

## ANALISIS FAKTOR DAN POLA KEJADIAN BANJIR DI BANDAR LAMPUNG MENGGUNAKAN ARIMA, RANDOM FOREST, DAN XGBOOST

Ahmad Suaif<sup>1</sup>, Eka Sylvianti Rahayu<sup>2</sup>

<sup>1</sup>Teknik Fisika, Fakultas Teknologi Industri, Institut Teknologi Sumatera, Lampung Selatan, Indonesia

<sup>2</sup> Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Bandung, Bandung, Indonesia  
ahmad.suaif@tf.itera.ac.id<sup>1</sup>, ekasylviantirahayu@gmail.com<sup>2</sup>

### Abstract

*Flooding is a significant environmental problem in Bandar Lampung City, influenced by various factors such as rainfall, humidity, etc. This study aims to analyze the factors that contribute to flooding and build a prediction model for flood patterns. The methods used include factor analysis with Random Forest Classifier and prediction model using ARIMA, Random Forest Regressor, and XGBoost Regressor. The results show that rainfall is the dominant factor with a feature importance value of 0.49. From the results of the comparison of prediction models, XGBoost Regressor provides the best performance with an RMSE value of 0.88, From the results of the comparison of prediction models, XGBoost Regressor provides the best performance with an RMSE value of 0.88 and MAE of 0.75, as well as a positive  $R^2$  value of 0.11. The conclusion of this study confirms that the ensemble learning-based machine learning method is superior to statistical models in predicting flood events.*

**Keywords:** Factor Analysis; Pattern; Flood; ARIMA; Random Forest Regressor

### Abstrak

Banjir merupakan permasalahan lingkungan yang signifikan di Kota Bandar Lampung, dipengaruhi oleh berbagai faktor seperti curah hujan, kelembaban, dan lainnya. Penelitian ini bertujuan untuk menganalisis faktor yang berkontribusi terhadap kejadian banjir serta membangun model prediksi pola kejadian banjir. Metode yang digunakan mencakup analisis faktor dengan Random Forest Classifier dan model prediksi menggunakan ARIMA, Random Forest Regressor, dan XGBoost Regressor. Hasil penelitian menunjukkan bahwa curah hujan merupakan faktor dominan dengan nilai *feature importance* sebesar 0.49. Dari hasil perbandingan model prediksi, XGBoost Regressor memberikan performa terbaik dengan nilai RMSE sebesar 0.88 dan MAE sebesar 0.75, serta nilai  $R^2$  positif sebesar 0.11. Kesimpulan dari penelitian ini menegaskan bahwa metode *machine learning* berbasis *ensemble learning* lebih unggul dibandingkan model statistik dalam prediksi kejadian banjir.

**Kata kunci:** Analisis Faktor; Pola Kejadian; Banjir; ARIMA; Random Forest Regressor

**Corresponding author :** ahmad.suaif@tf.itera.ac.id

## PENDAHULUAN

Banjir merupakan salah satu bencana alam yang rawan terjadi di Kota Bandar Lampung (Ramadhani et al., 2023), terutama akibat intensitas hujan yang tinggi dan faktor-faktor lainnya. Dampak banjir meliputi kerugian ekonomi, kerusakan infrastruktur, dan risiko kesehatan masyarakat (Agustri & Asbi, 2020). Oleh karena itu, penelitian terkait faktor penyebab banjir dan model prediksinya menjadi sangat penting untuk mitigasi dan pengambilan keputusan berbasis data.

Penelitian ini bertujuan untuk menganalisis faktor dominan yang mempengaruhi kejadian banjir dan membandingkan beberapa model prediksi. Selanjutnya, urgensi pada penelitian ini yaitu kebutuhan model prediksi yang dapat digunakan untuk mendukung upaya mitigasi banjir di masa mendatang. Sementara itu, kontribusi penelitian ini adalah memberikan wawasan tentang metode prediksi kejadian banjir dalam konteks peramalan bencana alam.

