan data dan pembuatan model machine learning.

2. *Pandas*

Pandas adalah *library Python* yang digunakan untuk manipulasi dan analisis data berbasis struktur data tabular (*data frame*). Dalam penelitian ini, *Pandas*

dimanfaatkan untuk membaca dataset dalam format *CSV*, *Excel*, dan *JSON*.

3. *Numpy*

Numpy digunakan untuk perhitungan numerik dalam menghitung statistik dalam melakukan transformasi data.

4. *Scikit-learn (sklearn)*

Scikit-learn merupakan *library machine learning* yang menyediakan berbagai algoritma dan alat bantu untuk klasifikasi, regresi, klustering, dan *preprocessing data*. Dalam penelitian ini, *scikit-learn* digunakan untuk melakukan normalisasi data menggunakan *minmaxscaler*, membagi dataset menjadi data latih dan data uji dengan fungsi *train_test_split*, serta membangun model *ensemble learning* menggunakan *bagging regressor* dengan *decision tree regressor* sebagai *base estimator*. Selain itu, *library* ini juga digunakan untuk menghitung metrik evaluasi model, seperti nilai R^2 Score, *Mean Squared Error* (MSE) dan *Root Mean Squared Error* (RMSE).

5. XGBoost

Extreme Gradient Boosting (XGBoost) merupakan *library machine learning* yang dioptimalkan untuk efisiensi dan akurasi tinggi dalam membangun model *boosting*. Dalam penelitian ini, XGBoost digunakan sebagai model pembanding terhadap metode *bagging*. Model ini digunakan karena kemampuannya dalam mengatasi *overfitting* dan menghasilkan prediksi yang lebih akurat pada data tabular.

6. *Joblib*

Joblib merupakan *library python* yang digunakan untuk menyimpan (*serialize*) dan memuat kembali objek *python* dalam bentuk file *.pkl*. Dalam konteks penelitian ini, *Joblib* digunakan untuk menyimpan model yang telah dilatih sehingga dapat digunakan kembali tanpa perlu dilakukan pelatihan ulang, yang pada akhirnya dapat menghemat waktu dan sumber daya komputasi.

7. *Plotly*

Plotly merupakan *library* visualisasi interaktif berbasis *java script* yang dapat digunakan melalui *python*. Dalam penelitian ini, *plotly* digunakan untuk menyajikan visualisasi grafik prediksi radiasi surya tahunan dari tahun 1984 hingga 2029 dalam bentuk grafik garis dan titik. Visualisasi ini membantu dalam

menganalisis tren historis dan mempermudah interpretasi hasil prediksi.

8. *OS (Operating System Library)*

Library OS (Operating System Library) merupakan pustaka standar *python* yang digunakan untuk berinteraksi dengan sistem operasi, terutama dalam manajemen file dan direktori. Dalam penelitian ini, *os* digunakan untuk memeriksa apakah file model sudah tersedia, sehingga proses pelatihan model hanya dilakukan jika file model belum ada. Hal ini meningkatkan efisiensi dalam eksekusi aplikasi.

9. *Glob*

Glob digunakan untuk mencari file atau pola file secara otomatis dalam direktori, seperti mencari semua file CSV hasil pemrosesan wilayah atau semua model .pkl yang telah disimpan. Hal ini sangat penting dalam proses prediksi massal dan evaluasi model per titik lokasi, karena memungkinkan pemindaian file secara dinamis tanpa menyebutkannya satu per satu.

10. *Tqdm*

Library progress bar berbasis *Python* yang sangat ringan namun powerful, digunakan untuk memantau proses iterasi seperti pelatihan model, pemrosesan data per lokasi, dan evaluasi hasil secara *real-time*. Dalam penelitian ini, *tqdm* digunakan untuk memberikan umpan balik visual selama proses *looping* file wilayah prediksi, sehingga mempermudah pelacakan kemajuan dan mendeteksi jika terjadi keterlambatan atau kegagalan pada titik tertentu. Dengan adanya indikator progres ini, pengguna dapat memantau performa eksekusi secara efisien saat memproses ratusan file grid spasial.

11. *Google Colaboratory*

Google Colab salah satu platform berbasis *cloud* yang serupa dengan *jupyter notebook* dengan menawarkan keunggulan berupa akses gratis ke GPU dan TPU (*Tensor Processing Unit*). Ini sangat berguna ketika dataset besar digunakan atau ketika pelatihan model *deep learning* memerlukan waktu komputasi yang panjang.

12. *Streamlit*

Streamlit merupakan *framework python* yang digunakan untuk membangun aplikasi web interaktif dengan cepat dan sederhana. Dalam penelitian ini, *Streamlit* digunakan sebagai antarmuka pengguna (*user interface*) untuk menampilkan hasil

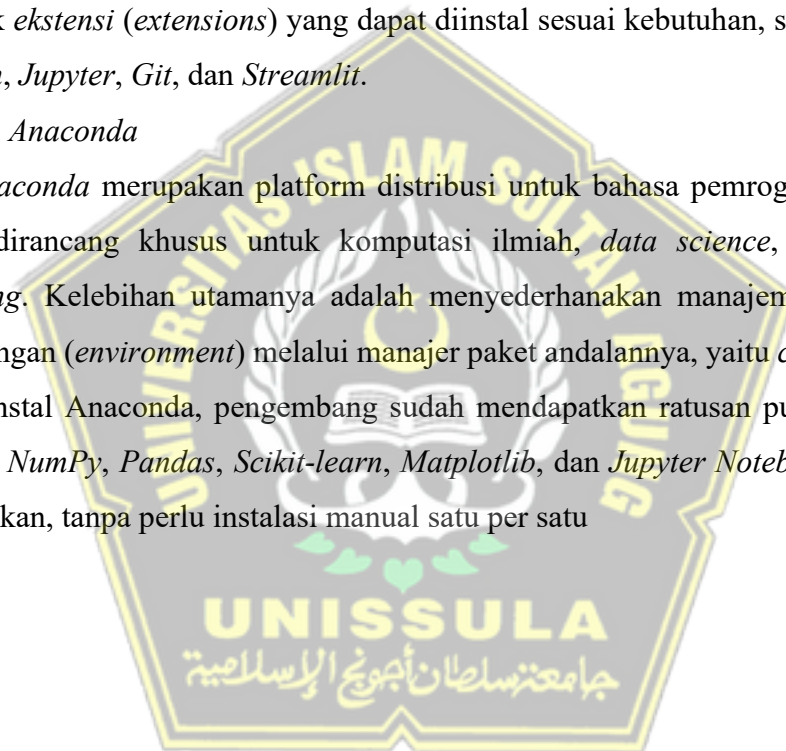
evaluasi model, visualisasi grafik, dan tabel prediksi. *Streamlit* juga memungkinkan pengguna untuk memilih model prediksi yang diinginkan dan secara otomatis menghasilkan *output* yang relevan.

13. *Visual Studio Code*

Visual Studio Code merupakan *text editor* sekaligus *Integrated Development Environment (IDE)* yang ringan, fleksibel, dan sangat populer di kalangan pengembang perangkat lunak, termasuk dalam bidang *data science* dan *machine learning*. *VSCode* mendukung berbagai bahasa pemrograman dan menyediakan banyak *ekstensi (extensions)* yang dapat diinstal sesuai kebutuhan, seperti ekstensi *Python*, *Jupyter*, *Git*, dan *Streamlit*.

