

BAB II TINJAUAN PUSTAKA

2.1 Teori dan Definisi

2.1.1 Curah Hujan

Curah Hujan dan Presipitasi adalah dua istilah penting dalam ilmu meteorologi yang sering digunakan untuk menggambarkan fenomena air yang jatuh ke permukaan bumi. Presipitasi adalah istilah yang lebih luas, mencakup semua bentuk air yang turun dari atmosfer ke bumi, termasuk hujan, salju, es, embun, dan kabut. Semua curah hujan adalah presipitasi, tetapi tidak semua presipitasi adalah curah hujan. Curah hujan adalah jumlah air dalam bentuk cair (hujan) yang jatuh ke tanah selama periode tertentu. Pengukuran dilakukan dalam satuan milimeter (mm), mewakili kedalaman air yang terkumpul di permukaan datar jika tidak ada penguapan atau peresapan.

Dalam hidrologi, curah hujan merujuk pada jumlah air yang jatuh ke permukaan bumi dalam bentuk hujan selama periode tertentu. Curah hujan diukur dalam satuan kedalaman, biasanya milimeter (mm) atau inci. Curah hujan mencakup berbagai bentuk presipitasi seperti hujan, salju, hujan es (hail), dan embun. Beberapa faktor yang mempengaruhi curah hujan antara lain geografi, iklim, dan siklus musiman. Curah hujan memainkan peran penting dalam

pengelolaan sumber daya air, pertanian, serta mitigasi bencana alam seperti banjir dan kekeringan (Allawi *et al.*, 2023).

2.1.2 Faktor yang Mempengaruhi Curah Hujan

Curah hujan dipengaruhi oleh sejumlah faktor meteorologis yang saling terkait. Penelitian ini akan menganalisis lima faktor utama yang memengaruhi curah hujan, yaitu temperatur rata-rata, kelembapan rata-rata, lamanya penyinaran matahari, kecepatan angin rata-rata, dan curah hujan itu sendiri.

1. Temperatur Rata-Rata



Gambar 2. 1 Thermometer

Sumber : (<https://staklim-sumsel.bmkg.go.id/termometer-tanah-digital/>)

Temperatur adalah ukuran rata-rata energi kinetik partikel dalam suatu substansi, yang menunjukkan seberapa panas atau dingin suatu zat tersebut. Dalam meteorologi, temperatur berperan penting dalam pola cuaca dan iklim karena dapat mempengaruhi pergerakan massa udara, proses pembentukan awan, dan intensitas curah hujan. Menurut teori kinetik gas, suhu suatu gas

berkorelasi langsung dengan energi kinetik partikel gas tersebut. Partikel gas bergerak lebih cepat pada suhu tinggi, meningkatkan kemungkinan terjadinya kondensasi yang akhirnya menghasilkan hujan. Perubahan suhu juga mempengaruhi tekanan atmosfer yang berperan dalam pergerakan angin (Yi Lim *et al.*, 2020)

2. Kelembapan Rata-Rata



Gambar 2. 2 Hygrometer

Sumber : (<https://envilife.co.id/alat-untuk-mengukur-kelembaban-udara/>)

Kelembapan dalam meteorologi merujuk pada jumlah uap air yang terdapat dalam udara. Kelembapan spesifik mengukur massa uap air yang terkandung dalam sejumlah udara, sedangkan kelembapan relatif menunjukkan perbandingan antara jumlah uap air yang ada dengan kapasitas maksimal udara pada suhu tertentu. Kelembapan berperan penting dalam proses atmosfer, seperti pembentukan awan dan curah hujan. Ketika udara mencapai kelembapan jenuh, uap air akan terkondensasi dan dapat membentuk presipitasi dalam

bentuk hujan, salju, atau bentuk lainnya. Proses kondensasi ini sangat dipengaruhi oleh suhu dan tekanan atmosfer (Zuo *et al.*, 2022)

3. Lamanya Penyinaran Matahari



Gambar 2. 3 Campbell-Stokes

Sumber : (<https://mardevitaislamiyakti.blogspot.com/2016/06/alat-agroklimatologi-solarimeter-type.html>)

Lamanya Penyinaran Matahari merujuk pada durasi waktu di mana permukaan bumi menerima sinar matahari secara langsung dalam sehari. Faktor ini berperan penting dalam pembentukan suhu udara, proses penguapan, serta dinamika atmosfer. Penyinaran matahari yang lebih lama meningkatkan pemanasan permukaan bumi, yang pada gilirannya mempercepat penguapan air dari permukaan tanah, laut, atau danau. Hal ini

berkontribusi pada peningkatan kelembapan udara dan pembentukan awan, yang dapat menyebabkan presipitasi (curah hujan).

Durasi penyinaran matahari merupakan salah satu faktor penting setiap hari, seperti pertanian atau energi terbarukan. Durasi penyinaran matahari juga dipengaruhi oleh posisi geografis dan waktu dalam setahun, karena daerah yang lebih dekat dengan ekuator menerima penyinaran yang lebih lama dan lebih intens dibandingkan daerah yang lebih dekat dengan kutub. Penelitian ini menunjukkan bahwa variasi panjang hari dan intensitas sinar matahari sangat mempengaruhi pola iklim dan cuaca, termasuk curah hujan, di berbagai belahan dunia. Panjang hari yang lebih panjang selama musim panas di wilayah dengan iklim temperate dapat meningkatkan peluang terbentuknya awan dan curah hujan. Sebaliknya, di wilayah dengan musim dingin yang panjang, waktu penyinaran yang lebih pendek dapat mengurangi jumlah penguapan dan pembentukan awan (Mulyadi and Djamal, 2019).