## KAJIAN PUSTAKA

Permasalahan hidrologi, telah menjadi perhatian penting dalam berbagai penelitian. Untuk memahami sistem hidrologi secara komprehensif, pendekatan statistika dan *machine learning* dapat digunakan dengan saling melengkapi.

Metode statistika seperti ARIMA (AutoRegressive Integrated Moving Average) digunakan secara luas untuk menganalisis tren, khususnya dalam prediksi deret waktu. ARIMA efektif dalam menangkap pola musiman dan tren jangka panjang. Namun, metode ini memiliki keterbatasan dalam menangani hubungan non-linier dan kompleks (Kontopoulou et al., 2023). Keterbatasan

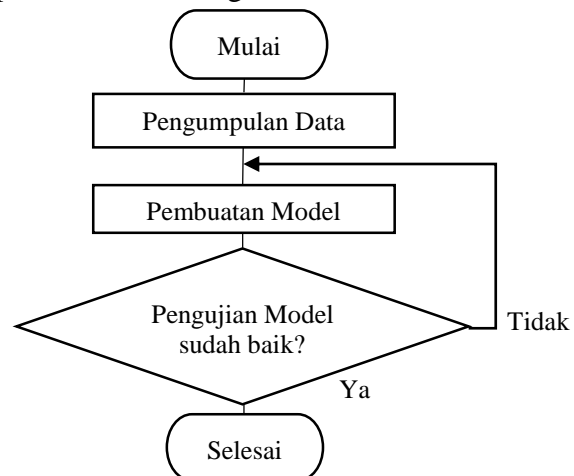
tersebut mendorong integrasi dengan pendekatan *machine learning* untuk memperoleh hasil yang lebih akurat dan adaptif.

Metode Random Forest dimanfaatkan untuk mengidentifikasi variabel penting dalam sistem hidrologi. Penelitian menunjukkan bahwa curah hujan merupakan faktor dominan dalam prediksi banjir (Islam et al., 2023). Sementara itu, XGBoost terbukti unggul dalam memprediksi kejadian hidrometeorologi secara akurat dan mampu menangani data kompleks (Adli Zakaria et al., 2023).

Studi-studi sebelumnya menunjukkan bahwa masing-masing metode memiliki kelebihan sesuai konteks. Penelitian ini bertujuan membandingkan pendekatan ARIMA, Random Forest, dan XGBoost dalam menganalisis pola dan faktor penyebab banjir di Bandar Lampung.

## METODE PENELITIAN

Tahapan kegiatan yang dilakukan dalam penelitian ini sebagaimana Gambar 1 berikut.



Gambar 1. Alur Penelitian

Berikut kegiatan-kegiatan yang dilakukan:

1. Pengumpulan Data

a. Sumber Data

Data bersumber dari gabungan data Badan Nasional Penanggulangan Bencana (BNPB) dan BMKG (Badan Meteorologi Klimatologi dan Geofisika (Nurpambudi & Aziz, 2022). Dataset berisi data dari tahun 2010-2020. Profil dari data sebagaimana Tabel 1.

Tabel 1. Profil Data

No.	Fitur	Jumlah	Tipe Data
1	Tanggal Banjir	50	datetime 64[ns]
2	Rata-Rata Suhu Udara (°C)	50	float64
3	Rata-Rata Kelembapan (%)	50	float64
4	Rata-Rata Angin Permukaan (°)	50	int64
5	Musim - Angin Monsun (°)	50	int64
6	Jumlah Curah Hujan (mm)	50	int64
7	Total Curah Hujan 1 Minggu (mm)	50	int64
8	Terjadi Banjir (1=Ya, 0=Tidak)	50	int64

b. Pemrosesan Data

Data dilakukan pemeriksaan dan penanganan dengan *Interquartile Range* (IQR) dengan hasil sebagaimana Tabel 2.