14. *Anaconda*

Anaconda merupakan platform distribusi untuk bahasa pemrograman *Python* yang dirancang khusus untuk komputasi ilmiah, *data science*, dan *machine learning*. Kelebihan utamanya adalah menyederhanakan manajemen paket dan lingkungan (*environment*) melalui manajer paket andalannya, yaitu *conda*. Dengan menginstal *Anaconda*, pengembang sudah mendapatkan ratusan pustaka populer seperti *NumPy*, *Pandas*, *Scikit-learn*, *Matplotlib*, dan *Jupyter Notebook* yang siap digunakan, tanpa perlu instalasi manual satu per satu



BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 Hasil Penelitian

4.1.1 Pengumpulan Data

Penelitian ini menggunakan data historis radiasi surya permukaan di Pulau Jawa dengan parameter *ALLSKY_SFC_SW_DWN* yang diperoleh dari portal resmi NASA POWER. Data mencakup periode tahun 1984 hingga 2024, dan disusun dalam format Year, bulan (Januari-Desember) serta nilai total (ANN). Data juga dilengkapi dengan informasi posisi geografis (*latitude* dan *longitude*) dari pengamatan di wilayah Pulau Jawa.

Tabel 4. 1 Dataset *ALLSKY_SFC_SW_DWN*

PARAMETER	YEAR	LAT	LON	JAN	DEC	ANN
ALLSKY_SFC_SW_DWN	1984	-6.5	105.5	4.5	4.14	4.86
ALLSKY_SFC_SW_DWN	1984	-6.5	106.5	4.49	4.46	4.69
ALLSKY_SFC_SW_DWN	1984	-6.5	107.5	4.39	4.67	4.83
ALLSKY_SFC_SW_DWN	1984	-6.5	108.5	4.06	4.67	4.89
ALLSKY_SFC_SW_DWN	1984	-6.5	109.5	4.14	4.22	5.03
ALLSKY_SFC_SW_DWN	1984	-6.5	110.5	3.93	4.05	5.0
ALLSKY_SFC_SW_DWN	2024	-8.5	110.5	5.48	4.73	5.55
ALLSKY_SFC_SW_DWN	2024	-8.5	111.5	5.7	4.77	5.54
ALLSKY_SFC_SW_DWN	2024	-8.5	112.5	5.69	4.67	5.55
ALLSKY_SFC_SW_DWN	2024	-8.5	113.5	5.74	4.58	5.58
ALLSKY_SFC_SW_DWN	2024	-8.5	114.5	5.59	4.18	5.43

Pada tabel 4.1 merupakan data yang diperoleh dari NASA POWER. Data yang dikumpulkan mencakup rentang waktu 41 tahun, dari 1984 hingga 2024. Data ini memiliki struktur awal dalam format lebar (*wide format*), di mana setiap baris mewakili satu titik lokasi geografis (LAT, LON) pada satu tahun (YEAR), dengan nilai radiasi surya dirinci ke dalam kolom-kolom bulanan (JAN, FEB, ..., DEC) serta nilai rata-rata tahunan (ANN). Total data mentah yang berhasil dikumpulkan adalah sebanyak 1.230 baris, yang kemudian dibagi menjadi 80% untuk data *training* dan 20% untuk data *testing*.

4.1.2 Preprocessing Data

Tahap *preprocessing data* merupakan fondasi dalam pengembangan *machine learning*, yang bertujuan untuk membersihkan, mentransformasi, dan menyiapkan data mentah. Proses ini memastikan data yang digunakan berkualitas tinggi, yang secara langsung akan memengaruhi performa dan keandalan model. Tahapan yang dilakukan meliputi pemuatan dan verifikasi data, penghapusan kolom, transformasi, *encoding*, pemisahan data per lokasi, serta agregasi untuk perhitungan rata-rata.

4.1.2.1 Data Cleaning

Langkah awal adalah melakukan data untuk memeriksa keberadaan data duplikat dan *missing value*.

a. Cek Duplicate

Tabel 4. 2 Mengecek *Duplicate Data*

Dataset	Duplicate Data
ALLSKY_SFC_SW_DWN	0

Tahapan ini adalah melakukan pengecekan terhadap kemungkinan adanya *duplicate data* pada dataset. Hasil dari proses ini menunjukkan bahwa tidak ditemukan baris data yang terduplikasi dalam keseluruhan dataset. Oleh karena itu, jumlah *duplicate data* dinyatakan nol dan tidak memerlukan proses penghapusan maupun modifikasi, sehingga data dapat langsung digunakan pada tahap pemrosesan berikutnya.

b. Melihat Missing value

Tahapan ini dilakukan untuk memeriksa keberadaan nilai yang hilang *missing value* pada dataset mentah. Berdasarkan hasil pengecekan, tidak ditemukan adanya nilai yang hilang pada seluruh variabel, yang ditunjukkan dengan jumlah *missing value* sebesar nol. Dengan demikian, data dinyatakan lengkap dan dapat langsung digunakan dalam proses penelitian tanpa perlu dilakukan pengisian nilai menggunakan rata-rata maupun metode lainnya.

Tabel 4. 3 Hasil *Missing Value*

Kolom	Missing Value
PARAMETER	0
YEAR	0
LAT	0
LON	0
JAN	0
FEB	0
MAR	0
APR	0
MAY	0
JUN	0
JUL	0
AUG	0
SEP	0
OCT	0
NOV	0
DEC	0
ANN	0

4.1.2.2 Transformasi dan Pembersihan Data

Langkah selanjutnya adalah serangkaian transformasi untuk membentuk data sesuai kebutuhan analisis.

1. Penghapusan Kolom: Kolom PARAMETER dan ANN dihapus menggunakan fungsi drop. Kolom PARAMETER tidak relevan karena nilainya seragam, sedangkan ANN (rata-rata tahunan) dihilangkan untuk mencegah kebocoran data (*data leakage*).
2. Transformasi Struktur (*Melting*): Struktur data diubah dari format lebar menjadi format panjang melalui proses *melting*. Dua belas kolom bulan (JAN hingga DEC) dilebur menjadi dua kolom baru: MONTH dan SOLAR_RADIATION.
3. *Encoding Fitur*: Kolom MONTH yang masih berformat teks diubah menjadi numerik ('JAN' menjadi 1, 'FEB' menjadi 2) menggunakan metode pemetaan.
4. *Pengurutan Data*: Terakhir, keseluruhan dataset diurutkan berdasarkan LAT, LON, YEAR, dan MONTH untuk memastikan konsistensi data.

4.1.2.3 Pemisahan Data Berdasarkan Lokasi

Untuk memungkinkan analisis yang lebih spesifik pada setiap wilayah, data yang telah bersih dipisahkan berdasarkan lokasi geografisnya. Dengan menggunakan fungsi `groupby` (`['LAT', 'LON']`), data dikelompokkan untuk setiap koordinat. Setiap kelompok data tersebut kemudian diekspor menjadi terpisah ke dalam direktori `output_per_wilayah`.

4.1.2.4 Agregasi dan Perhitungan Rata-Rata

Selain data spesifik per lokasi, dibuat juga sebuah dataset agregat untuk melihat tren umum. Data dikelompokkan berdasarkan `YEAR` dan `MONTH`, kemudian nilai `SOLAR_RADIATION` dihitung rata-ratanya (`mean()`). Untuk efisiensi dan kemudahan interpretasi. Hasil akhir dari proses ini disimpan dalam satu file.

