

PREDIKSI CURAH HUJAN KOTA BOGOR MENGGUNAKAN ALGORITMA RANDOM FOREST

Maesha Ayu Syaharani¹, Budi Arif Dermawan², Riza Ibnu Adam³

^{1,2,3}Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

maesyaayu62@gmail.com, budi.arif@staff.unsika.ac.id,

riza.adam@staff.unsika.ac.id

^{1,2,3}Jalan H.S Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Indonesia

Keywords:

*Rainfall, Prediction,
Random Forest, Grid
Search.*

Abstract

Bogor City is one of the regions in Indonesia with high rainfall intensity throughout the year. The high rainfall is one of the factors causing disasters such as floods and landslides. The development of a rainfall prediction model is an important need to support disaster risk planning and mitigation by the Bogor City local government. This study aims to build a daily rainfall prediction model. The research design uses the Knowledge Discovery in Database (KDD) methodology with the Random Forest algorithm as the main method in the analysis. The Random Forest model is applied with Quantile Transform adjustments and Grid Search to obtain the best parameter values. The tuning results show that the model performance is optimal with bootstrap: True, n_estimators: 500, max_depth: 5, and max_features: sqrt. The model was tested using K-Fold Cross Validation at several values of K = 3, 5, 7, 10, and 15 which were measured based on the difference between the predicted and actual values. The value of K = 15 shows the lowest error rate with an RMSE of 21,92 mm and an MAE of 15,21 mm.

Kata Kunci:

*Curah Hujan,
Prediksi, Random
Forest, Grid Search.*

Abstrak

Kota Bogor merupakan salah satu wilayah di Indonesia dengan intensitas curah hujan yang tinggi sepanjang tahun. Tingginya curah hujan tersebut menjadi salah satu faktor penyebab terjadinya bencana seperti banjir dan tanah longsor. Pengembangan model prediksi curah hujan menjadi kebutuhan penting untuk mendukung perencanaan dan mitigasi risiko bencana oleh pemerintah daerah Kota Bogor. Penelitian ini bertujuan untuk membangun model prediksi curah hujan harian. Rancangan penelitian menggunakan metodologi *Knowledge Discovery in Database* (KDD) dengan algoritma *Random Forest* sebagai metode utama dalam analisis. Model *Random Forest* diterapkan dengan penyesuaian *Quantile Transform* dan *Grid Search* untuk memperoleh nilai parameter terbaik. Hasil penyesuaian menunjukkan bahwa kinerja model optimal dengan *bootstrap: True*, *n_estimators: 500*, *max_depth: 5*, dan *max_features: sqrt*. Model diuji menggunakan *K-Fold Cross Validation* pada beberapa nilai K = 3, 5, 7, 10, dan 15 yang diukur berdasarkan selisih antara nilai prediksi dan aktual. Nilai K = 15 menunjukkan tingkat kesalahan terendah dengan RMSE sebesar 21,92 mm dan MAE sebesar 15,21 mm.

1. Pendahuluan

Indonesia berada pada garis khatulistiwa dan termasuk dalam wilayah beriklim tropis, yang membentang dari 23,5 derajat lintang utara hingga 23,5 derajat lintang selatan. Indonesia memiliki suhu yang relatif stabil namun curah hujannya bervariasi dan adanya pengaruh perubahan iklim yang dapat mengubah pola hujan di berbagai wilayah [1]. Salah satu contohnya adalah Kota Bogor di Provinsi Jawa Barat dengan luas daerah 111,39 km². Kota Bogor dikenal sebagai “Kota Hujan” karena intensitas curah hujannya yang tinggi. Dalam beberapa tahun terakhir, curah hujan di Kota Bogor tergolong tinggi setiap tahunnya, dengan jumlah sekitar 3.500 hingga lebih dari 4.000 mm/tahun.

