BAB I
PENDAHULUAN

1.1 Latar Belakang

Cuaca memiliki dampak signifikan pada kehidupan sehari-hari, memengaruhi sektor-sektor penting seperti pertanian dan transportasi. Kondisi ekstrem seperti hujan deras, angin kencang, atau panas berlebih dapat mengancam hasil pertanian dan meningkatkan risiko kecelakaan di jalan. Oleh karena itu, prediksi cuaca yang akurat sangat diperlukan untuk mengurangi risiko dan kerugian. Penerapan teknik komputasi canggih, khususnya *machine learning* (ML), telah meningkatkan akurasi peramalan cuaca. ML mampu menangani dataset besar dan kompleks dalam meteorologi, menawarkan cara untuk meningkatkan ketepatan prediksi serta mengurangi waktu peramalan. Kemampuan ML dalam mengenali pola dalam data historis dan real-time memperkuat model peramalan cuaca numerik atau *numerical weather prediction* (NWP) tradisional, memberikan solusi yang lebih efektif dan efisien (Iseh J & Y, 2013; Verma et al., 2024)

Kota Makassar, ibu kota provinsi Sulawesi Selatan, Indonesia, menjadi studi kasus yang relevan untuk penerapan teknik peramalan canggih semacam itu. Sebagai pusat perkotaan dan industri utama di Indonesia Timur, tatanan sosial-ekonominya sangat terkait erat dengan kondisi iklimnya (Rimba et al., 2023). Kota ini mengalami iklim hutan hujan tropis, yang ditandai dengan suhu tinggi yang konsisten, kelembapan, dan curah hujan tahunan yang substansial, dengan musim hujan dan kemarau yang berbeda (NearWeather.Com, n.d.). Profil iklim ini membuat Makassar rentan terhadap berbagai bahaya terkait cuaca, terutama banjir muson, yang dapat menyebabkan gangguan signifikan dan kerusakan ekonomi (Rimba et al., 2023).

Pentingnya strategis Makassar dalam mendukung pertumbuhan ekonomi Indonesia menggarisbawahi perlunya strategi manajemen bencana yang kuat, di mana peramalan cuaca yang akurat memainkan peran penting. Pembangunan sosial-ekonomi Makassar secara intrinsik terkait dengan ketahanannya terhadap gangguan terkait cuaca ini. Peramalan yang tidak akurat dapat memperburuk dampak bahaya, mengganggu produksi industri, rantai pasokan, dan kehidupan sehari-hari, sehingga menghambat stabilitas ekonomi (Rimba et al., 2023). Sebaliknya, peningkatan akurasi peramalan, seperti yang berpotensi ditawarkan oleh metode ML canggih seperti *Gradient Boosting Machines* (GBM), dapat mengarah pada kesiapsiagaan yang lebih baik, alokasi sumber daya yang lebih efektif untuk mitigasi dan respons, dan pada akhirnya, lingkungan perkotaan yang lebih tangguh yang mendukung aktivitas ekonomi berkelanjutan (Gede et al., 2025).

Sebelumnya telah dilakukan beberapa penelitian terkait dengan prediksi cuaca yaitu penelitian oleh (Luthfiarta et al., 2020) yang memahami dan memprediksi perubahan cuaca yang mungkin terjadi dengan menggunakan *supervised learning*, khususnya regresi linear berganda. Data yang digunakan berasal dari Stasiun Meteorologi Ahmad Yani Semarang tahun 2015-2017 yang dianalisis untuk mengukur hubungan antara variabel independen dan variabel dependen. Hasil penelitian menunjukkan bahwa nilai koefisien determinasi R^2 sebesar 25.5 persen, menandakan

i keempat variabel yang digunakan mampu menjelaskan sebesar
alam nilai curah hujan sebagai variabel dependen.
eh (Setiyaris et al., 2023) yang memberikan kontribusi yang
nperdalam pemahaman dan meningkatkan prediksi cuaca dengan
katan regresi linear berganda. Hasil penelitian yaitu variabel suhu,
n, dan kecepatan angin memiliki hubungan yang dapat diukur
ujan. Informasi ini menjadi dasar untuk peningkatan pemahaman
cuaca di masa depan, berkontribusi pada upaya mitigasi dan