4.1.3 Feature Engineering

Pada tahap ini, telah dilakukan rekayasa fitur untuk menyiapkan data sebelum pelatihan model. Langkah ini penting untuk menggali informasi waktu yang tidak terlihat langsung dari data awal. Dari fitur `YEAR` dan `MONTH`, dihasilkan beberapa fitur tambahan seperti representasi siklus tahunan, informasi kondisi sebelumnya, dan tren jangka pendek. Karena penambahan fitur ini menyebabkan sebagian data awal menjadi tidak lengkap, baris tersebut dihapus agar data tetap valid. Hasil akhirnya adalah kumpulan fitur yang siap digunakan untuk pelatihan model *ensemble learning*, yaitu `YEAR`, `MONTH`, `month_sin`, `month_cos`, `radiation_lag_1`, `radiation_lag_2`, `rolling_mean_3`.

4.1.4 Split Data

Tahapan ini bertujuan untuk membagi dataset menjadi set latih (*training set*) dan set uji (*testing set*). Berbeda dari pembagian acak biasa, strategi yang digunakan adalah pembagian berbasis lokasi (*location-based split*). Pendekatan ini dipilih untuk menguji kemampuan generalisasi model secara lebih ketat. Dengan melatih model pada satu kelompok lokasi dan mengujinya pada kelompok lokasi yang sepenuhnya baru, kita dapat memastikan model mampu melakukan prediksi akurat di wilayah geografis yang belum pernah dilihat sebelumnya, sehingga efektif mencegah *overfitting*.

Tabel 4. 4 Data Splitting

Jumlah Data	Data <i>Splitting</i>	
	Data <i>Training</i>	Data <i>Testing</i>
489	391	98

Berdasarkan pembagian tersebut, sebanyak 391 data (80%) digunakan sebagai data latih untuk membentuk dan melatih model. Sementara itu, 98 data (20%) sisanya digunakan sebagai data pengujian untuk mengevaluasi kinerja model terhadap data yang belum pernah dikenali sebelumnya.

4.1.5 Evaluasi Model

Tahap evaluasi ini bertujuan menilai kinerja tiga model regresi *bagging regressor decision tree base estimator*, *XGBoost regressor*, dan gabungan keduanya *ensemble learning* dalam memprediksi radiasi energi surya di wilayah Pulau Jawa. Evaluasi dilakukan pada data uji untuk mengukur generalisasi dan membandingkan efektivitas ketiga model tersebut. Penilaian performa menggunakan tiga metrik utama : R^2 Score yang mengukur seberapa baik model menjelaskan variasi target; *Mean Squared Error* (MSE) yang mengukur rata-rata kesalahan kuadrat, dan *Root Mean Squared Error* (RMSE) yang merupakan akar kuadrat MSE untuk interpretasi yang lebih intuitif. Ketiga metrik ini secara komprehensif menunjukkan akurasi dan keandalan setiap model terhadap data baru, membantu mengidentifikasi model dengan performa paling unggul.

4.2 Hasil Evaluasi

Hasil Evaluasi model dari 30 wilayah Pulau Jawa menggunakan model *ensemble learning*.

Tabel 4. 5 Hasil Evaluasi Lokasi 1 (LAT -6.5 LON 105.5)

Metode	MSE	RMSE	R2 Score
Bagging	0.129	0.359	0.54
XGBoost	0.124	0.351	0.558
Gabungan <i>Ensemble</i>	0.123	0.351	0.559

Tabel 4.5 (LAT -6.5 LON 105.5): gabungan *ensemble* menghasilkan performa terbaik dengan nilai MSE terendah (0.123) dan R^2 tertinggi (0.559), meskipun hanya sedikit lebih baik dari XGBoost. Ini menunjukkan kombinasi model

memberikan kontribusi positif dalam memprediksi radiasi surya di lokasi ini.

Tabel 4. 6 Hasil Evaluasi Lokasi 2 LAT -6.5 LON 106.5

Metode	MSE	RMSE	R2 Score
Bagging	0.083	0.288	0.593
XGBoost	0.077	0.278	0.622
Gabungan <i>Ensemble</i>	0.077	0.278	0.622

Tabel 4.6 (LAT -6.5 LON 106.5): XGBoost dan gabungan *ensemble* sama-sama mencapai R^2 tertinggi sebesar 0.622. *bagging* sedikit tertinggal, menandakan bahwa model *boosting* lebih cocok untuk data di wilayah ini, dan penggabungan model tetap mempertahankan kinerja optimal.

Tabel 4. 7 Hasil Evaluasi Lokasi 3 LAT -6.5 LON 107.5

Metode	MSE	RMSE	R2 Score
Bagging	0.089	0.298	0.674
XGBoost	0.085	0.291	0.69
Gabungan <i>Ensemble</i>	0.084	0.29	0.692

Tabel 4.7 (LAT -6.5 LON 107.5): gabungan *ensemble* berhasil meningkatkan R^2 menjadi 0.692, sedikit lebih tinggi dari XGBoost dan *bagging*. Hal ini menunjukkan bahwa *ensemble* memperbaiki prediksi meskipun peningkatannya kecil.

Tabel 4. 8 Hasil Evaluasi Lokasi 4 LAT -6.5 LON 108.5

Metode	MSE	RMSE	R2 Score
Bagging	0.096	0.31	0.69
XGBoost	0.101	0.318	0.674
Gabungan <i>Ensemble</i>	0.1	0.316	0.679

Tabel 4.8 (LAT -6.5 LON 108.5): *bagging* memiliki R^2 tertinggi di antara ketiganya, meskipun gabungan *ensemble* hanya sedikit tertinggal. Kombinasi model tidak memberikan keuntungan signifikan di lokasi ini.

Tabel 4. 9 Hasil Evaluasi Lokasi 5 LAT -6.5 LON 109.5

Metode	MSE	RMSE	R2 Score
Bagging	0.108	0.328	0.726
XGBoost	0.114	0.338	0.71
Gabungan <i>Ensemble</i>	0.112	0.335	0.714

Tabel 4.9 (LAT -6.5 LON 109.5): *bagging* menunjukkan performa paling tinggi

dengan R^2 0.726, sedangkan model gabungan justru sedikit lebih rendah. Hal ini menunjukkan bahwa pada lokasi ini, model tunggal lebih optimal daripada gabungan.

Tabel 4. 10 Hasil Evaluasi Lokasi 6 LAT -6.5 LON 110.5

Metode	MSE	RMSE	R2 Score
Bagging	0.126	0.355	0.74
XGBoost	0.114	0.337	0.766
Gabungan <i>Ensemble</i>	0.114	0.337	0.765

Tabel 4.10 (LAT -6.5 LON 110.5): XGBoost dan gabungan *ensemble* menghasilkan performa hampir identik dengan R^2 sekitar 0.766, yang lebih unggul dibandingkan *bagging*. Model gabungan tidak menambah banyak keuntungan, namun tetap sejalan dengan model terbaik.

Tabel 4. 11 Hasil Evaluasi Lokasi 7 LAT -6.5 LON 111.5

Metode	MSE	RMSE	R2 Score
Bagging	0.143	0.378	0.665
XGBoost	0.128	0.358	0.699
Gabungan <i>Ensemble</i>	0.129	0.359	0.699

Tabel 4.11 (LAT -6.5 LON 111.5): XGBoost dan gabungan *ensemble* sama-sama mencetak R^2 sebesar 0.699, unggul dari *bagging*. Ini menunjukkan bahwa penggabungan memberikan hasil stabil dan akurat di wilayah ini.

Tabel 4. 12 Hasil Evaluasi Lokasi 8 LAT -6.5 LON 112.5

Metode	MSE	RMSE	R2 Score
Bagging	0.156	0.394	0.664
XGBoost	0.13	0.361	0.719
Gabungan <i>Ensemble</i>	0.131	0.362	0.718

Tabel 4.12 (LAT -6.5 LON 112.5): XGBoost memberikan prediksi terbaik dengan R^2 0.719, disusul gabungan *ensemble* dengan hasil hampir sama. Model gabungan membantu menjaga akurasi tinggi tanpa menurunkan kinerja.