Tabel 2. Pemeriksaan Data

No.	Fitur	Range
1	Rata-Rata Suhu Udara (°C)	24.0 - 28.8
2	Rata-Rata Kelembapan (%)	68.0 - 98.5
3	Rata-Rata Angin Permukaan (°)	20 - 360
4	Musim - Angin Monsun (°)	130 - 330
5	Jumlah Curah Hujan (mm)	2 - 166
6	Total Curah Hujan 1 Minggu (mm)	0 - 264
7	Terjadi Banjir (1=Ya, 0=Tidak)	0 - 1

c. Penarikan Sampel

Untuk Analisis Faktor, model dibuat dengan data dibagi menjadi data latih (80%) dan data uji (20%). Sementara itu, untuk model Analisis Pola Kejadian, data latih berisi data ts[:3] dari 2010 hingga 2017 dan data uji ts[-3:] berisi data dari 2018 hingga 2020.

2. Pembuatan Model

a. Analisis Faktor

Model Random Forest Classifier digunakan untuk menghitung *feature importance* dari setiap variabel.

b. Analisis Pola Kejadian

Pada Analisis Pola Kejadian, digunakan Analisis Univariat untuk memahami karakteristik variabel.

Pada penelitian ini, dibuat juga model prediksi, dimana digunakan model statistika ARIMA yang dibandingkan dengan model *machine learning*, Random Forest dan XGBoost Regressor. Pada Model Statistika, Model ARIMA diterapkan berdasarkan metode yang terdiri dari komponen: *AutoRegressive* (AR), *Differencing* (I), dan *Moving Average* (MA). Sedangkan pada Model *Machine Learning*, *Random Forest Regressor* digunakan sebagai model regresi non-linear untuk menangkap hubungan kompleks antara variabel *input* dan *output*. XGBoost Regressor digunakan juga sebagai model berbasis *gradient boosting* untuk mempelajari interaksi non-linear (Syahreza et al., 2024).

### 3. Pengujian Model

Model dievaluasi dengan *root mean square error* (RMSE), *mean absolute error* (MAE), dan koefisien determinasi ( $R^2$ ).

### HASIL DAN PEMBAHASAN

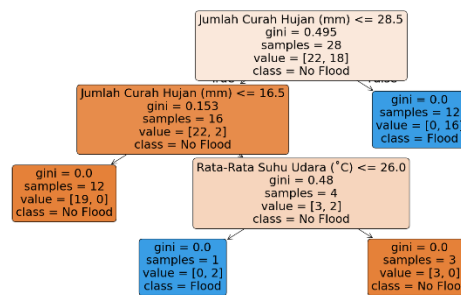
Untuk mengetahui faktor yang paling mempengaruhi kejadian banjir dilakukan perhitungan nilai *feature importance*. Perhitungan nilai ini dilakukan dengan membuat model *Random Forest Classifier*. Adapun hasil dari model ini adalah sebagaimana Tabel 3.

Tabel 3. Analisis Faktor

No.	Fitur	Feature Importance
1	Jumlah Curah Hujan (mm)	0.49
2	Rata-Rata Suhu Udara (°C)	0.16
3	Rata-Rata Kelembapan (%)	0.16
4	Total Curah Hujan 1 Minggu (mm)	0.09
5	Rata-Rata Angin Permukaan (°)	0.08
6	Musim - Angin Monsun (°)	0.02

Berdasarkan Tabel 3, faktor yang paling berpengaruh terhadap kejadian banjir di Kota Bandar Lampung adalah Jumlah Curah Hujan (mm) dengan tingkat skor *feature importance* yang jauh lebih tinggi (0.49) dibandingkan faktor lainnya. Faktor ini menunjukkan bahwa intensitas hujan harian memiliki hubungan yang kuat terhadap kemungkinan terjadinya banjir. Selain itu, Rata-Rata Suhu Udara (°C) (0.16) dan Rata-Rata Kelembapan (%) (0.16) juga berperan penting, mengindikasikan bahwa kondisi atmosfer turut memengaruhi pola banjir. Faktor lain seperti total curah hujan dalam satu minggu, kecepatan angin permukaan, dan angin monsun memiliki pengaruh yang lebih kecil, namun tetap berkontribusi terhadap dinamika kejadian banjir. Hal ini sejalan dengan penelitian sebelumnya (Gracia et al., 2023).