Tingginya intensitas curah hujan menjadi faktor utama penyebab bencana alam di Kota Bogor, seperti tanah longsor dan banjir [2]. Dalam 4 tahun terakhir, tanah longsor mengakibatkan 10 korban jiwa, 1.385 orang menderita, dan 170 rumah rusak, sedangkan banjir menyebabkan 7 korban jiwa, 1.659 orang menderita, dan 635 rumah terendam [3]. Agar masyarakat lebih siaga dan mampu mengurangi resiko terhadap bencana, diperlukan prediksi curah hujan yang akurat. Prediksi ini akan membantu masyarakat dan lembaga terkait sebagai upaya mitigasi bencana alam di Kota Bogor kedepannya.

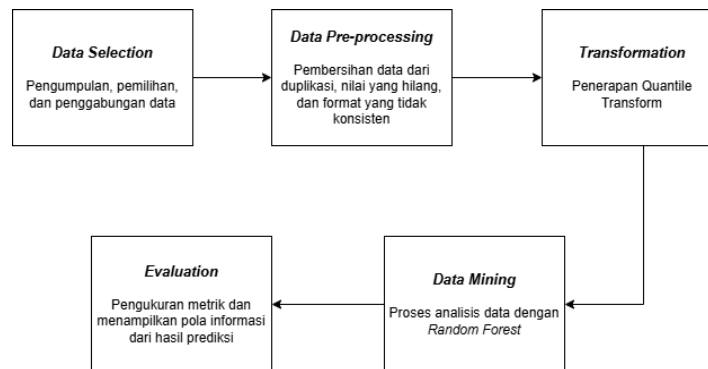
Berbagai penelitian terkait prediksi aspek lingkungan telah dilakukan sebelumnya. Penelitian [4] memprediksi kebakaran hutan di Serbia yang menunjukkan kinerja *Random Forest* lebih unggul dibandingkan *Logistic Regression* (LR). LR kurang optimal dalam memprediksi kebakaran yang berisiko tinggi karena keterbatasannya dalam menangani pola variabel yang kompleks. Penelitian lainnya [5] membandingkan *Random Forest* dan *Artificial Neural Network* (ANN) dalam memprediksi kualitas air. *Random Forest* menunjukkan kinerja yang lebih unggul dengan RMSE sebesar 27,99, sedangkan ANN menunjukkan performa yang kurang optimal yang disebabkan oleh sensitivitas terhadap pemilihan hyperparameter.

Penelitian [6] membandingkan *M5 Pruned* (M5P), *Random Forest*, *Random Tree*, dan *Reduced Error Pruning Tree* (REPT) dalam memprediksi ketinggian air Danau Zrebar di Iran. M5P menunjukkan performa yang lebih unggul, sedangkan REPT memiliki kinerja terendah karena keterbatasan model dalam mengatasi *outlier*. Kemudian, Penelitian [7] memprediksi kualitas udara di Yordania menggunakan algoritma *Random Forest*. Hasil penelitian menunjukkan *Random Forest* mampu memperkirakan nilai karbon monoksida (CO) secara akurat dengan error yang rendah meskipun memiliki jumlah fitur yang terbatas.

Berdasarkan uraian permasalahan dan penelitian sebelumnya, *Random Forest* akan digunakan dalam analisis prediksi curah hujan ini karena dinilai efektif dalam melakukan prediksi. Penerapan *Random Forest* disesuaikan dengan *Quantile Transform* untuk mengurangi pengaruh *outlier* serta *Grid Search* untuk mengatasi parameter yang kurang sesuai. Hasil prediksi ini diharapkan dapat digunakan untuk memahami pola curah hujan, memberikan informasi untuk perencanaan kebijakan, dan mitigasi bencana alam.

2. Metodologi Penelitian

Proses analisis dalam memprediksi curah hujan ini diterapkan menggunakan metodologi *Knowledge Discovery in Databases* (KDD). Metodologi *Knowledge Discovery in Databases* ini meliputi 5 tahapan yang ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

1. *Data Selection*

Tahap ini mencakup pemilihan, pengumpulan, dan penggabungan data. Data yang digunakan merupakan data cuaca Kota Bogor dari Stasiun Klimatologi Jawa Barat melalui situs resmi BMKG (dataonline.bmkg.id) periode 2019–2024. Data dikumpulkan secara bulanan dan digabungkan hingga menjadi dataset lengkap yang berjumlah 2.192 baris.