4. Kecepatan Angin Rata-Rata



Gambar 2. 4 Anemometer

Sumber : (<https://staklim-sumsel.bmkg.go.id/anemometer-0-5-m/>)

Kecepatan angin mengacu pada pergerakan udara di atmosfer dengan kecepatan tertentu, biasanya diukur dalam meter per detik (m/s) atau kilometer per jam (km/h). Dalam meteorologi, kecepatan angin memengaruhi distribusi uap air, perubahan suhu, dan tekanan atmosfer, yang bersama-sama berkontribusi pada pembentukan pola cuaca seperti hujan atau badai. Faktor-faktor seperti gradien tekanan, rotasi bumi (gaya Coriolis), serta gesekan dengan permukaan bumi memengaruhi kecepatan angin. Pemahaman tentang kecepatan angin sangat penting dalam mempelajari dinamika atmosfer dan dampaknya pada cuaca global (Soman *et al.*, 2010).

5. Curah Hujan



Gambar 2. 5 Rain Gauge

Sumber : (<https://www.mertani.co.id/post/ombrometer-pengertian-kelebihan-kekurangan-dan-peran-dalam-pertanian>)

Curah hujan merujuk pada jumlah air yang jatuh ke permukaan bumi dalam bentuk cair, seperti hujan, atau dalam bentuk padat, seperti salju dan hujan es. Pengukuran curah hujan dilakukan dalam satuan milimeter (mm), yang menunjukkan kedalaman air yang terkumpul di permukaan yang datar. Curah hujan dipengaruhi oleh sejumlah faktor, termasuk suhu udara, kelembapan, tekanan atmosfer, dan pergerakan angin. Ketika udara jenuh dengan uap air, kondensasi terjadi, membentuk awan yang akhirnya menghasilkan presipitasi. Faktor-faktor seperti topografi juga mempengaruhi pola curah hujan di suatu wilayah, terutama di daerah pegunungan yang cenderung menerima lebih banyak hujan (G *et al.*, 2023).

2.1.3 Alat Ukur Curah Hujan



Gambar 2. 6 Alat Ukur Curah Hujan Otomatis

Sumber : (<https://staklim-sumsel.bmkg.go.id/penakar-hujan-type-hellman/>)

Alat pengukur curah hujan (*rain gauge*) digunakan untuk mengumpulkan air hujan yang jatuh selama periode tertentu dan mengukur jumlahnya dalam milimeter (mm). Beberapa jenis alat yang sering digunakan antara lain:

1. **Tipping Bucket Rain Gauge (TBRG)**

Alat ini menggunakan ember kecil yang terbalik setiap kali penuh dengan air hujan, mengubahnya menjadi sinyal yang dicatat untuk mengukur curah hujan. TBRG cocok digunakan untuk pengukuran jangka panjang dengan interval waktu yang lebih kecil.

2. Standard Rain Gauge

Alat ini terdiri dari corong yang mengalirkan air hujan ke dalam tabung pengukur. Pengukuran dilakukan dengan menghitung tinggi kolom air dalam tabung. Alat ini sederhana namun efektif untuk pengukuran manual di stasiun cuaca atau daerah dengan cuaca stabil.

3. Optical Rain Gauge

Alat ini menggunakan sensor optik untuk mendeteksi tetesan hujan. Cahaya yang dipancarkan akan dipantulkan oleh tetesan hujan, memungkinkan pengukuran intensitas hujan dengan akurasi tinggi, terutama dalam kondisi hujan ringan hingga sedang.

Tantangan dalam Pengukuran Curah Hujan

1. Variasi Lokasi yaitu curah hujan dapat berbeda secara signifikan antar lokasi, dipengaruhi oleh faktor topografi dan kondisi atmosfer lokal.
2. Perubahan Iklim dapat mengubah pola curah hujan, sehingga memerlukan alat yang lebih adaptif dan akurat untuk mengukur curah hujan dalam kondisi yang berubah-ubah.

2.1.4 Stasiun Pengukur Cuaca (Automatic Weather Station (AWS))

Stasiun Pengukur Cuaca atau Automatic Weather Station (AWS) adalah perangkat yang dirancang untuk mengumpulkan data meteorologi secara otomatis menggunakan berbagai sensor canggih. AWS berfungsi untuk memantau parameter

cuaca seperti curah hujan, suhu, kelembapan, tekanan udara, kecepatan angin, dan radiasi matahari. Data yang dihasilkan dari alat ini sangat penting untuk mendukung analisis iklim dan peramalan cuaca di berbagai wilayah, termasuk Jawa Timur (Gaikwad *et al.*, 2023).

Di Jawa Timur, AWS tersebar di sejumlah lokasi strategis yang dikelola oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), universitas, dan lembaga riset lainnya. Beberapa kota yang memiliki stasiun AWS antara lain Surabaya, Malang, Kediri, Tulungagung, Nganjuk, dan Madiun. Di Kediri, AWS digunakan untuk mengamati data iklim secara berkesinambungan, mendukung pengelolaan sumber daya air, serta meningkatkan kemampuan prediksi fenomena cuaca ekstrem seperti hujan deras atau kekeringan.