Optimized using
trial version
www.balesio.com

ngan dampak perubahan cuaca. Pendekatan ini memanfaatkan data yang Badan Meteorologi untuk meningkatkan pemahaman terhadap fenomena nemberikan kontribusi positif pada pengembangan metode prediksi cuaca fektif.

iudian penelitian oleh (Insani et al., 2020) yang mengatasi perubahan cuaca yang cepat di Pekanbaru dengan menggunakan prediksi cuaca. Proses prediksi cuaca melibatkan beberapa kriteria, termasuk suhu udara, kelembaban, kecepatan angin, dan tekanan. Algoritma genetika digunakan sebagai alat untuk mencari nilai optimum pada masing-masing individu, yang kemudian digunakan sebagai batas fungsi keanggotaan dalam Fuzzy Tsukamoto. Hasil penelitian menunjukkan optimasi Fuzzy Tsukamoto dan Algoritma Genetika menghasilkan akurasi prediksi mencapai 72%, dengan probabilitas *crossover* sebesar 0,6 dan probabilitas mutasi sebesar 0,4. Hasil kombinasi Fuzzy Tsukamoto dan Algoritma Genetika terbukti mampu menghasilkan hasil akhir baik sehingga meningkatkan akurasi prediksi cuaca dan memungkinkan untuk mengantisipasi perubahan cuaca yang cepat di Pekanbaru.

Lalu penelitian oleh (Sunardi et al., 2020) yang merancang dan mengimplementasikan sistem prediksi curah hujan dengan menggunakan pengembangan Siklus Hidup Perangkat Lunak (SDLC) model *Waterfall*, serta menerapkan jaringan saraf tiruan (ANN) dengan algoritma *Backpropagation*. Hasil penelitian menunjukkan bahwa pengujian pada sistem yang dibangun untuk pelatihan dan validasi menghasilkan akurasi pelatihan sebesar 93.92% dengan validasi sebesar 73.04%, menunjukkan bahwa sistem ini dapat digunakan dan telah berjalan sesuai harapan.

Terakhir penelitian oleh (Dwiyanti & Prianto, 2023) yang fokus pada prediksi cuaca di Kota Jakarta menggunakan pendekatan *Random Forest* dan memanfaatkan data cuaca historis yang diperoleh dari OpenData Jakarta. Hasil evaluasi menunjukkan bahwa model *Random Forest* memberikan prediksi cuaca yang sangat baik, dengan tingkat akurasi, presisi, dan *recall* mencapai 0.71. F1-*score* mencapai 0.70, sementara ROC-AUC mencapai 0.92. Evaluasi ini mencerminkan kinerja model dalam mengklasifikasikan kondisi cuaca dengan mempertimbangkan keakuratan, ketepatan, serta keseimbangan antara presisi dan *recall*. Model ini berhasil mengidentifikasi berbagai kelas cuaca dengan akurat, memberikan informasi yang sangat berharga untuk mendukung pengambilan keputusan terkait kondisi cuaca di Kota Jakarta.

Algoritma Gradient Boosting Machine (GBM) muncul sebagai solusi yang menjanjikan, didukung oleh berbagai penelitian yang menunjukkan keunggulannya dalam prediksi cuaca (Gede et al., 2025). GBM adalah teknik pembelajaran mesin yang tangguh; ia bekerja dengan membangun model prediktif secara sekuensial dengan mengintegrasikan beberapa model prediktif lemah (umumnya pohon keputusan) menjadi satu model yang lebih kuat, di mana setiap model berikutnya berupaya memperbaiki kesalahan dari model sebelumnya. Proses ini melatih sistem untuk mengidentifikasi pola dalam data pelatihan dan memberikan respons yang tepat, baik terhadap pola yang dikenal maupun pola serupa lainnya. Peningkatan akurasi peramalan, sebagaimana yang berpotensi ditawarkan oleh metode ML canggih seperti GBM, dapat mengarah pada kesiapsiagaan yang lebih baik, alokasi sumber daya yang lebih efektif untuk mitigasi dan respons, dan pada akhirnya, terciptanya lingkungan perkotaan yang lebih tangguh yang mendukung aktivitas ekonomi berkelanjutan (Rimba et al., 2023). Berdasarkan urgensi ini, penelitian ini akan melakukan prediksi cuaca Kota Makassar menggunakan Algoritma *Gradient Boosting Machine***,** dengan harapan dapat berkontribusi dalam perencanaan dan penanggulangan dampak perubahan cuaca di masa depan bagi Kota Makassar.