2. *Data pre-processing*

Tahap ini data akan dibersihkan dari adanya duplikat, *missing value*, kesalahan nilai, dan penyesuaian format data. Pendekatan yang digunakan untuk mengatasi *missing value* adalah melakukan imputasi menggunakan nilai rata-rata (*mean imputation*). Selain itu, kesalahan nilai juga dibersihkan agar sesuai dengan kebutuhan proses analisis pada tahap berikutnya.

3. *Transformation*

Transformasi data diartikan sebagai proses mengubah bentuk atau format data yang kompleks menjadi lebih sederhana dan mudah diolah untuk analisis. Keberadaan *outlier* dapat memengaruhi kinerja model sehingga diperlukan teknik *feature scaling* agar memiliki skala yang sama dan lebih stabil. Oleh karena itu, akan diterapkan *Quantile Transform* pada fitur yang memiliki *outlier*.

4. *Data Mining*

Tahap ini dilakukan analisis data menggunakan algoritma *Random Forest regression*. Penerapan model akan disesuaikan dengan *hyperparameter* yang optimal. Penentuan kombinasi *hyperparameter* terbaik diterapkan menggunakan *Grid Search*. Model terbaik yang dihasilkan akan dievaluasi menggunakan *K-Fold Cross Validation* dengan menguji pada beberapa nilai K untuk menilai kinerja model pada berbagai pembagian data.

5. *Evaluation*

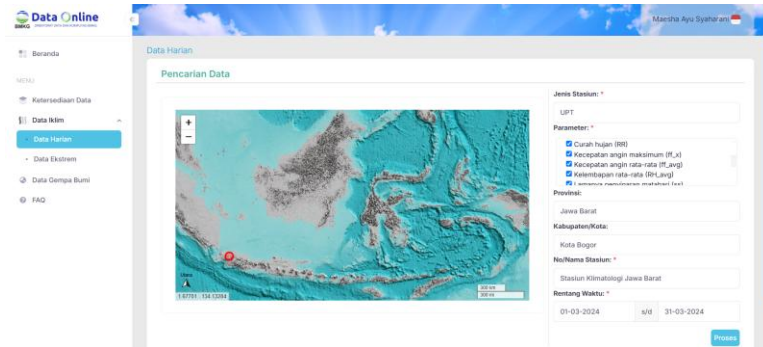
Tahap ini dilakukan pengkajian ulang terhadap model yang telah dikembangkan. Evaluasi pada model ini dilakukan dengan mengukur metrik seperti RMSE (*Root Mean Squared Error*) dan MAE (*Mean Absolute Error*) pada berbagai nilai K, menampilkan nilai aktual dan nilai estimasi, mengukur nilai prediktor serta nilai prediksi selama 1 tahun kedepan.

3. Hasil dan Pembahasan

Pada bagian ini akan menjelaskan hasil dari penelitian sesuai dengan rancangan yang telah dipaparkan sebelumnya.

3.1 Data Selection

Langkah pertama dilakukan dengan memilih stasiun cuaca dan variabel yang sesuai dengan keperluan analisis pada situs resmi BMKG, dataonline.bmkg.id. Tampilan proses pengumpulan data dapat dilihat pada Gambar 2.



Gambar 2. Tampilan saat mengumpulkan data

Atribut yang diperoleh meliputi temperatur (minimum, maksimum, dan rata-rata), kelembapan relatif, curah hujan, lamanya penyinaran, kecepatan angin (maksimum dan rata-rata), serta arah angin. Kemudian, data yang diperoleh disusun menjadi satu dataset lengkap dengan jumlah 2.192 baris.