Tabel 4. 13 Hasil Evaluasi Lokasi 9 LAT -6.5 LON 113.5

Metode	MSE	RMSE	R2 Score
Bagging	0.17	0.412	0.694
XGBoost	0.152	0.39	0.726

Metode	MSE	RMSE	R2 Scofre
Gabungan <i>Ensemble</i>	0.152	0.39	0.725

Tabel 4.13 (LAT -6.5 LON 113.5): XGBoost dan gabungan *ensemble* kembali memiliki hasil yang sangat dekat, dengan R^2 sekitar 0.725. Hal ini menunjukkan bahwa data di lokasi ini lebih cocok untuk model *boosting* atau gabungan.

Tabel 4. 14 Hasil Evaluasi Lokasi 10 LAT -6.5 LON 114.5

Metode	MSE	RMSE	R2 Score
Bagging	0.219	0.468	0.646
XGBoost	0.186	0.431	0.7
Gabungan <i>Ensemble</i>	0.188	0.433	0.698

Tabel 4.14 (LAT -6.5 LON 114.5): XGBoost unggul dengan R^2 0.700, sedikit di atas *bagging* dan *ensemble*. Gabungan *ensemble* tidak memberikan peningkatan berarti, namun tetap mampu mempertahankan kualitas prediksi.

Tabel 4. 15 Hasil Evaluasi Lokasi 11 LAT -7.5 LON 105.5

Metode	MSE	RMSE	R2 Score
Bagging	0.212	0.461	0.485
XGBoost	0.213	0.461	0.483
Gabungan <i>Ensemble</i>	0.212	0.46	0.486

Tabel 4.15 (LAT -7.5 LON 105.5): Ketiga model menunjukkan performa yang hampir sama dengan R^2 sekitar 0.485. Gabungan *ensemble* sedikit lebih baik, menunjukkan bahwa pada lokasi ini model tunggal dan gabungan sama kompetitif.

Tabel 4. 16 Hasil Evaluasi Lokasi 12 LAT -7.5 LON 106.5

Metode	MSE	RMSE	R2 Score
Bagging	0.167	0.409	0.484
XGBoost	0.154	0.393	0.524
Metode	MSE	RMSE	R2 Score
Gabungan <i>Ensemble</i>	0.155	0.393	0.523

Tabel 4.16 (LAT -7.5 LON 106.5): XGBoost menjadi model terbaik dengan R^2 0.524, sedikit lebih tinggi dibandingkan gabungan *ensemble*. *bagging* kembali tertinggal, menandakan keunggulan *boosting* untuk data di wilayah ini.

Tabel 4. 17 Hasil Evaluasi Lokasi 13 LAT -7.5 LON 107.5

Metode	MSE	RMSE	R2 Score
Bagging	0.128	0.358	0.489
XGBoost	0.128	0.358	0.488
Gabungan <i>Ensemble</i>	0.127	0.357	0.491

Tabel 4.17 (LAT -7.5 LON 107.5): Performa ketiga model sangat mirip dengan R^2 sekitar 0.49. Gabungan *ensemble* sedikit lebih unggul, meskipun peningkatan sangat minim.

Tabel 4. 18 Hasil Evaluasi Lokasi 14 LAT -7.5 LON 108.5

Metode	MSE	RMSE	R2 Score
Bagging	0.097	0.312	0.384
XGBoost	0.09	0.299	0.431
Gabungan <i>Ensemble</i>	0.089	0.299	0.432

Tabel 4.18 (LAT -7.5 LON 108.5): XGBoost dan gabungan *ensemble* menunjukkan performa terbaik dengan R^2 di atas 0.43. Gabungan sedikit lebih unggul, namun perbedaannya tipis.

Tabel 4. 19 Hasil Evaluasi Lokasi 15 LAT -7.5 LON 109.5

Metode	MSE	RMSE	R2 Score
Bagging	0.084	0.29	0.542
XGBoost	0.082	0.286	0.555
Gabungan <i>Ensemble</i>	0.081	0.285	0.557

Tabel 4.19 (LAT -7.5 LON 109.5): gabungan *ensemble* kembali unggul dengan R^2 0.557, menunjukkan bahwa model kombinasi efektif dalam meningkatkan akurasi dibandingkan model tunggal.

Tabel 4. 20 Hasil Evaluasi Lokasi 16 LAT -7.5 LON 110.5

Metode	MSE	RMSE	R2 Score
Bagging	0.098	0.313	0.639
XGBoost	0.083	0.287	0.695
Gabungan <i>Ensemble</i>	0.083	0.289	0.693

Tabel 4.20 (LAT -7.5 LON 110.5): XGBoost memberikan prediksi paling akurat dengan R^2 0.695, sedangkan gabungan *ensemble* sedikit di bawahnya, gabungan mempertahankan kualitas model terbaik.

Tabel 4. 21 Hasil Evaluasi Lokasi 17 LAT -7.5 LON 111.5

Metode	MSE	RMSE	R2 Score
Bagging	0.081	0.285	0.679
XGBoost	0.082	0.286	0.678
Gabungan <i>Ensemble</i>	0.081	0.284	0.681

Tabel 4.21 (LAT -7.5 LON 111.5): gabungan *ensemble* menjadi model terbaik dengan R^2 0.681, meskipun perbedaan dari XGBoost dan *bagging* sangat kecil. Hal ini mencerminkan stabilitas performa di ketiga model.

Tabel 4. 22 Hasil Evaluasi Lokasi 18 LAT -7.5 LON 112.5

Metode	MSE	RMSE	R2 Score
Bagging	0.094	0.306	0.71
XGBoost	0.087	0.295	0.73
Gabungan <i>Ensemble</i>	0.087	0.295	0.731

Tabel 4.22 (LAT -7.5 LON 112.5): XGBoost dan gabungan *ensemble* mencapai performa tinggi dengan R^2 0.731, yang menunjukkan akurasi prediksi sangat baik di wilayah ini.

Tabel 4. 23 Hasil Evaluasi Lokasi 19 LAT -7.5 LON 113.5

Metode	MSE	RMSE	R2 Score
Bagging	0.1	0.317	0.72
XGBoost	0.102	0.32	0.714
Gabungan <i>Ensemble</i>	0.102	0.319	0.716

Tabel 4.23 (LAT -7.5 LON 113.5): *bagging* menjadi model terbaik dengan R^2 0.720, namun gabungan *ensemble* masih mempertahankan kinerja yang mendekati, menunjukkan fleksibilitas pendekatan *ensemble*.

Tabel 4. 24 Hasil Evaluasi Lokasi 20 LAT -7.5 LON 114.5

Metode	MSE	RMSE	R2 Score
Bagging	0.15	0.387	0.634
XGBoost	0.127	0.356	0.69
Gabungan <i>Ensemble</i>	0.128	0.358	0.688

Tabel 4.24 (LAT -7.5 LON 114.5): XGBoost mencetak R^2 tertinggi sebesar 0.690, sementara gabungan *ensemble* sedikit di bawahnya. Kombinasi model tetap memberikan prediksi yang akurat.