Optimized using
Trial Version
www.balesio.com

ksi

dalam fisika melibatkan penggunaan berbagai metode dan model untuk i fenomena fisik di masa depan berdasarkan data dan prinsip ilmiah yang lah aspek penting dari berbagai bidang, termasuk meteorologi, klimatologi, hidrologi, dan ilmu lingkungan. Proses peramalan bergantung pada analisis data historis dan penerapan prinsip-prinsip fisik untuk membuat prediksi tentang kejadian atau kondisi di masa depan. Dalam konteks cuaca dan iklim, peramalan memainkan peran penting dalam memahami dan memprediksi fenomena atmosfer seperti suhu, curah hujan, dan pola badai (Barlow & Torres, 2021; Blanchard-Wrigglesworth et al., 2023; Brodie et al., 2021; Grooms et al., 2023; Huva et al., 2021). Selain itu, peramalan meluas ke sistem ekologi, yang digunakan untuk memprediksi distribusi spesies dan fenomena ekologi berdasarkan faktor klimatologi dan lingkungan (Johnson et al., 2023; Mu et al., 2023).

Model peramalan berbasis fisika, seperti model prediksi cuaca numerik (NWP), memanfaatkan hukum dan persamaan fisika untuk mensimulasikan dan memprediksi kondisi atmosfer. Model ini menggabungkan prinsip dinamika fluida, termodinamika, dan transfer radiasi untuk memperkirakan pola dan fenomena cuaca (Haupt et al., 2020; Huva et al., 2021). Selain itu, integrasi metode berbasis fisika dengan kecerdasan buatan dan teknik pembelajaran mesin telah menghasilkan kemajuan dalam peramalan probabilistik dan energi terbarukan, sehingga meningkatkan akurasi dan keandalan prediksi (Johnson & Wang, 2020; Kang et al., 2021).

Keakuratan dan keandalan perkiraan dipengaruhi oleh berbagai faktor, termasuk kualitas kondisi awal, model fisika, dan asimilasi data observasi (Haupt et al., 2020; Kaaya & Ascencio-Vásquez, 2022; Ye et al., 2023). Selain itu, pengembangan model peramalan berdasarkan informasi fisika, seperti yang digunakan untuk tenaga angin dan energi fotovoltaik, menekankan pentingnya mengintegrasikan prinsip-prinsip fisik dengan teknik komputasi canggih untuk meningkatkan akurasi peramalan (Xu et al., 2022).

Singkatnya, peramalan dalam fisika mencakup berbagai aplikasi, mulai dari prediksi cuaca dan iklim hingga peramalan ekologi dan energi terbarukan. Ini melibatkan integrasi prinsip-prinsip fisika, metode komputasi tingkat lanjut, dan data observasi untuk membuat prediksi tentang fenomena fisik di masa depan(Huva et al., 2021).

Maka dapat disimpulkan bahwa prediksi adalah teknik untuk meramalkan atau memperkirakan sesuatu yang belum terjadi. Proses ini melibatkan ramalan terstruktur berdasarkan informasi masa lampau dan saat ini dengan tujuan untuk mengurangi kesalahan. Prediksi dapat diartikan sebagai upaya memperkirakan keadaan masa depan dengan menggunakan informasi dari masa lampau, dan hal ini dilakukan untuk meminimalkan potensi kesalahan (Barlow & Torres, 2021).

1.3 Cuaca

Cuaca dalam fisika mengacu pada kondisi atmosfer dan fenomena yang terjadi di atmosfer bumi, termasuk suhu, kelembaban, angin, dan curah hujan. Unsur-unsur tersebut dipengaruhi oleh faktor fisik seperti intensitas cahaya, suhu, dan angin. Lebih lanjut, cuaca tidak hanya terbatas pada atmosfer bumi saja, namun juga mencakup cuaca antariksa, yang mencakup aktivitas matahari dan dampaknya terhadap ruang antarplanet dan magnetosfer bumi. Fenomena cuaca antariksa meliputi letusan matahari, lontaran massa koronal, dan partikel energik matahari, yang merupakan

Optimized using
Trial Version
www.balesio.com

ral dari proses fisik terkait cuaca antariksa. Selain itu, studi cuaca melibatkan model numerik, seperti model Weather Research and Forecasting (WRF), mulasikan dan memprediksi kondisi atmosfer.