AWS biasanya dilengkapi dengan alat pengukur curah hujan otomatis (rain gauge), sensor suhu dan kelembapan, serta barometer untuk tekanan udara. Data yang diperoleh dapat diproses menggunakan berbagai metode analisis, termasuk teknik regresi lanjutan seperti Advanced Regression, untuk membangun model prediksi yang lebih akurat. Misalnya, di Kediri, data AWS dapat digunakan untuk memprediksi curah hujan berdasarkan variabel suhu dan kelembapan, mendukung keputusan terkait pengelolaan irigasi dan pencegahan banjir.

Keunggulan AWS terletak pada kemampuannya mengurangi ketergantungan pada pengukuran manual, sehingga mempercepat proses pengumpulan data dan meningkatkan akurasi analisis. Data AWS juga sering kali

dikombinasikan dengan citra satelit dan model numerik untuk memberikan hasil yang lebih komprehensif.

2.1.5 BMKG (Badan Meteorologi, Klimatologi, dan Geofisika)

BMKG adalah lembaga pemerintah yang berfungsi sebagai pusat pengumpulan, pengelolaan, dan analisis data cuaca, iklim, dan geofisika di Indonesia. Dengan jaringan stasiun meteorologi yang tersebar di seluruh wilayah nusantara, BMKG memiliki peran strategis dalam menyediakan informasi cuaca yang mendukung perencanaan, mitigasi bencana, dan pembangunan nasional. Data yang dikumpulkan meliputi suhu udara, curah hujan, kelembapan, kecepatan angin, arah angin, serta durasi penyinaran matahari. Informasi ini sangat penting untuk berbagai sektor, termasuk transportasi, pertanian, dan perencanaan infrastruktur.

BMKG mengoperasikan berbagai model prakiraan cuaca berbasis data numerik yang diperoleh dari stasiun pengamatan. Misalnya, data suhu digunakan untuk memprediksi pembentukan awan, sementara data kelembapan dapat memproyeksikan potensi hujan di suatu wilayah. Selain itu, kecepatan angin yang dikumpulkan membantu mempelajari pola sirkulasi udara dan mendeteksi perubahan cuaca ekstrem. Teknologi yang digunakan BMKG terus diperbarui agar hasil prediksi semakin akurat, termasuk penggunaan algoritma berbasis machine learning (Furizal, 2024).

Data BMKG juga sangat relevan dalam penelitian ini. Dengan memanfaatkan informasi meteorologi dari BMKG, model prediksi curah hujan berbasis Advanced Regression dapat dikembangkan untuk menghasilkan hasil yang

lebih presisi. Model ini memungkinkan peneliti untuk menganalisis hubungan antar variabel iklim dengan lebih mendalam. Selain itu, data yang disediakan BMKG mendukung mitigasi risiko bencana alam, seperti banjir dan kekeringan, yang sering melanda wilayah tertentu, termasuk Kediri Raya. Dengan demikian, peran BMKG tidak hanya terbatas pada penyediaan data, tetapi juga mendukung pengambilan keputusan berbasis bukti untuk pembangunan yang lebih berkelanjutan.

2.1.6 Meteorologi

Meteorologi adalah cabang ilmu yang mempelajari atmosfer Bumi, termasuk fenomena fisik dan kimia yang memengaruhi pola cuaca dan iklim. Ilmu ini mencakup berbagai aspek, mulai dari analisis tekanan udara, suhu, dan kelembapan hingga radiasi matahari yang memengaruhi perubahan atmosfer. Meteorologi menjadi dasar penting dalam memahami proses yang mendasari terjadinya hujan, badai, kekeringan, dan fenomena cuaca ekstrem.

Secara umum, meteorologi terbagi menjadi beberapa subbidang utama. Meteorologi Dinamis mengkaji interaksi gaya-gaya fisik, seperti gaya Coriolis dan gradien tekanan, yang memengaruhi pola sirkulasi udara dan distribusi suhu. Meteorologi Fisik berfokus pada proses kondensasi, penguapan, dan interaksi antara energi matahari dengan atmosfer, yang penting untuk memahami pembentukan awan dan curah hujan. Sementara itu, Meteorologi Kimia mempelajari reaksi kimia di atmosfer, termasuk pengaruh polusi udara, ozon, dan gas rumah kaca terhadap kualitas udara dan perubahan iklim (Xiao *et al.*, 2024).

Dalam konteks penelitian prediksi cuaca, meteorologi menjadi dasar ilmiah untuk mengembangkan model prediksi yang lebih akurat. Data yang dikumpulkan melalui pengamatan meteorologi digunakan untuk menganalisis pola cuaca dan perubahan iklim, memberikan wawasan yang penting bagi mitigasi risiko bencana seperti banjir atau badai. Dengan pendekatan berbasis meteorologi, penelitian ini bertujuan untuk memahami lebih dalam dinamika atmosfer dan bagaimana faktor-faktor iklim memengaruhi curah hujan di Kediri Raya. Hal ini tidak hanya mendukung pengelolaan sumber daya alam, tetapi juga membantu dalam pengambilan keputusan berbasis bukti untuk pembangunan yang lebih berkelanjutan.