3.2 Data Pre-processing

Data yang dikumpulkan masih belum ideal untuk dianalisis karena terdapat beberapa ketidaksesuaian yang dapat memengaruhi kinerja model. Tahap ini dilakukan penanganan sesuai dengan kebutuhan analisis dan karakteristik model. Dataset akan dilakukan penyesuaian format, identifikasi nilai yang tidak konsisten, duplikat, *missing value*, serta seleksi fitur agar mendukung kinerja model secara optimal.

1. Penyesuaian format data

Bagian ini diawali dengan mengubah nama atribut. Nama atribut dalam dataset awal berupa singkatan atau kode tertentu sehingga diperlukan mengubah nama atribut agar lebih mudah dipahami dan konsisten dengan konteks data.

Tabel 1. Penyesuaian nama atribut

Sebelum	Sesudah
Tn	suhu_min
Tx	suhu_max
Tavg	suhu_avg
RH_avg	kelembapan_avg
RR	curah_hujan
ss	sinar_mthr
ff_x	angin_max

ddd_x	arah_angin
ff_avg	angin_avg

Setelah menyesuaikan nama atribut, dilakukan pengecekan tipe data atribut untuk memastikan bahwa setiap kolom dapat digunakan untuk analisis. Hasil pengecekan ditemukan atribut curah hujan memiliki tipe data *object* yang ditunjukkan pada Gambar 3. Ketidaksesuaian ini disebabkan oleh keberadaan karakter non-numerik di dalam kolom tersebut.

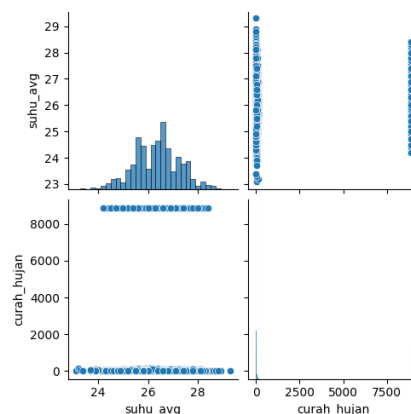
Tanggal	object
suhu_min	float64
suhu_max	float64
suhu_avg	float64
kelembapan_avg	float64
curah_hujan	object
sinar_mthr	float64
angin_max	float64
arah_angin	float64
angin_avg	float64

Gambar 3. Informasi tipe data atribut

Atribut curah hujan seharusnya memuat nilai bertipe numerik secara keseluruhan karena digunakan sebagai variabel target pada analisis. Oleh karena itu, karakter non-numerik dibersihkan dengan menghapus dan menggantinya menjadi nilai kosong.

2. Identifikasi ketidaksesuaian nilai

Tahap ini dilakukan dengan mengidentifikasi sebaran data atribut pada dataset. Gambar 4 dibawah menunjukkan grafik *scatterplot matrix* dalam sebaran nilai atribut suhu rata-rata dengan curah hujan.



Gambar 4. Sebaran suhu rata-rata dengan curah hujan

Sebaran data pada sebagian besar atribut dalam dataset terlihat stabil namun terdapat penyimpangan pada nilai curah hujan. Gambar 4 menunjukkan sebaran nilai curah hujan cenderung menyimpang ke sisi atas, bawah, dan samping. Hal ini disebabkan karena

adanya nilai 8888 yang menandakan data tidak terukur (BMKG, 2023) sehingga nilai curah hujan lebih terlihat ke sisi yang lebih tinggi. Oleh karena itu, dilakukan pembersihan dengan menghapus nilai 8888 dan menggantinya menjadi nilai kosong (*null*).

3. Identifikasi duplikat

Duplikat diartikan sebagai data yang muncul lebih dari satu kali dengan nilai identik di seluruh kolom. Proses identifikasi data duplikat pada dataset ditunjukkan pada Gambar 5.

```
#mengecek duplikat
total_duplikat = df.duplicated().sum()
print(f"Jumlah duplikat pada data: {total_duplikat}")
```

Jumlah duplikat pada data: 0

Gambar 5. Identifikasi duplikat

Hasil pengecekan duplikat pada Gambar 5, menunjukkan tidak adanya baris yang duplikat atau diartikan dengan 2.192 baris unik sesuai dengan jumlah data.