1.3.1. Suhu

Suhu adalah parameter yang mengukur tingkat panas atau kehangatan udara di sekitar. Metodenya melibatkan penggunaan termometer, dan perubahan suhu sepanjang hari dapat direkam melalui perangkat seperti termograf atau termogram. Suhu tertinggi biasanya terjadi pada jam 13.00 atau 14.00, sementara suhu terendah dapat tercatat pada jam 04.00 atau 05.00.

1.3.2. Kelembapan

Kelembapan udara adalah ukuran sejauh mana udara di sekitar mengandung uap air. Hal ini bisa diukur sebagai persentase dari maksimum jumlah uap air yang udara dapat menampung pada suhu tertentu. Kelembapan udara merupakan faktor penting dalam menentukan kenyamanan cuaca dan memiliki dampak pada berbagai aspek, seperti kesehatan manusia, pertanian, dan kondisi lingkungan.

1.3.3. Curah Hujan

Hujan adalah tetesan air di udara yang jatuh ke permukaan bumi. Proses terjadinya hujan melibatkan pembentukan awan, pemadatan titik air di awan menjadi tetesan air, dan akhirnya jatuh ke bumi. Alat pengukur hujan disebut fluviog.

1.3.4. Angin

Perbedaan tekanan udara di berbagai lokasi menciptakan aliran udara atau angin, mengalir dari daerah bertekanan tinggi ke daerah bertekanan rendah (Sayeed et al., 2023).

1.3.5. Tujuan

Tujuan penelitian yaitu:

1. Menganalisis dan mengevaluasi performa algoritma Gradient Boosting Machine dalam memprediksi kondisi cuaca di Kota Makassar.
2. Mengidentifikasi variabel cuaca yang paling signifikan dalam mempengaruhi hasil prediksi menggunakan model GBM.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan beberapa manfaat. Bagi peneliti dan akademisi, penelitian ini tidak hanya memberikan sumbangan ilmiah di bidang ilmu komputer dan meteorologi, khususnya dalam penerapan algoritma GBM pada prakiraan cuaca, tetapi juga dapat menjadi acuan bagi penelitian sejenis di masa mendatang. Bagi lembaga terkait seperti BMKG dan pemerintah daerah, hasil penelitian ini diharapkan dapat digunakan untuk meningkatkan akurasi sistem prakiraan cuaca guna mendukung pengambilan keputusan untuk mengurangi risiko bencana akibat cuaca ekstrem di Kota Makassar. Manfaat lainnya ditujukan bagi masyarakat umum, yaitu masyarakat dapat memiliki informasi prakiraan cuaca yang lebih akurat untuk merencanakan kegiatan sehari-hari dengan lebih baik dan mengurangi dampak perubahan cuaca yang tiba-tiba.

Optimized using
Trial Version
www.balesio.com

penelitian ini juga bermanfaat bagi pengembang sistem teknologi cuaca, at memberikan wawasan tentang variabel meteorologi yang paling besar a terhadap hasil prakiraan, yang dapat digunakan untuk menyederhanakan timalkan model prakiraan cuaca di masa mendatang.

www.balesio.com
Trial Version
Optimized using

BAB II
METODE PENELITIAN

itma Gradient Boosting Machine

Gradient Bosting Machine (GBM) adalah metode pembelaran mesin berbasis penggabungan (*ensemble).* GBM bekerja dengan menggabungkan sejumlah model sederhana secara bertahap. Setiap model baru yang ditambahkan berfokus untuk memperbaiki galat prediksi dari model-model sebelumnya, sehingga hasil akhirnya menjadi lebih akurat. Model yang digunakan umumnya adalah pohon keputusan (*decision tree)* yang relatif sederhana atau lemah, namun ketika digabungkan secara iteratif, mampu menghasilkan model prediksi yang sangat kuat dan akurat. Secara umum GBM dapat direpresentasikan pada berikut :