2.1.7 Advanced Regression

Advanced Regression adalah teknik statistik lanjutan yang digunakan untuk menganalisis hubungan antara variabel dalam dataset yang kompleks. Teknik ini mencakup metode seperti ridge regression, lasso regression, dan model ensemble seperti random forest. Advanced Regression memberikan fleksibilitas dalam memodelkan hubungan non-linear antar variabel, menjadikannya alat yang sangat berguna dalam memprediksi pola cuaca dan iklim. Pendekatan ini dirancang untuk menangani dataset dengan multikolinearitas tinggi, variabel dengan distribusi non-normal, dan hubungan yang tidak linier. Dalam konteks prediksi cuaca, Advanced Regression dapat digunakan untuk mengidentifikasi faktor-faktor iklim utama, seperti suhu, kelembapan, dan kecepatan angin, yang memengaruhi curah hujan secara langsung. Misalnya, dalam studi iklim, metode regresi penalti seperti lasso digunakan untuk memilih variabel paling relevan dari sejumlah besar data

klimatologi, sedangkan model ensemble membantu meningkatkan akurasi dengan menggabungkan prediksi dari berbagai model individu.

Keunggulan Advanced Regression terletak pada kemampuannya mengatasi keterbatasan model konvensional. Misalnya, regresi linear standar tidak dapat menangani hubungan non-linear, sedangkan Advanced Regression dapat mengintegrasikan berbagai teknik statistik untuk menghasilkan prediksi yang lebih akurat. Hal ini penting dalam studi iklim, di mana interaksi variabel sering kali sangat kompleks. Dengan kemampuan ini, Advanced Regression menjadi metode yang sangat cocok untuk analisis data cuaca berskala besar dan berperan penting dalam mendukung mitigasi bencana berbasis data. Pendekatan ini memberikan kontribusi signifikan dalam penelitian iklim di Kediri Raya, membantu memodelkan hubungan kompleks antara curah hujan dan variabel iklim lainnya, sekaligus mendukung pengelolaan sumber daya alam secara lebih efektif (Das *et al.*, 2024).

2.1.8 Generalized Linear Model (GLM)

Generalized Linear Model (GLM) adalah metode statistik yang memperluas regresi linear klasik dengan memungkinkan data respons mengikuti distribusi non-normal, seperti distribusi Poisson, binomial, atau gamma. Dalam penelitian iklim, GLM sering digunakan untuk memodelkan hubungan antara curah hujan dan variabel iklim lainnya, seperti suhu, kelembapan, dan tekanan udara. GLM sangat efektif untuk memprediksi curah hujan karena fleksibilitasnya dalam menangani data dengan variasi besar dan distribusi yang tidak seragam.

Pendekatan GLM mencakup tiga komponen utama: fungsi hubungan (link function), distribusi data respons, dan regresi linear untuk prediktor. Dalam konteks prediksi curah hujan, fungsi hubungan log-link sering digunakan untuk menangkap hubungan eksponensial antara faktor-faktor iklim dan curah hujan. Distribusi gamma, misalnya, dapat digunakan untuk memodelkan data curah hujan yang biasanya memiliki variansi tinggi.

Keunggulan GLM adalah kemampuannya untuk mengakomodasi kompleksitas data tanpa kehilangan interpretabilitas model. Dalam penelitian ini, GLM diterapkan untuk memodelkan hubungan antara curah hujan dan variabel iklim di Kediri Raya, memungkinkan analisis yang lebih terfokus pada faktor utama yang memengaruhi pola curah hujan. Dengan demikian, GLM memberikan dasar yang kuat untuk memahami dinamika cuaca lokal sekaligus menghasilkan prediksi yang lebih akurat untuk mendukung mitigasi risiko terkait perubahan iklim (Ahmed and Mohamed, 2021).

2.1.9 Random Forest (RF)

Random Forest adalah algoritme pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi. Algoritme ini bekerja dengan membangun sejumlah besar pohon keputusan (decision trees) selama pelatihan dan menghasilkan output dengan cara agregasi, seperti pemungutan suara mayoritas untuk klasifikasi atau rata-rata untuk regresi. Pendekatan ini meningkatkan akurasi prediksi dan membantu mencegah overfitting dengan menggabungkan hasil dari berbagai pohon yang dilatih pada subset data yang berbeda. Dalam konteks prediksi curah hujan,

Random Forest telah diterapkan untuk menganalisis hubungan antara variabel iklim, seperti suhu, kelembapan, tekanan udara, dan kecepatan angin, dengan tingkat curah hujan. Studi menunjukkan bahwa Random Forest dapat menghasilkan prediksi yang lebih akurat dibandingkan dengan beberapa algoritme lain.

Keunggulan Random Forest meliputi kemampuannya dalam menangani data dengan variabel input yang banyak dan beragam, serta kemampuannya untuk mengidentifikasi interaksi kompleks antara variabel. Selain itu, algoritme ini relatif tahan terhadap overfitting, terutama ketika jumlah pohon yang digunakan cukup besar. Namun, salah satu kelemahannya adalah kebutuhan komputasi yang lebih tinggi dibandingkan dengan model yang lebih sederhana, terutama ketika diterapkan pada dataset yang sangat besar (Nandini *et al.*, 2022).