4. Missing value

Adanya *missing value* atau nilai hilang ini dapat memengaruhi kinerja model terutama jika dalam jumlah yang besar. Untuk itu, dilakukan pengecekan pada setiap atribut dengan jumlah nilai hilang ditunjukkan pada Tabel 2.

Tabel 2. Jumlah *missing value* pada atribut

Atribut	Jumlah nilai hilang
suhu_min	193
suhu_max	81
suhu_avg	45
kelembapan_avg	47
curah_hujan	710
sinar_mthr	57
angin_max	6
arah_angin	7
angin_avg	6

Tabel 2 menunjukkan setiap atribut memiliki nilai yang hilang. Atribut curah hujan sebagai target memiliki jumlah nilai hilang terbanyak. Penanganan nilai hilang pada target ini dilakukan dengan menghapus baris-baris tersebut agar mencerminkan kondisi sebenarnya untuk menjaga keakuratan model. Sementara, nilai hilang pada fitur diatasi dengan mengganti nilai tersebut menggunakan rata-rata dari masing-masing atribut atau disebut *mean imputation*.

5. Seleksi fitur

Pemilihan fitur ini dilakukan dengan mengukur korelasi antara masing-masing fitur terhadap target menggunakan fungsi korelasi yang tersedia pada *library pandas*. Nilai korelasi antara fitur terhadap target ditunjukkan pada Tabel 3.

Tabel 3. Nilai korelasi fitur terhadap target

Fitur	Nilai korelasi
kelembapan_avg	0.250481
angin_max	0.039211
sinar_mthr	0.016823
arah_angin	0.016023
suhu_min	-0.003715
suhu_max	-0.033312
angin_avg	-0.038898
suhu_avg	-0.117251

Sebagian besar fitur menunjukkan korelasi yang relatif rendah terhadap target. Fitur diseleksi berdasarkan 5 nilai korelasi tertinggi yang meliputi kelembapan rata-rata, angin maksimum, sinar matahari, arah angin, dan suhu minimum. Berdasarkan tahapan *pre-processing* yang telah dilakukan, dataset dapat digunakan untuk analisis pada tahap selanjutnya dengan jumlah 1.482 baris dan 5 fitur yang telah diseleksi.

3.3 Transformation

Transformasi data bertujuan untuk meningkatkan efektivitas model sehingga dapat meningkatkan kinerja prediksi. Tahap transformation ini diterapkan teknik *feature scaling* menggunakan *Quantile Transform* untuk mengubah skala data menjadi lebih stabil dan tidak terpengaruh oleh nilai ekstrem. Metode ini diterapkan pada fitur yang mengandung *outlier*. Hasil identifikasi menunjukkan fitur yang mengandung *outlier* meliputi suhu minimum, kelembapan rata-rata, dan angin maksimum. Oleh karena itu, diterapkan *Quantile Transform* pada fitur yang disebutkan. Hasil penerapan ini ditunjukkan pada Gambar 6.

```

suhu_min  suhu_min_qt  kelembapan_avg  kelembapan_avg_qt  angin_max  \
0      22.5    -0.016310         89.0         0.895780        10.0
1      23.4     0.941700         84.0        -0.186756         5.0
2      22.9     0.414267         76.0        -1.453280         3.0
7      22.5    -0.016310         84.0        -0.186756         2.0
8      23.1     0.686350         86.0         0.253606         4.0

angin_max_qt
0      1.486537
1      0.292676
2     -1.027247
7     -5.199338
8     -0.261387

```

Gambar 6. Hasil penerapan *Quantile Transform*

Gambar 6 diatas menampilkan sampel hasil transformasi pada beberapa fitur. Penerapan ini menggunakan kelas *quantile transformer* pada *library scikit learn*. *Quantile Transform* bekerja dengan menerapkan fungsi kuantil dengan mengubah nilai berdasarkan peringkatnya dalam distribusi data.