$$F_m(x) = F_{m-1}(x) + v \cdot h_m(x) \quad (2.1)$$

dimana $F_m(x)$ dan $F_{m-1}(x)$ berturut turut adalah prediksi dari model *ensemble* pada iterasi ke-m dan prediksi iterasi sebelum m. Parameter $v > 0$ adalah laju pembelajaran (*learning rate*), yang umumnya berupa bilangan desimal kecil.Fungsi $h_m(x)$ merupakan model lemah (*weak learner*) yang baru ditambahkan pada iterasi ke-mmm. Model ini biasanya berupa pohon keputusan sederhana yang dilatih untuk memperbaiki kesalahan prediksi dari model sebelumnya. Untuk melakukan hal ini $h_m(x)$, dilatih menggunakan gradien negatif dari fungsi kerugian (*loss function*), yang secara intuitif menunjukkan arah terbaik untuk mengurangi kesalahan model. Artinya, pada setiap iterasi algoritma GBM menghitung seberapa besar dan ke arah mana model saat ini salah, lalu membuat model kecil baru yang mencoba mengoreksi kesalahan tersebut. Proses ini dilakukan secara berulang, sehingga model akhir merupakan kombinasi dari banyak model kecil yang bersama-sama memberikan prediksi yang lebih akurat (Natekin & Knoll, 2013).

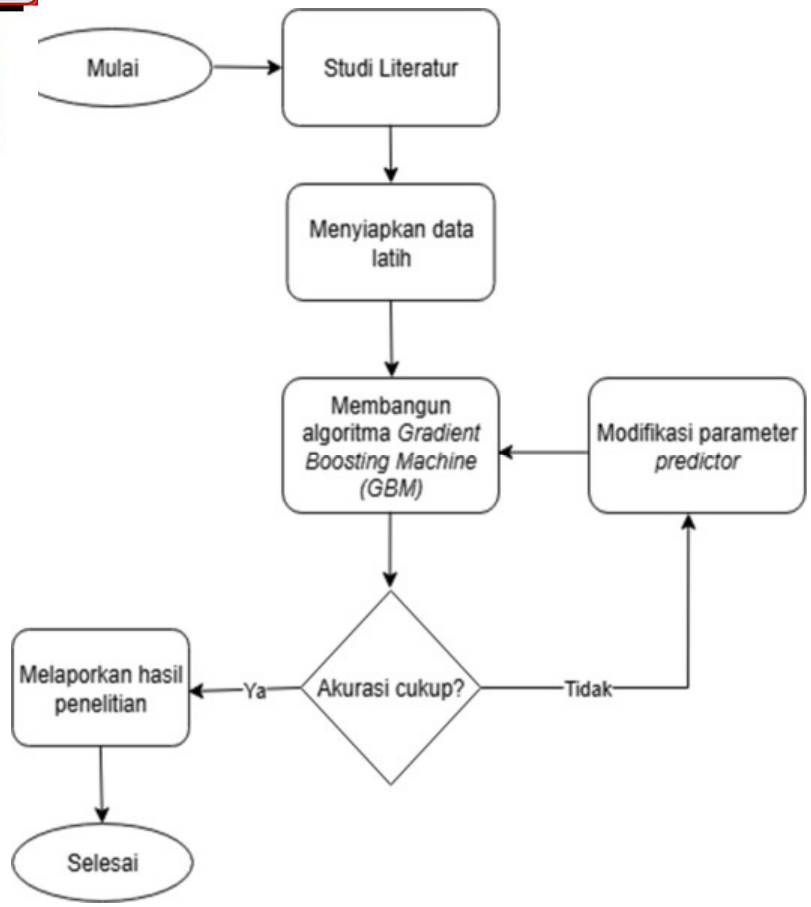
2.2 *Diagram Alir Penelitian*

Penelitian ini diawali dengan melakukan studi literatur terkait prediksi cuaca dan penggunaan algoritma *Gradient Boosting Machine (*GBM) dalam konteks data meteorologi. Tujuan dari studi ini adalah untuk memahami pendekatan-pendekatan yang telah digunakan sebelumnya serta menentukan posisi dan kontribusi dari penelitian ini. Model pembelajaran mesin dengan algoritma GBM ini dibuat dengan menggunakan bahasa pemrograman python. Adapun tahapan-tahapan dalam penelitian ini meliputi:

1. Membangun model pembelajaran mesin dengan algoritma *Gradient Boosting Machine*
2. Menguji kode program pada data BMKG yang telah diperoleh dan menyempurnakan hasilnya sehingga diperoleh hasil akurasi yang cukup
3. Menganalisis data prediksi dan aktual yang diperoleh dan membuat metrik evaluasinya
4. Terakhir, melaporkan hasil penelitian berdasarkan *output* model yang telah dibuat.