2.1.10 Vector Autoregression (VAR)

Vector Autoregression (VAR) adalah metode statistik untuk menganalisis data deret waktu dengan banyak variabel, yang memungkinkan pemodelan hubungan dinamis antar variabel tersebut. Dalam konteks penelitian iklim, VAR digunakan untuk memahami bagaimana perubahan pada variabel iklim seperti suhu, kelembapan, dan kecepatan angin dapat memengaruhi curah hujan di waktu mendatang. VAR sangat efektif dalam menangkap pengaruh lag atau waktu tunda antar variabel, yang penting dalam analisis iklim.

Model VAR bekerja dengan memodelkan setiap variabel dalam sistem sebagai fungsi dari nilai lag-nya sendiri dan variabel lainnya. Sebagai contoh, curah hujan pada suatu hari dapat dipengaruhi oleh suhu dan kelembapan pada hari

sebelumnya. Proses ini memungkinkan VAR untuk menangkap interaksi dinamis antar variabel secara menyeluruh. Selain itu, metode ini mampu mengakomodasi hubungan timbal balik antar variabel, sehingga memberikan gambaran yang lebih realistis tentang pola iklim.

Dalam penelitian ini, VAR digunakan untuk mengeksplorasi hubungan antara variabel iklim di Kediri Raya dan memberikan prediksi yang lebih terarah untuk curah hujan. Dengan mempertimbangkan pengaruh waktu tunda, VAR dapat membantu memprediksi pola cuaca ekstrem dan memberikan informasi penting untuk mitigasi risiko bencana seperti banjir. Metode ini tidak hanya meningkatkan akurasi prediksi, tetapi juga mendukung perencanaan sumber daya air yang lebih strategis (Md. *et al.*, 2023).

Perbedaan utama antara Gaussian Generalized Linear Model (GLM), Random Forest (RF), dan Vector Autoregression (VAR) dalam konteks analisis statistik dan pemodelan data dapat dirangkum dalam tabel berikut :

Tabel 2. 1 Perbedaan GLM, RF dan VAR

Aspek	GLM	RF	VAR
Konsep Dasar	Memperluas regresi linear klasik untuk menangani berbagai	Algoritme ensemble berbasis pohon keputusan yang menggabungkan	Bukan model regresi, tetapi teknik analisis varians untuk mengevaluasi

	<p>distribusi dari variabel dependen.</p> <p>Menggunakan fungsi link untuk menghubungkan prediktor linear dengan ekspektasi variabel respons.</p>	<p>hasil dari banyak pohon (dengan bootstrap) untuk meningkatkan akurasi prediksi.</p>	<p>kontribusi variabel prediktor terhadap variabilitas data.</p>
Kegunaan	<p>Berguna untuk data dengan distribusi non-normal, seperti distribusi Poisson, binomial, atau gamma. Ideal untuk hubungan linier yang kompleks dengan distribusi tertentu.</p>	<p>Cocok untuk menangani data besar dan non-linear, serta dapat menangkap hubungan kompleks antara variabel prediktor dan respons.</p>	<p>Berguna untuk menilai pengaruh variabel prediktor terhadap variabilitas variabel dependen.</p> <p>Memberikan wawasan tentang signifikansi faktor.</p>
Penerapan dalam Data Curah Hujan	<p>Sangat efektif untuk memodelkan</p>	<p>Dapat digunakan untuk memprediksi</p>	<p>Digunakan untuk menentukan variabel iklim</p>

	<p>curah hujan dengan distribusi tidak seragam, seperti distribusi gamma atau eksponensial. Fungsi link seperti log dapat menangkap hubungan non-linear.</p>	<p>curah hujan dengan memanfaatkan hubungan kompleks antar variabel iklim. Memberikan hasil yang sering kali lebih akurat pada dataset besar dan kompleks.</p>	<p>mana yang paling memengaruhi variabilitas curah hujan, memberikan wawasan penting tentang pengaruh variabel.</p>
Keunggulan	<p>Mampu menangani data dengan distribusi kompleks sambil mempertahankan interpretabilitas model.</p>	<p>Menghasilkan prediksi yang robust dengan overfitting minimal. Dapat menangani data dengan banyak fitur dan interaksi non-linear.</p>	<p>Memberikan wawasan tentang signifikansi variabel prediktor terhadap variabilitas data. Cocok untuk analisis pengaruh variabel dalam dataset.</p>
Kelemahan	<p>Membutuhkan pemahaman</p>	<p>Memerlukan daya komputasi tinggi</p>	<p>Tidak digunakan langsung untuk</p>

	mendalam tentang distribusi data dan fungsi link yang sesuai.	untuk dataset besar. Interpretasi model lebih sulit dibandingkan GLM atau VAR.	prediksi, lebih fokus pada analisis hubungan antar variabel dalam dataset.
--	---	--	--

Tabel 2. 2 Perbandingan GLM, RF dan VAR

Aspek	GLM	RF	VAR
Jenis Model	Model regresi yang diperluas	Algoritme ensemble berbasis pohon keputusan	Analisis variansi, bukan regresi
Distribusi	Beragam distribusi (Poisson, gamma, binomial)	Tidak bergantung pada asumsi distribusi	Tidak fokus pada distribusi
Aplikasi dalam Curah Hujan	Memodelkan curah hujan dengan distribusi non-normal	Memprediksi curah hujan dengan hubungan non-linear kompleks	Menilai signifikansi variabel prediktor

Fokus Utama	Memperluas regresi untuk distribusi lain	Menangkap hubungan kompleks antar variabel	Mengukur kontribusi variabel terhadap variabilitas
-------------	--	--	--

Berdasarkan analisis persamaan dan perbedaan antar metode, saya memilih Generalized Linear Model (GLM), Random Forest (RF), dan Vector Autoregression (VAR) karena ketiganya menawarkan pendekatan yang sesuai dengan kebutuhan penelitian prediksi curah hujan.