3.4 Data Mining

Tahap ini dilakukan analisis menggunakan *Random Forest Regressor*. Input dalam analisis ini meliputi 'suhu_min_qt', 'kelembapan_avg_qt', sinar_mthr, angin_max_qt, dan arah_angin. Sementara, target pada analisis ini yaitu curah hujan. Penerapan algoritma *Random Forest* disesuaikan dengan *hyperparameter* menggunakan *Grid Search* untuk memperoleh parameter yang optimal. *GridSearch* diimplementasikan menggunakan *GridSearchCV* yang tersedia dalam *library scikit learn*. *Hyperparameter* terbaik ditentukan dengan menguji berbagai kombinasi nilai dari *bootstrap*, *n_estimators*, *max_depth*, dan *max_features*. Pencarian *hyperparameter* ditunjukkan pada Tabel 4.

Tabel 4. Pencarian nilai *hyperparameter*

<i>Hyperparameter</i>	Nilai
<i>bootstrap</i>	[True]
<i>n_estimators</i>	[100, 200, 300, 400, 500]
<i>max_depth</i>	[None, 5, 10, 15]
<i>max_features</i>	[3, sqrt, log2]

Tabel 4 diatas menunjukkan nilai yang digunakan dalam proses pencarian *hyperparameter* optimal. Metode *GridSearchCV* menguji seluruh kombinasi nilai yang ada untuk menemukan nilai terbaik. Pada proses ini, dilakukan pencarian terhadap 60 kombinasi *hyperparameter* di mana setiap kombinasi divalidasi sebanyak 10 kali sehingga terdapat 600 proses validasi. Kemudian, hasil dari proses pencarian kombinasi *hyperparameter* terbaik dengan *bootstrap:True*, *n_estimators: 500*, *max_depth: 5*, dan *max_features: sqrt*. Model akan mencapai kinerja optimal dengan menerapkan teknik *bootstrap sampling*, membangun 500 pohon keputusan, membatasi kedalaman maksimum setiap pohon hingga 5 tingkat dari akar ke daun, serta secara acak memilih sejumlah fitur pada setiap pemisahan (split) sebanyak akar kuadrat dari total fitur yang tersedia. Kemudian, *hyperparameter* tersebut disimpan menjadi model terbaik yang akan digunakan untuk proses evaluasi akhir.

3.5 Evaluation

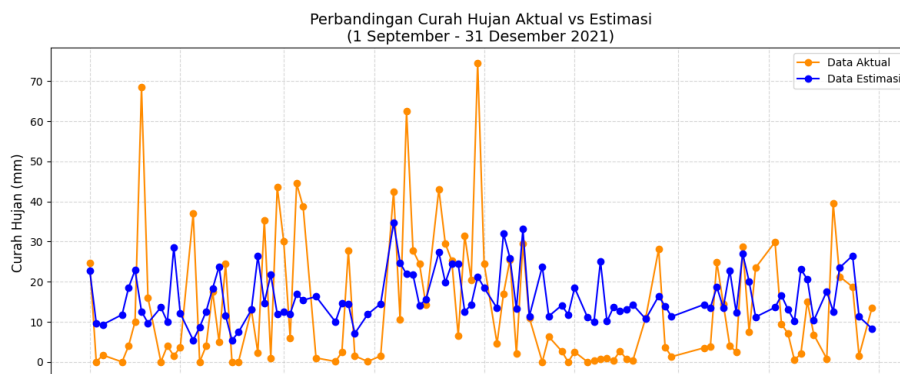
Setelah membangun model pada tahap *data mining*, model tersebut dievaluasi untuk melihat kinerja pada setiap pembagian data menggunakan beberapa nilai *K-Fold Cross Validation*. Pengujian dilakukan dengan mengukur kinerja model terbaik hasil *hyperparameter* menggunakan metrik RMSE (*Root Mean Squared Error*) dan MAE (*Mean Absolut Error*) pada berbagai nilai K.