Tahapan-tahapan inti penelitian ini secara lengkap diberikan dalam diagram alir yang ditunjukkan pada Gambar 2.1 berikut:

www.balesio.com
Optimized using
Trial Version



Gambar 1. Diagram Alir Penelitian

Seluruh tahapan tersebut akan dilaksanakan di Laboratorium Fisika Teori dan Komputasi, Departemen Fisika, Universitas Hasanuddin (UNHAS). Implementasi dan proses komputasi intensif dilakukan menggunakan perangkat PC pribadi dengan spesifikasi sebagai berikut: prosesor Intel® Core™ i5-12400F, memori 16 GB DDR4, penyimpanan SSD 1 TB, dan kartu grafis NVIDIA RTX 3060 dengan VRAM 8 GB. Perangkat ini digunakan untuk mempercepat proses pelatihan model Gradient Boosting Machine (GBM) pada dataset berukuran besar.

2.3 *Data Cuaca Harian*

Data dalam penelitian ini mencakup data historis cuaca Kota Makassar yang terdiri atas suhu udara, tekanan udara, kelembapan udara, dan data lain yang relevan untuk membangun model prediksi cuaca. Data ini diambil dari pangkalan data BMKG untuk data harian parameter cuaca di Kota Makassar melalui stasiun Paotere (*Data Online - Direktorat Data Dan Komputasi BMKG*, n.d.).

Data ini terdiri dari beberapa parameter. Parameter-parameter ini memberikan gambaran komprehensif tentang kondisi atmosfer pada setiap harinya. Penjelasan tentang parameter diberikan pada Tabel 2.1.

Optimized using
trial version
www.balesio.com

skripsi parameter data BMKG

	Penjelasan Singkat	Satuan
	Tanggal pengamatan data cuaca	-
	Suhu Minimum	◦C
Tx	Suhu Maksimum	◦C
Tavg	Suhu Rata-rata	◦C
RH_avg	Kelembapan Relatif Rata-rata	%
RR	Curah Hujan	mm
ss	Penyinaran Matahari	Jam
FF_x	Kecepatan Angin Maksimum	m/s
DDD_x	Arah Angin saat Kecepatan Maksimum	Derajat (0-360)
FF_avg	Kecepatan Angin Rata-rata	m/s
DDD_car	Arah Angin Rata-rata (Cardinal Direction)	(N, NE, E, dll.)

Untuk menunjukkan data tidak terukur diberikan nilai 8888, dan 9999 jika tidak ada data pengukuran sama sekali (tidak dilakukan pengukuran).

2.4 *Pra Pemrosesan Data*

Pra-pemrosesan data merupakan tahap penting dalam membangun model prediksi yang andal. Data cuaca historis sering kali mengandung nilai hilang, anomali, atau format yang tidak konsisten. Oleh karena itu, dilakukan beberapa tahapan preprocessing untuk membersihkan data, mengembangkan fitur baru, dan menyiapkan data agar sesuai untuk pelatihan model.

2.4.1 Pembersihan Data (Data Cleaning)

Pada tahap ini, penanganan nilai hilang dan tidak valid menjadi fokus utama. Dalam dataset BMKG, nilai khusus seperti 8888 atau 9999 digunakan untuk menandai data yang hilang atau error. Nilai-nilai ini tidak dapat langsung digunakan oleh model dan harus digantikan melalui strategi interpolasi. Untuk memastikan bahwa data bersih tanpa kehilangan karakteristik temporal alaminya, dilakukan beberapa tahap interpolasi sebagai berikut :

1. *Forward Fill* – Mengisi nilai kosong dengan nilai valid terakhir sebelumnya.
2. *Backward Fill* – Mengisi nilai kosong dengan nilai valid berikutnya.
3. Rata-Rata Bergerak 5 Hari (*5-Day Rolling Mean*) – Digunakan untuk menghaluskan data dan menggantikan nilai kosong berdasarkan tren lokal.
4. *Median Imputation* – Digunakan sebagai opsi terakhir jika nilai masih hilang, dengan mengganti menggunakan nilai median dari kolom terkait.