1. GLM sangat baik untuk analisis yang membutuhkan fleksibilitas terhadap distribusi non-normal dan kemudahan interpretasi model.
2. Random Forest (RF) memberikan keunggulan dalam menangani hubungan kompleks dan menghasilkan prediksi yang robust pada dataset yang besar.
3. Vector Autoregression (VAR) terbukti paling efektif dalam memberikan hasil prediksi dengan akurasi tertinggi berdasarkan metrik evaluasi, menjadikannya metode yang ideal untuk menilai kontribusi variabel terhadap variabilitas curah hujan.

Metode ini dipadukan untuk memastikan akurasi prediksi yang optimal sekaligus memberikan wawasan yang mendalam terhadap faktor-faktor yang memengaruhi curah hujan dalam penelitian ini.

2.1.11 Validasi Data

Validasi data adalah proses sistematis yang bertujuan untuk memastikan bahwa data yang digunakan dalam analisis atau proses pengambilan keputusan memenuhi standar kualitas tertentu. Data yang divalidasi memiliki karakteristik yang relevan, akurat, lengkap, dan konsisten sehingga dapat mendukung pengambilan keputusan yang lebih tepat. Proses ini dilakukan untuk mendeteksi kesalahan, kekurangan, atau ketidaksesuaian data sebelum data tersebut digunakan lebih lanjut dalam analisis, pemodelan, atau pelaporan.

Validasi data juga mencakup proses pengujian terhadap keandalan data, baik secara manual maupun otomatis, dengan memanfaatkan perangkat lunak atau algoritma tertentu. Dengan validasi yang baik, analisis yang dilakukan dapat menghasilkan prediksi atau kesimpulan yang lebih andal dan akurat, meminimalkan risiko kesalahan yang dapat memengaruhi keputusan atau hasil akhir.

Transformasi data adalah proses mengubah atau memodifikasi data mentah menjadi bentuk yang lebih sesuai untuk dianalisis. Transformasi dilakukan untuk meningkatkan kualitas dan konsistensi data, sehingga memungkinkan analisis yang lebih mendalam dan komprehensif. Selain itu, transformasi data juga bertujuan untuk:

- Mengubah skala pengukuran data agar sesuai dengan asumsi analisis.
- Menjadikan distribusi data lebih normal untuk mempermudah penerapan metode statistik tertentu.
- Mengatasi masalah asimetri atau heteroskedastisitas dalam data.

- Menyelaraskan berbagai metrik sehingga memungkinkan analisis yang lebih seragam.
- Memudahkan aksesibilitas data untuk mendukung pengambilan keputusan berbasis data.

Transformasi data sering kali menjadi bagian dari proses ETL (Ekstraksi, Transformasi, Pemuatan) atau ELT. Untuk memastikan transformasi data dilakukan secara benar, penting untuk menyesuaikan teknik transformasi dengan karakteristik data dan tujuan analisis. Konsultasi dengan ahli statistik atau analis data dapat membantu memilih pendekatan yang paling tepat.

Transformasi logaritma adalah salah satu metode transformasi yang sering digunakan untuk menangani data yang tidak memenuhi asumsi pengaruh aditif. Teknik ini mengubah data mentah ke skala logaritmik, yang dapat membantu mengurangi pengaruh nilai ekstrem, menormalkan distribusi data, dan mengatasi pola hubungan non-linear.

2.2. Penelitian Terdahulu

Tabel 2. 3 Penelitian Terdahulu

No.	Author	Judul	Metode	Hasil
1.	K. Vasanth; All.	Weighted Hybrid Machine Learning Algorithm to predict Rainfall	Weighted Hybrid Machine Learning	Hasil penelitian menghasilkan akurasi 75% dengan data curah hujan aktual

		for Hyderabad Region		real-time untuk bulan April dan Mei 2020.
2.	K. Likhith; All.	Improved Accuracy for Exploring Text - Based Emotion Recognition in Social Media Conversation Generalized Linear Model Compared with Decision Tree	Generalized linear model (GLM)	Metodologi Generalized Linear Model (GLM) mencapai akurasi 93,01% ketika dikorelasikan dengan algoritma Decision Tree.
3.	Soham Vyas; Sanskar Bhuwania; All.	The Impacts of Maintenance Weather and Aging on Solar Power Generation Forecasting and Prediction	Vector Autoregression (VAR)	Model VAR mampu memprediksi aktivitas pembangkit listrik tenaga surya dan pembangkitan listrik dengan RMSPE di bawah 10%, mendukung pengelolaan yang efisien.

4.	Mohammad Kaiser Azad; All.	Support vector regression based electricity peak load forecasting	Vector Autoregression (VAR)	Penelitian menggunakan data harian konsumsi beban puncak, suhu, dan kelembaban relatif selama 2014- 2016. Hasilnya menunjukkan efektivitas regresi vektor pendukung untuk prediksi beban puncak.
5.	Nur Dalilla Nordin; All.	Fast temperature extraction approach for BOTDA using Generalized Linear Model	Generalized Linear Model (GLM)	Metode ini lebih cepat karena menghilangkan langkah penentuan BFS, dengan akurasi distribusi suhu GLM setara metode konvensional.