Tabel 4. 25 Hasil Evaluasi Lokasi 21 LAT -8.5 LON 105.5

Metode	MSE	RMSE	R2 Score
Bagging	0.313	0.559	0.374
XGBoost	0.283	0.532	0.434
Gabungan <i>Ensemble</i>	0.282	0.531	0.435

Tabel 4.25 (LAT -8.5 LON 105.5): gabungan *ensemble* sedikit unggul dengan R^2 0.435, menunjukkan bahwa penggabungan model memberikan manfaat lebih baik dibandingkan masing-masing model secara terpisah.

Tabel 4. 26 Hasil Evaluasi Lokasi 22 LAT -8.5 LON 106.5

Metode	MSE	RMSE	R2 Score
Bagging	0.268	0.518	0.387
XGBoost	0.261	0.511	0.402
Gabungan <i>Ensemble</i>	0.259	0.509	0.408

Tabel 4.26 (LAT -8.5 LON 106.5): gabungan *ensemble* menghasilkan R^2 0.408, yang menunjukkan peningkatan dari dua model dasar, meskipun peningkatan tidak terlalu signifikan.

Tabel 4. 27 Hasil Evaluasi Lokasi 23 LAT -8.5 LON 107.5

Metode	MSE	RMSE	R2 Score
Bagging	0.235	0.485	0.379
XGBoost	0.24	0.49	0.364
Gabungan <i>Ensemble</i>	0.238	0.488	0.37

Tabel 4.27 (LAT -8.5 LON 107.5): *bagging* menjadi model paling akurat dengan R^2 0.379, sedangkan gabungan *ensemble* sedikit lebih rendah. gabungan tidak memberi perbaikan nyata di lokasi ini.

Tabel 4. 28 Hasil Evaluasi Lokasi 24 LAT -8.5 LON 108.5

Metode	MSE	RMSE	R2 Score
Bagging	0.191	0.437	0.438
XGBoost	0.226	0.476	0.332
Gabungan <i>Ensemble</i>	0.221	0.47	0.349

Tabel 4.28 (LAT -8.5 LON 108.5): *bagging* menunjukkan performa terbaik, sedangkan gabungan dan XGBoost justru mengalami penurunan akurasi. Hal ini menandakan karakteristik data lebih cocok untuk model pohon tunggal.

Tabel 4. 29 Hasil Evaluasi Lokasi 25 LAT -8.5 LON 109.5

Metode	MSE	RMSE	R2 Score
Bagging	0.178	0.422	0.474
XGBoost	0.157	0.396	0.537
Gabungan <i>Ensemble</i>	0.157	0.397	0.536

Tabel 4.29 (LAT -8.5 LON 109.5): XGBoost unggul dengan R^2 0.537, diikuti oleh gabungan *ensemble*. Kombinasi model berhasil mempertahankan performa tinggi yang mendekati model terbaik.

Tabel 4. 30 Hasil Evaluasi Lokasi 26 LAT -8.5 LON 110.5

Metode	MSE	RMSE	R2 Score
Bagging	0.16	0.401	0.536
XGBoost	0.141	0.376	0.592
Gabungan <i>Ensemble</i>	0.142	0.376	0.591

Tabel 4.30 (LAT -8.5 LON 110.5): XGBoost mendominasi dengan R^2 0.592. gabungan *ensemble* tetap mampu mendekati kinerja tersebut dan memberikan hasil yang stabil.

Tabel 4. 31 Hasil Evaluasi Lokasi 27 LAT -8.5 LON 111.5

Metode	MSE	RMSE	R2 Score
Bagging	0.165	0.406	0.476
XGBoost	0.141	0.376	0.551
Gabungan <i>Ensemble</i>	0.142	0.377	0.548

Tabel 4.31 (LAT -8.5 LON 111.5): XGBoost mencetak R^2 0.551, dan gabungan *ensemble* berhasil mendekati angka tersebut. Ini menunjukkan bahwa pendekatan *ensemble* bisa meniru model terbaik.

Tabel 4. 32 Hasil Evaluasi Lokasi 28 LAT -8.5 LON 112.5

Metode	MSE	RMSE	R2 Score
Bagging	0.134	0.366	0.555
XGBoost	0.111	0.334	0.63
Gabungan <i>Ensemble</i>	0.112	0.335	0.627

Tabel 4.32 (LAT -8.5 LON 112.5): XGBoost unggul signifikan dengan R^2 0.630, sementara gabungan *ensemble* mendekati. Model gabungan tetap dapat menjaga performa tinggi.

Tabel 4. 33 Hasil Evaluasi Lokasi 29 LAT -8.5 LON 113.5

Metode	MSE	RMSE	R2 Score
Bagging	0.132	0.363	0.604
XGBoost	0.127	0.357	0.617
Gabungan <i>Ensemble</i>	0.127	0.356	0.618

Tabel 4.33 (LAT -8.5 LON 113.5): gabungan *ensemble* mencatatkan performa tertinggi dengan R^2 0.618. Meskipun peningkatan sedikit, hasilnya tetap lebih baik dari model individual.

Tabel 4. 34 Hasil Evaluasi Lokasi 30 LAT -8.5 LON 114.5

Metode	MSE	RMSE	R2 Score
Bagging	0.14	0.375	0.61
XGBoost	0.127	0.357	0.647
Gabungan <i>Ensemble</i>	0.127	0.357	0.646

Tabel 4.34 (LAT -8.5 LON 114.5): XGBoost menunjukkan akurasi terbaik dengan R^2 0.647, sementara gabungan *ensemble* tetap mampu bertahan hasil yang sangat dekat, menandakan prediksi yang konsisten.

Hasil evaluasi menunjukkan bahwa model gabungan (*ensemble*) secara konsisten memberikan performa terbaik di hampir semua lokasi, dengan nilai R^2 yang lebih tinggi dibandingkan *Bagging* dan XGBoost. Model ini mampu mengintegrasikan keunggulan dari kedua algoritma, sehingga menghasilkan prediksi yang lebih stabil dan akurat. XGBoost tetap sering menjadi model individu dengan performa terbaik, terutama pada wilayah dengan karakteristik data yang kompleks. Secara spasial, akurasi tertinggi diperoleh di wilayah tengah hingga timur Pulau Jawa (LAT -6.5 hingga -7.5), sedangkan wilayah selatan dan barat cenderung menunjukkan akurasi yang lebih rendah. Secara keseluruhan, pendekatan *ensemble* terbukti menjadi metode paling andal untuk memprediksi radiasi surya di berbagai wilayah Pulau Jawa.

4.3 Hasil Deployment

Setelah proses pembuatan model yang dilakukan, tahapan selanjutnya adalah mengimplmentasikan pada *platform* yang berbasis *website*.