2.4.2 Rekayasa Fitur (Feature Engineering)

Feature engineering merupakan proses transformasi dan penciptaan fitur-fitur baru dari data mentah untuk meningkatkan kemampuan model dalam mengenali pola. Pada

Optimized using
trial version
www.balesio.com

ni, berbagai kategori fitur dikembangkan secara sistematis untuk tasikan dinamika cuaca yang kompleks, baik secara temporal, fisis, maupun

r Temporal

Cuaca memiliki sifat musiman yang berulang, terutama dalam konteks tahunan. Untuk menangkap siklus waktu tersebut, digunakan *encoding* siklis terhadap informasi kalender seperti bulan dan hari ke-berapa dalam setahun. Transformasi siklis ini dilakukan dengan fungsi sinus dan cosinus agar model dapat memahami hubungan siklikal antara waktu dan cuaca.

$$Month_{sin} = sin\ sin\left(\frac{2\pi \cdot Month}{12}\right), \qquad Month_{cos} = cos\ cos\left(\frac{2\pi \cdot Month}{12}\right) \tag{2.2}$$

$$DayOfYear_{sin} = sin\ sin\left(\frac{2\pi \cdot DayOfYear}{365.25}\right), \quad DayOfYear_{cos} = cos\ cos\left(\frac{2\pi \cdot DayOfYear}{365.25}\right) \tag{2.3}$$

Dimana $DayOfYear$ adalah bilangan urutan hari ke-berapa dalam setahun untuk suatu tanggal tertentu, dimulai dari 1 Januari sebagai hari ke-1, hingga 31 desember sebagai hari ke 365 (atau ke-366 jika tahun kabisat).

2.4.2.2 Turunan Meteorologis

Beberapa parameter baru diturunkan dari parameter cuaca yang tersedia untuk merepresentasikan kondisi atmosfer yang lebih kompleks. Fitur turunan yang dimaksud adalah sebagai berikut :

1. Rentang Suhu

$$TempRange = T_{max} - T_{min} \tag{2.4}$$

2. Titik Embun

$$DewPoint = T_{avg} - \frac{100 - RH_{avg}}{5} \tag{2.5}$$

3. Indeks Panas

$$HeatIndex = T_{avg} + 0.05 \cdot RH_{avg} \tag{2.6}$$

Fitur-fitur diatas membantu memperkaya informasi termal dan kelembapan dalam dataset yang penting untuk menetukan prediksi presipitasi, suhu, dan kecepatan angin.

2.4.2.3 Fitur Interaksi Nonlinier

Untuk menangkap hubungan kompleks antar variabel meteorologi, dibuat fitur interaksi:

1. Interaksi suhu dan kelembapan:

$$Temp_Humidity = \left(\frac{T_{max} + T_{min}}{2}\right) \cdot RH_{avg} \tag{2.7}$$

2. Interaksi rentang suhu dan kelembapan:

$$Temp_Range_RH = Temp_Range \cdot RH_{avg} \tag{2.8}$$

Optimized using
Trial Version
www.balesio.com

memperkaya informasi termal dan kelembapan dalam dataset yang penting ntukan prediksi presipitasi, suhu, dan kecepatan angin.

r Historis : Lag dan Rolling

merupakan data yang bersifat deret waktu (*time series)*, sehingga penting untuk mengakomodasi ketergantungan nilai saat ini terhadap nilai historisnya:

1. Fitur Lag:

$$Lag_k(x_t) = x_{t-k}, untuk\ k \in \{1,2,3,5,7,14\} \quad (2.9)$$

2. Rolling Mean dan Rolling Standar Deviation

$$RollingMean_w(x_t) = \frac{1}{w}\sum_{i=0}^{w-1} \left(x_{t-i} - \underline{x}\right)^2 \quad (2.10)$$

$$RollingStd_w(x_t) = \sqrt{\frac{1}{w}\sum_{i=0}^{w-1} \left(x_{t-i} - \underline{x}\right)^2} \quad (2.11)$$

Nilai rata-rata dan standar deviasi harian pada jendela waktu tertentu (3,7, dan 14 hari) digunakan untuk mencerimkan tren lokal bergerak dan fluktuasi data pada rentang tersebut.