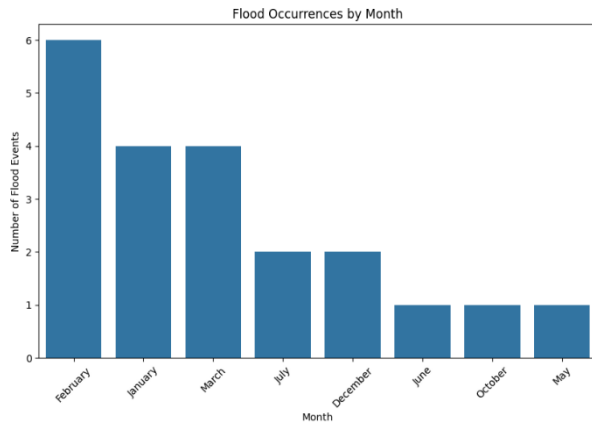
Untuk mengetahui lebih lanjut bagaimana diperoleh hasil *feature importance*, dilakukan visualisasi hasil *Random Forest Classifier* sebagaimana pada Gambar 2 berikut.



Gambar 2. Model Random Forest Classifier untuk Analisis Faktor Banjir

Berdasarkan model yang dihasilkan, terlihat bahwa jika jumlah curah hujan melebihi 28,5 mm, maka kemungkinan terjadi banjir sangat tinggi. Sebaliknya, jika jumlah curah hujan kurang dari 16,5 mm, kejadian banjir hampir tidak terjadi. Selain itu, suhu udara rata-rata juga berkontribusi sebagai faktor sekunder, di mana suhu di bawah 26°C memiliki kecenderungan lebih tinggi terhadap kejadian banjir. Dengan pola ini, dapat diketahui bahwa pemantauan curah hujan dan suhu udara dapat menjadi indikator penting dalam sistem peringatan dini banjir (Gracia et al., 2023)

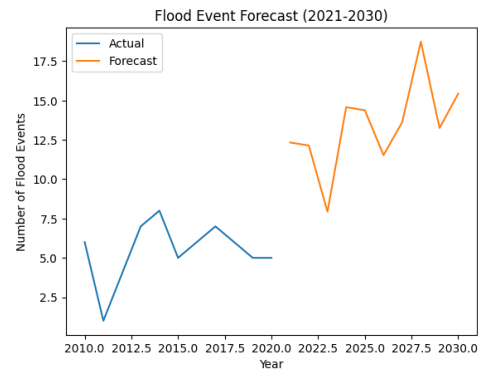
Untuk Analisis Pola Kejadian dilakukan Analisis Univariat. Hasil Analisis sebagaimana Gambar 3.



Gambar 3. Analisis Univariat untuk Analisis Pola Kejadian Banjir

Berdasarkan hasil analisis univariat terhadap kejadian banjir di Kota Bandar Lampung, pola musiman terlihat jelas dengan frekuensi kejadian banjir yang lebih tinggi pada bulan Januari, Februari, dan Maret. Puncak kejadian banjir terjadi pada bulan Februari dengan enam kali kejadian, diikuti oleh Januari dan Maret dengan masing-masing empat kejadian. Sementara itu, bulan-bulan dengan kejadian banjir paling sedikit adalah Mei, Juni, Oktober, dan Desember, yang masing-masing hanya mengalami satu hingga dua kejadian. Pola ini menunjukkan bahwa musim hujan, terutama di awal tahun, memiliki pengaruh signifikan terhadap peningkatan risiko banjir (Alawiyah & Harintaka, 2021).

Untuk Analisis Pola Kejadian lanjutan, dibuat dua kelompok model. Pertama Model Statistik Konvensional ARIMA dengan profil sebagaimana Gambar 4.