Tabel 5. Evaluasi metrik model optimasi *Grid Search*

Nilai K-	RMSE	MAE
3	22,1522	15,2439
5	22,0953	15,2455
7	22,0811	15,2487

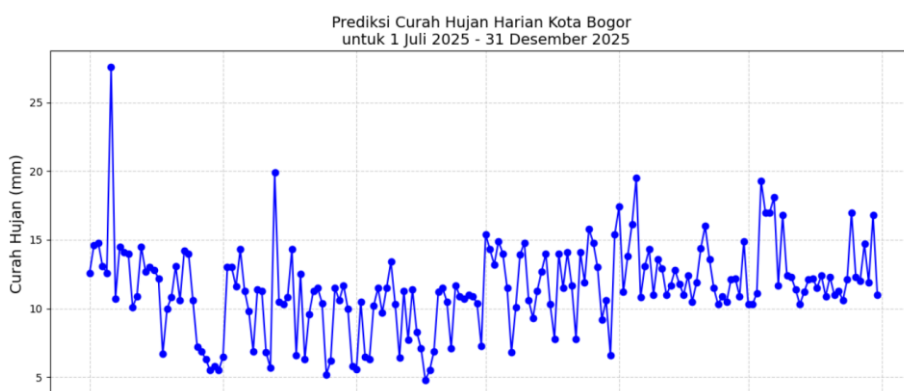
10	21,9812	15,2155
15	21,9267	15,2142

Dari sejumlah nilai K yang diuji, terlihat bahwa semakin tinggi nilai K maka nilai *error* cenderung menurun meskipun tidak terlalu signifikan. Nilai optimal pada kedua metrik tersebut diperoleh saat K = 15 dengan RMSE sebesar 21,92 dan MAE sebesar 15,21. Nilai RMSE sebesar 21,92 mm menunjukkan tingkat kesalahan prediksi yang masih dapat diterima untuk skala harian, mengingat variasi curah hujan di Kota Bogor yang tinggi. Hasil ini sejalan dengan temuan Dewi et al. (2024) bahwa optimasi *hyperparameter* dapat meningkatkan akurasi model Random Forest pada data lingkungan. Meskipun penurunan *error* tidak terlalu signifikan, model dengan optimasi *hyperparameter* tetap menunjukkan kinerja yang lebih baik dibandingkan model *default*. Dengan K = 15 yang menunjukkan kinerja terbaik, maka ditampilkan pola data aktual dan estimasi untuk mengukur ketepatan model yang ditunjukkan pada Gambar 7.



Gambar 7. Pola data aktual dan estimasi

Gambar 7 diatas menampilkan perbandingan antara nilai aktual dan nilai estimasi pada bulan September hingga Desember 2021. Pola tersebut menunjukkan beberapa titik estimasi mendekati nilai aktualnya meskipun cenderung terbatas dalam memperkirakan nilai yang sangat tinggi. Untuk melihat kinerja model lebih lanjut, evaluasi juga dilakukan dengan menampilkan hasil prediksi curah hujan untuk setahun kedepan. Data yang diprediksi mulai dari 1 januari 2025 hingga 31 desember 2025. Prediksi diterapkan dengan model optimasi *hyperparameter* menggunakan Grid Search yang diperoleh pada data pemodelan. Kemudian, hasil prediksi untuk periode bulan juli hingga desember 2025 ditampilkan untuk menggambarkan pola prediksi yang dapat dilihat pada Gambar 8.



Gambar 8. Prediksi curah hujan harian periode Juli hingga Desember 2025

Hasil prediksi menunjukkan pola yang bersifat fluktuatif dengan nilai curah hujan bervariasi antara sekitar 5 mm hingga mendekati 30 mm. Pola ini mencerminkan prediksi bulan Juli hingga Desember 2025 kedepan berada pada intensitas hujan ringan hingga hujan sedang. Meskipun sebagian besar prediksi berada pada intensitas ringan, pola tersebut menunjukkan model dengan optimasi hyperparameter yang diterapkan mampu menghasilkan prediksi harian selama beberapa bulan kedepan.