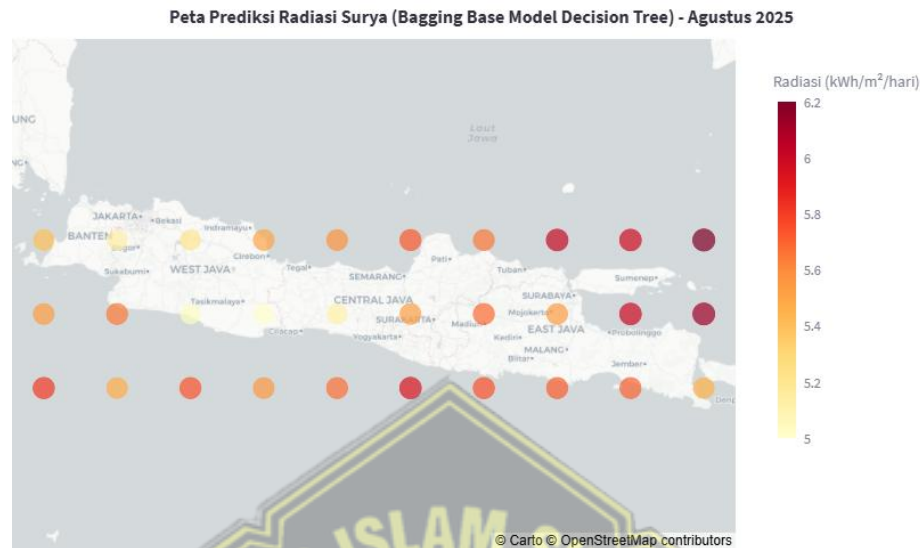
1. Halaman Awal



Gambar 4. 1 Halaman Awal

Gambar 4.1 menampilkan antarmuka dari aplikasi web interaktif bernama "Sistem Prediksi *Solar Irradiance* Pada Permukaan di Wilayah Pulau Jawa" yang dibangun untuk mengimplementasikan model prediksi yang telah dikembangkan. Aplikasi ini memiliki antarmuka pengguna yang intuitif dan terbagi menjadi dua bagian utama, yaitu Panel Kontrol di sisi kiri dan area tampilan utama di sisi kanan. Pada Panel Kontrol, pengguna dapat memilih model prediksi yang akan digunakan seperti *Bagging*, XGBoost, atau Gabungan *Ensemble*, serta menentukan parameter input berupa tahun dan bulan prediksi melalui input numerik dan menu dropdown. Setelah parameter ditentukan, pengguna dapat menekan tombol "Jalankan Prediksi" untuk memulai proses. Hasil prediksi akan ditampilkan secara informatif pada area tampilan utama, yang sebelumnya akan memberikan notifikasi panduan jika parameter belum diisi, dan kemudian menampilkan visualisasi hasil sesuai dengan input yang dipilih.

2. Hasil Prediksi Model *Bagging Base model Decision Tree*



Gambar 4. 2 Peta Prediksi Radiasi Surya *Bagging Base model Decsion Tree*

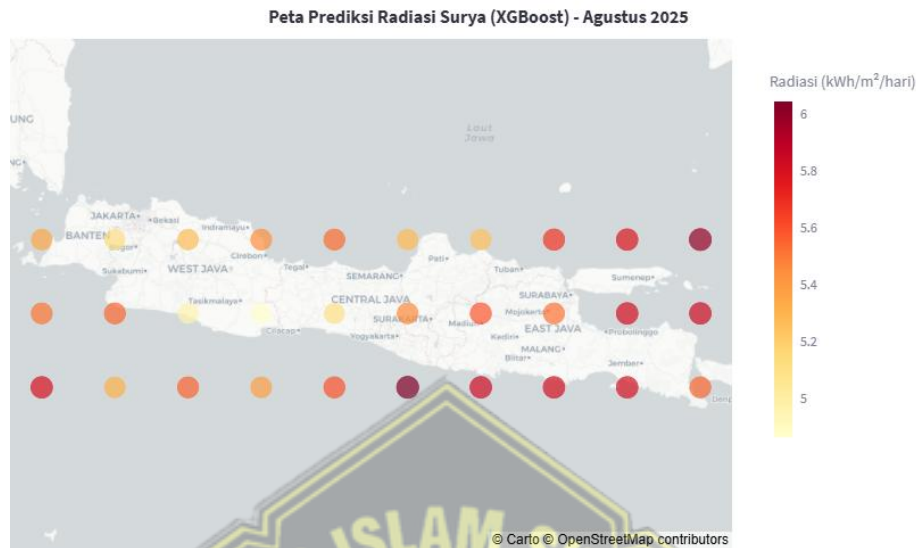
Berdasarkan hasil pada gambar 4.2 disajikan peta sebaran spasial prediksi radiasi surya di perairan selatan Pulau Jawa. Dari peta tersebut, terlihat adanya variasi nilai radiasi surya di seluruh titik prediksi. Wilayah dengan potensi radiasi tertinggi, yang ditandai dengan warna merah tua (mendekati 6.2 kWh/m²/hari), secara konsisten teridentifikasi pada area laut yang lebih jauh dari garis pantai, khususnya di perairan selatan Provinsi Jawa Tengah dan Jawa Timur. Sebaliknya, nilai radiasi yang relatif lebih rendah (ditandai dengan warna kuning hingga oranye, berkisar antara 5.0 hingga 5.8 kWh/m²/hari) teramati pada titik-titik yang lokasinya lebih dekat dengan pesisir, terutama di sebelah selatan Provinsi Banten dan Jawa Barat.

LAT	LON	PREDICTION
-6.5	105.5	5.40
-6.5	106.5	5.14
-6.5	107.5	5.21
-6.5	108.5	5.51
-6.5	109.5	5.58
-6.5	110.5	5.73
-6.5	111.5	5.65
-6.5	112.5	6.01
-6.5	113.5	5.98
-6.5	114.5	6.20
-7.5	105.5	5.54
-7.5	106.5	5.65
-7.5	107.5	5.03
-7.5	108.5	5.00
-7.5	109.5	5.12
-7.5	110.5	5.53
-7.5	111.5	5.69
-7.5	112.5	5.54
-7.5	113.5	5.99
-7.5	114.5	6.12
-8.5	105.5	5.81
-8.5	106.5	5.48
-8.5	107.5	5.74
-8.5	108.5	5.56
-8.5	109.5	5.67
-8.5	110.5	5.94
-8.5	111.5	5.75
-8.5	112.5	5.72
-8.5	113.5	5.70
-8.5	114.5	5.46

Gambar 4. 3 Hasil Prediksi Radiasi Surya *Bagging Base model Decsion Tree*

Gambar 4.3 Berdasarkan data hasil prediksi *bagging base model decision tree* menghasilkan rentang nilai antara 5.00 kWh/m²/hari hingga 6.20 kWh/m²/hari. Nilai prediksi tertinggi (6.20 kWh/m²/hari) teridentifikasi pada koordinat (-6.5, 114.5), yang berlokasi di perairan lepas pantai sebelah tenggara Pulau Jawa. Sementara itu, nilai prediksi terendah (5.00 kWh/m²/hari) ditemukan pada koordinat (-7.5, 107.5), di perairan selatan Jawa Barat. Secara keseluruhan, model Bagging menunjukkan kecenderungan untuk memberikan prediksi dengan variasi yang cukup tinggi dan mampu mencapai nilai puncak yang paling optimis, yang mengindikasikan potensi energi surya yang sangat signifikan di wilayah timur.