Tabel ini menjelaskan penelitian terdahulu yang relevan dengan penelitian yang sedang dilakukan. Penelitian-penelitian dalam tabel tersebut mencakup berbagai metode seperti Weighted Hybrid Machine Learning, Generalized Linear

Model (GLM), dan Vector Autoregression (VAR), yang digunakan untuk memprediksi fenomena terkait data klimatologi, energi, dan sosial. Hasil penelitian terdahulu menunjukkan keberhasilan masing-masing metode dalam memodelkan dan memprediksi data spesifik, seperti :

Penelitian dari (Vasanth *et al.*, 2022)

Dalam studi mengenai prediksi curah hujan, penggunaan advanced regression memiliki sejumlah keuntungan dibandingkan dengan Weighted Hybrid Machine Learning Algorithm (WHMLA), khususnya dalam meningkatkan pemahaman terhadap faktor-faktor yang mempengaruhi curah hujan. Model regresi lanjutan mampu memetakan hubungan linier antara variabel iklim, seperti suhu, kelembaban, dan tekanan atmosfer, dengan curah hujan secara lebih jelas. Hal ini memungkinkan identifikasi variabel yang paling mempengaruhi prediksi curah hujan.

Sementara itu, WHMLA mengkombinasikan beberapa algoritma pembelajaran mesin yang lebih kompleks, membutuhkan lebih banyak data, waktu pemrosesan yang lebih lama, dan pengaturan parameter yang lebih sulit. Walaupun WHMLA dapat memberikan hasil lebih tepat dalam beberapa situasi, kompleksitasnya justru mengurangi kemampuan interpretasi dan bisa kurang ideal jika tujuan utamanya adalah untuk memahami faktor-faktor yang mempengaruhi curah hujan.

Advanced Regression lebih cocok digunakan untuk penelitian yang mengutamakan efisiensi dan pemahaman, terutama bila data yang tersedia terbatas

dan interpretasi model sangat penting. Model ini memungkinkan analisis mendalam mengenai pengaruh faktor-faktor iklim terhadap curah hujan tanpa menghadapi kompleksitas model pembelajaran mesin. Di Kediri Raya, penggunaan advanced regression bisa mempercepat adaptasi terhadap perubahan iklim, serta memberikan pemahaman lebih mendalam tentang dinamika cuaca dan curah hujan. Secara keseluruhan, advanced regression menawarkan keseimbangan antara akurasi dan kemudahan interpretasi, menjadikannya pilihan yang tepat untuk penelitian ini.

Penelitian dari (Likhith, Rekha and Kalaiyarasi, 2023)

Penelitian ini membahas perbandingan antara Generalized Linear Model (GLM) dan Decision Tree Algorithm dalam pengenalan emosi berbasis teks. Dalam penelitian ini, GLM menunjukkan akurasi yang lebih baik dibandingkan dengan algoritma pohon keputusan. GLM, sebagai model statistik yang dapat menangani data dengan berbagai distribusi, menunjukkan kemampuannya dalam meningkatkan akurasi prediksi, terutama dalam aplikasi pengenalan emosi dari teks yang tidak terstruktur.

Dalam konteks penelitian ini, Advanced Regression yang digunakan untuk memodelkan hubungan antara variabel iklim dan curah hujan mengadopsi prinsip yang serupa dengan GLM. Advanced Regression memungkinkan peneliti untuk memodelkan hubungan linier antara faktor-faktor iklim seperti suhu, kelembaban, dan tekanan atmosfer dengan curah hujan secara lebih transparan. Dengan demikian, seperti GLM yang meningkatkan akurasi dalam pengenalan emosi, Advanced Regression juga dapat menghasilkan prediksi yang lebih akurat dalam

konteks curah hujan, dengan memanfaatkan data iklim yang lebih jelas dan interpretatif.

Salah satu keunggulan dari Advanced Regression dibandingkan dengan metode yang lebih kompleks seperti machine learning adalah kemampuannya dalam menyederhanakan proses analisis tanpa mengorbankan akurasi. GLM yang digunakan dalam penelitian pengenalan emosi berbasis teks juga menunjukkan bagaimana teknik ini bisa diadaptasi untuk memodelkan data iklim, dengan keuntungan dalam hal kejelasan dan interpretasi hasil model. Hal ini sangat relevan dengan penelitian prediksi curah hujan, di mana pemahaman yang mendalam tentang faktor-faktor yang mempengaruhi curah hujan sangat diperlukan untuk merencanakan mitigasi bencana dan pengelolaan sumber daya alam.

Sebagaimana GLM memberikan hasil yang lebih baik dalam pengenalan emosi, Advanced Regression juga menawarkan metode yang efektif dan efisien dalam memodelkan prediksi curah hujan, memungkinkan untuk prediksi yang lebih tepat dan memberikan wawasan yang lebih jelas mengenai dinamika cuaca dan iklim di daerah tertentu, seperti Kediri Raya.