Gambar 4. Model Prediksi Pola Kejadian Banjir menggunakan Statistika ARIMA

Gambar 4 menunjukkan prediksi pola kejadian banjir di Bandar Lampung untuk periode 2021-2030 menggunakan model ARIMA. Grafik ini membandingkan data aktual (*Actual*) hingga sekitar tahun 2020 dengan hasil prediksi (*Forecast*) untuk tahun setelahnya. Terlihat bahwa model memperkirakan tren kejadian banjir yang cenderung meningkat, dengan fluktuasi yang cukup besar di beberapa tahun mendatang.

Adapun hasil Evaluasi Model Prediksi berbasis Statistika sebagaimana Tabel 4.

Tabel 4. Evaluasi Model Prediksi berbasis Statistika

No.	Model	RMSE	MAE	R <sup>2</sup>
1	ARIMA	3.56	3.34	-13.26

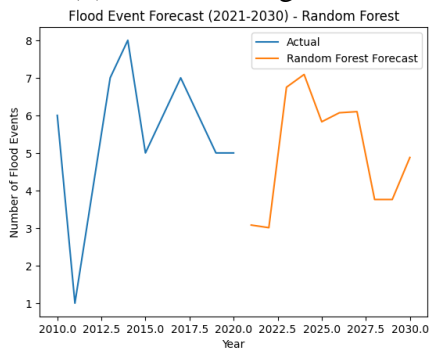
Dari hasil evaluasi model, diperoleh metrik sebagai berikut:

- RMSE (*Root Mean Square Error*): 3.56  
Hal ini menunjukkan bahwa rata-rata kesalahan prediksi model sekitar 3.56 kejadian banjir.
- MAE (*Mean Absolute Error*): 3.34  
Dalam hal ini nilai kesalahan absolut rata-rata sebesar 3.34 kejadian banjir.
- R<sup>2</sup>: -13.26  
Hal ini menunjukkan bahwa model tidak mampu menjelaskan variabilitas data dengan baik, bahkan lebih buruk

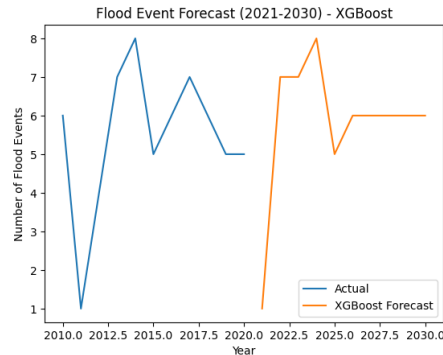
dibandingkan sekadar menggunakan rata-rata sebagai prediksi.

Nilai  $R^2$  yang negatif mengindikasikan bahwa model ARIMA yang digunakan kemungkinan tidak cocok untuk data ini, atau terdapat pola yang tidak dapat ditangkap dengan baik oleh pendekatan statistik konvensional. Hal ini dapat terjadi jika terdapat pola non-linier atau faktor eksternal yang signifikan yang tidak dimasukkan dalam model. Oleh karena itu, diperlukan pendekatan tambahan, seperti model berbasis *machine learning* atau *deep learning*, untuk meningkatkan kemampuan prediksi (Bzdok et al., 2018). Oleh karena inilah, dibuat Kelompok Model Kedua, yaitu model *machine learning*, yaitu Random Forest dan XGBoost Regressor.

Gambar 5 berikut merupakan hasil dari Model Prediksi Pola Kejadian Banjir menggunakan *Machine Learning* (a) Random Forest, dan (b) XGBoost Regressor.



(a)



(b)

Gambar 5. Model Prediksi Pola Kejadian Banjir menggunakan Machine Learning (a) Random Forest, dan (b) XGBoost Regressor

Berdasarkan Gambar 5, terlihat bahwa model berbasis *machine learning* menunjukkan performa yang lebih baik dibandingkan dengan model statistik ARIMA dimana model Random Forest dan XGBoost menampilkan pola berulang yang lebih baik dibandingkan model statistika ARIMA. Dari gambar ini pula dapat diketahui bahwa intensitas kejadian banjir diperkirakan mencapai puncaknya sekitar tahun 2025 dengan jumlah kejadian tertinggi dibandingkan tahun-tahun lainnya. Hal ini perlu menjadi perhatian.