3. Hasil Prediksi Model XGBoost



Gambar 4. 4 Peta Prediksi Radiasi Surya XGBoost

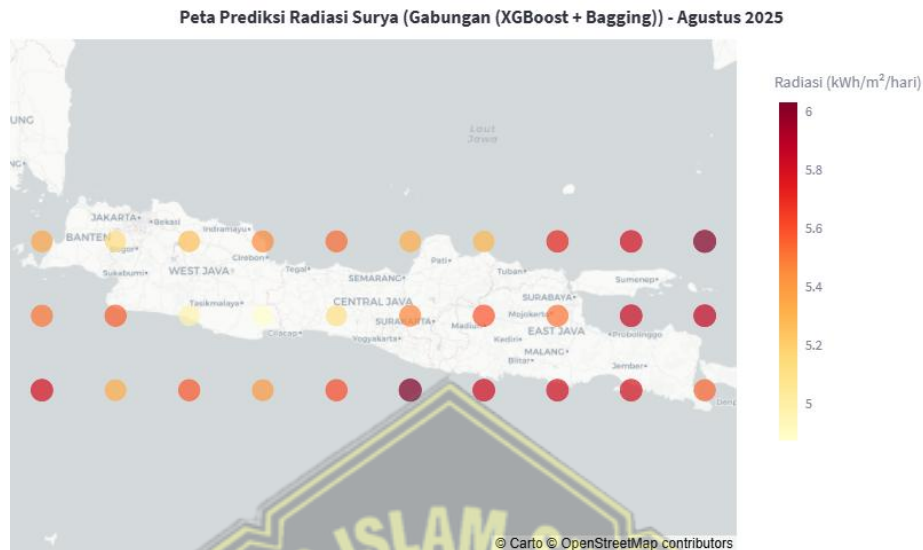
Gambar 4.4 menampilkan peta sebaran spasial hasil prediksi radiasi surya di perairan selatan Pulau Jawa untuk Agustus 2025, yang dihasilkan menggunakan model XGBoost. Peta ini mengilustrasikan variasi potensi energi surya dengan rentang nilai antara 5.0 kWh/m²/hari (ditandai warna kuning) hingga 6.0 kWh/m²/hari (ditandai warna merah tua). Berdasarkan pola yang terlihat, wilayah dengan potensi radiasi tertinggi secara konsisten berada di area laut yang lebih jauh dari garis pantai, khususnya di perairan lepas pantai Jawa Tengah dan Jawa Timur. Sebaliknya, nilai radiasi yang lebih rendah ditemukan pada titik-titik yang lebih dekat dengan pesisir, terutama di selatan Banten dan Jawa Barat, yang menunjukkan bahwa lokasi lepas pantai lebih optimal untuk pemanfaatan energi surya.

LAT	LON	PREDICTION
-6.5	105.5	5.35
-6.5	106.5	5.12
-6.5	107.5	5.27
-6.5	108.5	5.45
-6.5	109.5	5.53
-6.5	110.5	5.28
-6.5	111.5	5.25
-6.5	112.5	5.67
-6.5	113.5	5.78
-6.5	114.5	5.99
-7.5	105.5	5.51
-7.5	106.5	5.54
-7.5	107.5	4.96
-7.5	108.5	4.86
-7.5	109.5	5.10
-7.5	110.5	5.46
-7.5	111.5	5.58
-7.5	112.5	5.51
-7.5	113.5	5.81
-7.5	114.5	5.83
-8.5	105.5	5.80
-8.5	106.5	5.31
-8.5	107.5	5.55
-8.5	108.5	5.38
-8.5	109.5	5.59
-8.5	110.5	6.04
-8.5	111.5	5.83
-8.5	112.5	5.81
-8.5	113.5	5.80
-8.5	114.5	5.54

Gambar 4. 5 Hasil Prediksi XGBoost

Gambar 4.5 menyajikan hasil prediksi radiasi surya menggunakan model XGBoost. Berdasarkan data yang ditampilkan, model ini menghasilkan prediksi dengan rentang nilai antara 4.86 kWh/m²/hari hingga 6.04 kWh/m²/hari. Nilai prediksi tertinggi, yaitu 6.04 kWh/m²/hari, teridentifikasi pada koordinat (-8.5, 110.5), yang berlokasi di perairan selatan Jawa Tengah. Sementara itu, titik dengan nilai prediksi terendah, yaitu 4.86 kWh/m²/hari, ditemukan pada koordinat (-7.5, 107.5) di perairan selatan Jawa Barat. Meskipun nilai puncaknya sedikit lebih konservatif dibandingkan model Bagging, XGBoost menunjukkan kemampuannya dalam mengidentifikasi titik dengan potensi terendah secara lebih tajam, yang berguna untuk analisis variabilitas energi.

4. Hasil Prediksi Gabungan *Ensemble Learning*



Gambar 4. 6 Grafik Hasil Prediksi Penggabungan *Ensemble Learning*

Gambar 4.6 Sebagai hasil akhir dari proses pemodelan, menampilkan peta prediksi radiasi surya yang dihasilkan oleh model gabungan (*ensemble*) antara XGBoost dan *bagging base model decision tree*. Model gabungan ini bertujuan untuk mengkombinasikan keunggulan dari masing-masing model individu untuk menghasilkan prediksi yang lebih robust dan akurat. Hasil visualisasi menunjukkan pola sebaran yang merupakan perpaduan dari kedua model sebelumnya, dengan rentang nilai prediksi antara 5.0 hingga 6.0 kWh/m²/hari.

Secara spasial, potensi radiasi tertinggi (ditandai warna merah tua, mendekati 6.0 kWh/m²/hari) teridentifikasi secara dominan di perairan selatan Provinsi Jawa Timur dan sebagian Jawa Tengah. Sementara itu, nilai yang lebih rendah terdistribusi di bagian barat, khususnya di perairan selatan Banten dan Jawa Barat, serta pada beberapa titik yang lebih dekat dengan garis pantai di utara grid prediksi. Pola ini mengkonfirmasi bahwa wilayah laut lepas di bagian timur secara konsisten menunjukkan potensi energi surya tertinggi. Hasil dari model gabungan ini menjadi output utama penelitian yang dapat digunakan sebagai rekomendasi paling akurat untuk penentuan lokasi optimal instalasi PLTS terapung.

LAT	LON	PREDICTION
-6.5	105.5	5.36
-6.5	106.5	5.13
-6.5	107.5	5.26
-6.5	108.5	5.46
-6.5	109.5	5.53
-6.5	110.5	5.32
-6.5	111.5	5.29
-6.5	112.5	5.71
-6.5	113.5	5.80
-6.5	114.5	6.01
-7.5	105.5	5.51
-7.5	106.5	5.56
-7.5	107.5	4.97
-7.5	108.5	4.87
-7.5	109.5	5.10
-7.5	110.5	5.47
-7.5	111.5	5.59
-7.5	112.5	5.51
-7.5	113.5	5.83
-7.5	114.5	5.86
-8.5	105.5	5.80
-8.5	106.5	5.33
-8.5	107.5	5.57
-8.5	108.5	5.40
-8.5	109.5	5.60
-8.5	110.5	6.03
-8.5	111.5	5.82
-8.5	112.5	5.80
-8.5	113.5	5.79
-8.5	114.5	5.54

Gambar 4. 7 Hasil Prediksi Penggabungan *Ensemble Learning*

Sebagai tahap akhir, dilakukan pendekatan model gabungan (*ensemble*) dengan mengkombinasikan hasil dari model *Bagging* dan XGBoost, seperti terlihat pada data di gambar 4.7 Rincian data dari model gabungan ini menghasilkan prediksi dengan rentang nilai antara 4.87 kWh/m²/hari hingga 6.03 kWh/m²/hari. Hasil ini merefleksikan karakteristik perpaduan dari kedua model dasar, di mana nilai puncaknya (6.03 kWh/m²/hari) dan nilai terendahnya (4.87 kWh/m²/hari) berada di antara atau sedikit menyesuaikan nilai ekstrem dari model-model individu. Dengan menyeimbangkan hasil prediksi, model gabungan ini bertujuan untuk menghasilkan output yang lebih stabil dan robust, mengurangi potensi overfitting atau bias dari satu model tunggal, sehingga dapat dianggap sebagai representasi prediksi yang paling andal dalam penelitian ini.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. Implementasi *bagging* dengan *base model decision tree* menghasilkan prediksi radiasi surya dengan rentang nilai antara 5.00 kWh/m²/hari hingga 6.20 kWh/m²/hari. Model ini terbukti efektif dalam mengidentifikasi wilayah dengan potensi energi surya yang sangat signifikan, khususnya di perairan sebelah tenggara Pulau Jawa.
2. Implementasi dengan XGBoost menghasilkan prediksi dengan rentang antara 4.86 kWh/m²/hari hingga 6.04 kWh/m²/hari. Keunggulan model ini adalah ketajamannya yang lebih baik dalam mendeteksi titik dengan potensi radiasi terendah, seperti di perairan selatan Jawa Barat, yang krusial untuk analisis variabilitas.
3. Evaluasi kinerja model gabungan terbaik hasil evaluasi komprehensif menggunakan MSE, RMSE, dan R² Score menunjukkan bahwa model gabungan memberikan performa paling unggul. Model gabungan ini diimplementasikan menggunakan metode *weighted average*. Hasil ini menegaskan bahwa pendekatan *ensemble* dengan pembobotan yang tepat adalah strategi terbaik untuk mencapai akurasi prediksi tertinggi pada data penelitian ini.