Penelitian dari (Vyas *et al.*, 2023)

Penelitian ini membahas penggunaan Vector Auto Regression (VAR) dalam prediksi energi surya, dengan fokus pada pemanfaatan data dinamis, termasuk pemeliharaan terjadwal dan tidak terjadwal serta kondisi cuaca yang berubah-ubah. VAR digunakan untuk memodelkan hubungan antar variabel yang mempengaruhi

energi surya, seperti suhu, kelembaban, dan kecepatan angin. Model ini sangat efektif dalam menghasilkan prediksi yang akurat, mengingat data yang digunakan bersifat kompleks dan dinamis. Pendekatan ini memberikan wawasan penting tentang bagaimana faktor-faktor eksternal dapat mempengaruhi produksi energi surya, dan hasilnya relevan dengan berbagai sektor yang bergantung pada data iklim untuk perencanaan dan pengambilan keputusan.

Metode VAR, yang digunakan dalam penelitian energi surya, memiliki kesamaan dengan Advanced Regression yang diterapkan dalam prediksi curah hujan. Kedua metode ini dapat memodelkan hubungan antara faktor-faktor lingkungan yang saling berkaitan, seperti suhu dan kelembaban, untuk memberikan hasil yang lebih transparan dan akurat. Dalam konteks curah hujan, Advanced Regression menggunakan data klimatologi untuk memprediksi pola curah hujan berdasarkan variabel-variabel iklim, memungkinkan peneliti untuk menangkap dinamika cuaca secara lebih jelas dan menyeluruh.

Salah satu keuntungan utama dari Advanced Regression dibandingkan dengan metode pembelajaran mesin yang lebih kompleks adalah kemampuannya untuk memberikan hasil yang lebih mudah diinterpretasikan, tanpa mengorbankan akurasi prediksi. Sama seperti VAR yang menangani data energi surya dengan cara yang sederhana namun efektif, Advanced Regression dapat digunakan untuk memprediksi curah hujan dengan lebih efisien, tanpa memerlukan data yang sangat besar atau model yang rumit. Dengan demikian, Advanced Regression menjadi

pilihan yang sangat cocok dalam konteks penelitian perubahan iklim dan prediksi cuaca ekstrem, seperti yang terjadi di Kediri Raya.

Penelitian dari (Azad, Uddin and Takruri, 2018)

Penelitian ini mengkaji penggunaan Support Vector Regression (SVR) dalam peramalan beban puncak listrik, menyoroti kemampuan model regresi untuk menganalisis hubungan antara data historis dan faktor lingkungan seperti suhu dan kelembaban. Dalam konteks penelitian ini, Advanced Regression dapat diterapkan untuk mengidentifikasi pola hubungan yang serupa antara faktor-faktor iklim seperti suhu, kelembaban, dan curah hujan. Model ini memiliki keunggulan dalam menangani data yang kompleks, memberikan hasil yang lebih akurat, dan memungkinkan prediksi yang konsisten berdasarkan data historis. Penelitian menggunakan dataset selama tiga tahun berturut-turut ini menekankan pentingnya data historis dalam menghasilkan model yang dapat diandalkan untuk prediksi.

Dalam hal ini, penelitian ini memanfaatkan dataset iklim untuk menggambarkan pola curah hujan di Kediri Raya dan menggunakan pendekatan Advanced Regression untuk memperoleh hasil yang lebih presisi dibandingkan dengan metode tradisional. Advanced Regression memungkinkan pengambilan keputusan berbasis data yang lebih tepat untuk mitigasi risiko terkait perubahan iklim. Model ini juga memberikan solusi yang lebih strategis dalam pengelolaan sumber daya alam, khususnya di wilayah dengan dinamika iklim seperti Kediri Raya. Dengan demikian, Advanced Regression berfungsi tidak hanya sebagai alat

analisis, tetapi juga sebagai landasan untuk pengambilan keputusan berbasis bukti yang mendukung kesejahteraan masyarakat.

Penelitian dari (Nordin *et al.*, 2020)

Penelitian ini mengungkapkan penggunaan Generalized Linear Model (GLM) dalam ekstraksi suhu pada sistem Brillouin Optical Time Domain Analyzer (BOTDA), yang menunjukkan bahwa GLM lebih efisien dan cepat dibandingkan dengan metode konvensional. Studi ini menekankan bahwa GLM dapat menyederhanakan proses yang rumit, seperti penghitungan Brillouin Frequency Shift (BFS), sambil tetap menghasilkan akurasi yang setara.

Pendekatan ini sangat relevan dengan Advanced Regression yang digunakan dalam penelitian ini, karena keduanya memanfaatkan metode regresi untuk menangani kompleksitas data dan efisiensi waktu dalam analisis. Advanced Regression juga dapat menyederhanakan analisis data iklim untuk memprediksi curah hujan dengan lebih akurat dan dapat diandalkan. Keunggulan Advanced Regression terletak pada kemampuannya memodelkan hubungan non-linear serta menangani variabilitas dalam dataset iklim, seperti suhu dan kelembaban, yang sangat penting dalam prediksi curah hujan. Selain itu, pendekatan ini memungkinkan pemrosesan data dalam skala besar dengan efisiensi tinggi, menjadikannya solusi ideal untuk penelitian berbasis data historis jangka panjang.

Dengan menggunakan Advanced Regression, model prediksi curah hujan menjadi lebih akurat, mengurangi waktu pemrosesan, dan meningkatkan

pengambilan keputusan yang berbasis data, seperti perencanaan pengelolaan sumber daya air dan mitigasi risiko banjir. Model ini tidak hanya memberikan hasil yang lebih baik, tetapi juga menjadi solusi strategis untuk menghadapi tantangan perubahan iklim di Kediri Raya.