4. Kesimpulan dan Saran

Berdasarkan hasil pemaparan dan penelitian yang telah dilakukan maka dapat disimpulkan bahwa prediksi curah hujan Kota Bogor dapat dilakukan secara efektif menggunakan algoritma *Random Forest Regression* yang disesuaikan dengan *Quantile Transform* dan optimasi *hyperparameter Grid Search*. Metode ini dapat mengurangi pengaruh *outlier*. Optimasi *Grid Search* menghasilkan nilai *hyperparameter* terbaik dengan *bootstrap: True*, *n_estimators: 500*, *max_depth: 5*, dan *max_features: sqrt*. Evaluasi model dengan *K-Fold Cross Validation* pada $K = 15$ menunjukkan bahwa model memiliki tingkat kesalahan terendah dengan nilai RMSE sebesar 21,92 mm dan MAE sebesar 15,21 mm. Hasil penelitian menunjukkan bahwa model mampu memperkirakan nilai estimasi dengan aktual serta memprediksi curah hujan harian selama setahun kedepan. Temuan ini memberikan kontribusi terhadap penerapan *machine learning* pada mitigasi bencana berbasis data klimatologi lokal.

Rekomendasi yang dapat dipertimbangkan penelitian selanjutnya adalah menambahkan fitur evaporasi dan fenomena iklim global (*El Niño dan La Niña*) sehingga lebih memengaruhi curah hujan. Selain itu, menggunakan *Local Distribution-Based Adaptive Oversampling (LDAO)* untuk menangani ketidakseimbangan distribusi target pada regresi. Serta menggunakan *Random Search* sebagai metode *hyperparameter tuning* untuk mengurangi durasi pemrosesan dan komputasi yang kompleks. Kemudian penelitian selanjutnya dapat mengintegrasikan model *Random Forest* dengan data spasial (misalnya data citra satelit) untuk meningkatkan presisi prediksi.

Referensi

- [1] Molle, B. A., & Larasati, A. F. (2020). Analisis Anomali Pola Curah Hujan Bulanan Tahunan Tahun 2019 Terhadap Normal Curah Hujan (30 Tahun) di Kota Manado dan Sekitarnya. *Jurnal Meteorologi Klimatologi Dan Geofisika*, 7(1).
- [2] BPBD Kota Bogor. (2020). Riwayat Kejadian Bencana. Badan Penanggulangan Bencana Daerah.
- [3] BNPB. (2024). Data Informasi Bencana Indonesia. Badan Nasional Penanggulangan Bencana. <https://dibi.bnpb.go.id/>.
- [4] Milanović, S., Milanović, S. D., Marković, N., Pamučar, D., Gigović, L., & Kostić, P. (2021). Forest Fire Probability Mapping in Eastern Serbia: Logistic Regression Versus Random. *Forest*, 12(1). <https://doi.org/10.3390/f12010005>.
- [5] Dewi, D. A., Wei, A. S., Lin, L. C., & Heng, C. D. (2024). Water Quality Prediction using Random Forest Algorithm and Optimization. *Journal of Applied Data Sciences*, 5(3), 1354-1362.

[6] Nhu, V. H., Shahabi, H., Nohani, E., Shirzadi, A., Al-Ansari, N., Bahrami, S., Miraki, S., Geertsema, M., & Nguyen, H. (2020). Daily Water Level Prediction of Zrebar Lake (Iran): A Comparison Between M5p, Random Forest, Random Tree and Reduced Error Pruning Trees Algorithms. ISPRS International of Geo-Information, 9(8). <https://doi.org/10.3390/ijgi9080479>.

[7] Alzu'bi, F., Al-Rawabdeh, A., & Almagbile, A. (2024). Predicting air quality using random forest: A case study in Amman-Zarqa. The Egyptian Journal of Remote Sensing and Space Sciences, 27(3), 604-613.