5.2 Saran

1. Penelitian selanjutnya disarankan untuk menggabungkan metode lain seperti LSTM atau model untuk menangani data *time series* yang lebih kompleks.
2. Model dapat dikembangkan lebih lanjut dengan penambahan fitur atmosferik lain, seperti kelembapan, suhu, atau tutupan awan untuk meningkatkan akurasi.
3. Platform prediksi dapat dikembangkan menjadi aplikasi web interaktif yang dapat digunakan oleh instansi atau masyarakat untuk perencanaan energi terbarukan.

DAFTAR PUSTAKA

- Algiffary, A., & Sutabri, T. (2023). Analisis Perbandingan Kinerja Metode Ensemble Bagging dan Boosting pada Klasifikasi Bantuan Subsidi Listrik di Kabupaten/Kota Bogor. *Indonesian Journal of Computer Science*, 12(2), 284–301. <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- Ali, A. Al, Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique. *Applied Sciences (Switzerland)*, 13(4). <https://doi.org/10.3390/app13042272>
- Annisa Nur Oktaviani. (2025). *Analisis Prediksi Radiasi Matahari Dengan Algoritma Machine Learning Dan Implementasi Bayesian Optimization*. 8, 45–58.
- Arif, A. F. N., & Nurwati, N. (2022). Pengaruh Konsentrasi Penduduk Indonesia Di Pulau Jawa Terhadap Kesejahteraan Masyarakat. *Jurnal Ilmu Kesejahteraan Sosial HUMANITAS*, 4(I), 54–70. <https://doi.org/10.23969/humanitas.v4ii.3920>
- Aziz, M. I. (2025). *Implementasi Machine Learning Untuk Deteksi Anomali Kesehatan Janin Menggunakan Metode Ensemble Berbasis Decision Tree masa kehamilan yang selanjutnya atau kedepannya (Salwa Darin L , 2018). daripada bentuk dan suatu tren yang cenderung untuk menganalisis*.
- Deqita, A. (2022). Artikel Analisis Intensitas Radiasi Matahari Dan Peningkatan Suhu Lingkungan. *Jurnal Pendidikan Fisika Dan Sains (JPFS)*, 5(2), 76–82. <https://doi.org/10.52188/jpfs.v5i2.237>
- Dwisari, V., Sudarti, S., & Yushardi, Y. (2023). Pemanfaatan Energi Matahari: Masa Depan Energi Terbarukan. *OPTIKA: Jurnal Pendidikan Fisika*, 7(2), 376–384. <https://doi.org/10.37478/optika.v7i2.3322>
- Hanif, H., & Utomo, D. W. (2025). Prediksi Diabetes menggunakan Metode Ensemble Learning dengan Teknik Soft Voting. *Infotekmesin*, 16(01), 127–134. <https://doi.org/10.35970/infotekmesin.v16i1.2534>
- Helmut, A., & Triantoro Murdiansyah, D. (2020). *Klasifikasi email multi kelas menggunakan ensemble bagging*. 7(1), 2498–2505.

- Ismayanti, R., & Baihaqi, W. M. (2024). *Prediksi Potensi Suatu Wilayah Untuk Menjadi PLTS Dengan Machine Learning*. 1(2), 66–72.
- Majid, A. M., & Nawangsih, I. (2024). Perbandingan Metode Ensemble Untuk Meningkatkan Akurasi Algoritma Machine Learning Dalam Memprediksi Penyakit Breast Cancer (Kanker Payudara). *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 23(1), 97. <https://doi.org/10.53513/jis.v23i1.9563>
- Mia, Nur Masruriyah, A. F., & Pratama, A. R. (2024). Komparasi Model DecisionTree dan Random Forest untuk Memprediksi Penyakit Jantung. *Scientific Student Journal for Information, Technology and Science*, V(2), 123–130.
- Mohbey, K. K., Khan, M. Z., & Indian, A. (2022). Credit Card Fraud Prediction Using XGBoost. *International Journal of Information Retrieval Research*, 12(2), 1–17. <https://doi.org/10.4018/ijirr.299940>
- Mukaromah, H., Studi, P., Teknik, M., Meneng, G., & Lampung, K. B. (2024). *Komparasi Teknik Bagging Dan Adaboost Pada Decision Tree Dan Naive Bayes Untuk Prediksi Stroke* 1,2. 167–180.
- Murdiansyah, D. T. (2024). *Prediksi Stroke Menggunakan Extreme Gradient Boosting Stroke Prediction Using Extreme Gradient Boosting*. 2, 419–426. <https://doi.org/10.26798/jiko.v8i2.1295>
- Rama, M., Suryanto, H., & Utomo, D. W. (2024). *Pembelajaran Ensemble Untuk Klasifikasi Ulasan Pelanggan E-commerce Menggunakan Teknik Boosting*. 15(02), 238–244. <https://doi.org/10.35970/infotekmesin.v15i2.2314>
- Rayadin, M. A., Musaruddin, M., Saputra, R. A., & Isnawaty, I. (2024). Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 5(2), 111–119.
- Rifai, A., Permana, S. E., Hamonangan, R., Informatika, T., Informatika, M., Forest, R., Parameters, E. O., & Utara, M. (2024). *METODE ENSEMBLE DAN ALGORITMA RANDOM FOREST*. 8(4), 8226–8234.
- Rizkiani, E., & Brahma, D. (2024). *Perbandingan Performa Algoritma Metode*

Bagging dan Boosting pada Prediksi Konsentrasi PM 10 di Jakarta Utara. 01, 74–81.

Sepbriant, G. D., & Utomo, D. W. (2024). Ensemble Learning pada Kategorisasi Produk E-Commerce Menggunakan Teknik Boosting. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(2), 123–133. <https://doi.org/10.14421/jiska.2024.9.2.123-133>

Sindar, A., Ramen, S., & Mulyani, S. (2024). *Teknik Ensemble Dalam Machine Learning Untuk Menentukan Tingkat Akurasi Perkembangan Motorik Bayi 0-12 Bulan. 23(MI), 418–425.*

Siregar, A. M. (2020). Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning. *Petir, 13(2)*, 138–147. <https://doi.org/10.33322/petir.v13i2.998>

Soelistijadi, R., Dwiati Wismarini, T., Eniyati, S., & I4, S. (2024). *Pemodelan Prediktif Menggunakan Metode Ensemble Learning XGBoost dalam Peningkatan Akurasi Klasifikasi Penyakit Ginjal. 5(4), 1866–1875.*

Sudarman, E. J., & Budi, S. (2023). *Pengembangan Model Kecerdasan Mesin Extreme Gradient Boosting untuk Prediksi Keberhasilan Studi Mahasiswa. 5(November), 297–314.*