2.4.2.5 Fitur Pola (Pattern Features)

Pola berturut-turut dalam peristiwa hujan atau kering dapat menjadi sinyal penting bagi model:

1. *Rain Streak* (Hujan berturut-turut):

$$Rain_Streak_t = max\ \{k | x_{t-k+1}, x_{t-k+2}, \dots, x_t > 0\} \quad (2.12)$$

2. *Dry Streak* (Cerah berturut-turut):

$$Dry_Streak_t = max\ \{k | x_{t-k+1}, x_{t-k+2}, \dots, x_t > 0\} \quad (2.13)$$

2.4.2.6 Konversi arah mata angin

Data arah mata angin seringkali dituliskan dalam bentuk *string*, misalnya N untuk *North* atau Utara, S untuk *South* atau selatan, W untuk *West*, dll. Sehingga diperlukan proses konversi atau *mapping* data ini ke tipe data numerik sehingga dapat diolah dalam pembelajaran mesin. Tabel 1 berisi *mapping* arah mata angin untuk 16 arah mata angin:

Tabel 2. Konversi arah mata angin

Arah	Derajat (°)	Arah	Derajat (°)
N	0.0	S	180.0
NNE	22.5	SSW	202.5
NE	45.0	SW	225.0
ENE	67.5	WSW	247.5
E	90.0	W	270.0

Optimized using
Trial Version
www.balesio.com

Derajat (°)	Arah	Derajat (°)
112.5	WNW	292.5
135.0	NW	315.0
157.5	NNW	337.5

2.4.3 Normalisasi Data

Dalam pembelajaran mesin, normalisasi data merupakan tahap penting, terutama ketika fitur memiliki skala nilai yang berbeda secara signifikan. Pada penelitian ini digunakan berbagai macam parameter dengan rentang nilai yang berbeda beda, sehingga normalisasi data perlu dilakukan untuk mempercepat proses pelatihan model pembelajaran mesin algoritma GBM.

2.4.3.1. Normalisasi Fitur

Semua variabel prediktor (variabel independen) dinormalisasi menggunakan teknik standardisasi, yaitu transformasi data ke dalam distribusi normal standar dengan rata-rata 0 dan standar deviasi 1. Proses ini dilakukan dengan menggunakan rumus:

$$x' = \frac{x - \mu}{\sigma} \tag{2.14}$$

Dimana x adalah nilai asli dari fitur atau parameter, μ adalah rata rata dari fitur, dan σ adalah standar deviasi dari fitur tersebut.

Proses normalisasi data dilakukan menggunakan algoritma StandarScaler dari pustaka Scikit-learn. Normalisasi hanya dilakukan berdasarkan data latih (*training set*), sedangkan data uji (*test set*) ditransformasi menggunakan parameter (μ, σ) dari data latih untuk menghindari kebocoran data (*data leakage)*.

2.4.3.2. Normalisasi target

Transformasi target dilakukan dengan pendekatan yang sama seperti pada fitur yaitu:

$$y' = \frac{y - \mu}{\sigma} \tag{2.15}$$

dimana y adalah nilai asli dari target, μ adalah rata rata dari target, dan σ adalah standar deviasi dari target tersebut. Normalisasi dilakukan hanya pada saat pelatihan model. Setelah dilakukan prediksi dari model yang dibuat, hasil yang diperoleh kemudian ditransformasi kembali ke skala aslinya menggunakan proses *inverse transform*:

$$y = y' \cdot \sigma + \mu \tag{2.16}$$

2.4.4 Seleksi Fitur

2.4.4.1 Seleksi berdasarkan korelasi Pearson

Memilih fitur yang penting dan berpengaruh pada model tidak bisa dilakukan dengan sebarang cara. Salah satu cara untuk seleksi fitur adalah dengan menggunakan korelasi Pearson. Setiap fitur akan dipasangkan dengan variabel target. Fitur dengan korelasi yang sangat rendah yaitu dengan nilai $|r| < 0.05$ akan dieliminasi dari fitur awal. Nilai r dapat dihitung dengan rumus (Blanchard-Wrigglesworth et al., 2023):

$$r_{x,y} = \frac{\sum(x_i - \underline{x}) \cdot \left(y_i - \underline{y}\right)}{\sqrt{\sum(x_i - \underline{x})^2 \cdot \sum\left(y_i - \underline{y}\right)^2}} \tag{2.17}$$