Adapun skor dari metrik Evaluasi Model Prediksi berbasis Machine Learning tampak sebagaimana Tabel 5.

Tabel 5. Evaluasi Model Prediksi berbasis Machine Learning

No.	Model	RMSE	MAE	$R^2$
1	Random Forest	0.88	0.75	0.11
2	XGBoost	0.99	0.99	0.12

Dari hasil evaluasi model, dapat dilihat bahwa model Random Forest memiliki performa terbaik dengan nilai RMSE sebesar 0.88 dan MAE sebesar 0.75, serta nilai  $R^2$  positif sebesar 0.11, yang menunjukkan kemampuan model dalam menangkap pola

data dengan lebih baik dibandingkan model lainnya. Dari hasil ini, dapat disimpulkan bahwa model berbasis machine learning, khususnya Random Forest, memberikan prediksi yang lebih akurat dibandingkan model statistika ARIMA.

### **KESIMPULAN DAN SARAN**

Penelitian ini menunjukkan bahwa analisis faktor dengan Random Forest Classifier dapat mengidentifikasi faktor utama penyebab banjir di Kota Bandar Lampung, dengan curah hujan sebagai faktor dominan. Prediksi menggunakan model machine learning Random Forest Regressor dan XGBoost Regressor memberikan hasil yang lebih akurat dibandingkan metode statistika ARIMA, dengan XGBoost menghasilkan RMSE sebesar 0.88 dan MAE sebesar 0.75, serta nilai  $R^2$  positif sebesar 0.11. Hasil ini sejalan dengan penelitian terdahulu yang menyatakan bahwa model berbasis ensemble learning lebih efektif untuk prediksi kejadian bencana. Temuan ini dapat dimanfaatkan dalam perencanaan mitigasi banjir dan pengambilan kebijakan berbasis data.

### **DAFTAR PUSTAKA**

Adli Zakaria, M. N., Ahmed, A. N., Abdul Malek, M., Birima, A. H., Hayet Khan, M. M., Sherif, M., & Elshafie, A. (2023). Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia. *Heliyon*, 9(7). <https://doi.org/10.1016/j.heliyon.2023.e17689>

Agustri, M. P., & Asbi, A. M. (2020). Tingkat Risiko Bencana Banjir Di Kota Bandar Lampung Dan Upaya Pengurangannya Berbasis Penataan Ruang. *Jurnal Dialog Penanggulangan Bencana*, Vol. 11, No. 1. <https://jdpb.bnppb.go.id>

Alawiyah, A. M., & Harintaka, H. (2021). Identifikasi Genangan Banjir di Wilayah DKI Jakarta Menggunakan Citra Satelit Sentinel-1. *JGISE: Journal of Geospatial Information Science and Engineering*, 4(2), 95. <https://doi.org/10.22146/jgise.68353>

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of Significance: Statistics versus machine learning. In *Nature Methods* (Vol. 15, Issue 4, pp. 233–234). Nature Publishing Group. <https://doi.org/10.1038/nmeth.4642>

Gracia, C., Dewi, C., & Anisa, R. (2023). Analisis Pengaruh Intensitas Curah Hujan terhadap Luas Daerah Genangan Berpotensi Banjir di Kota Bandar Lampung. *Journal of Geodesy and Geomatics*, Vol. 3 No. 2. <https://journal.eng.unila.ac.id>

Islam, K. I., Elias, E., Carroll, K. C., & Brown, C. (2023). Exploring Random Forest Machine Learning and Remote Sensing Data for Streamflow Prediction: An Alternative Approach to a Process-Based Hydrologic Modeling in a Snowmelt-Driven Watershed. *Remote Sensing*, 15(16). <https://doi.org/10.3390/rs15163999>

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. In *Future Internet* (Vol. 15, Issue 8). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/fi15080255>

Nurpambudi, R., & Aziz, R. A. (2022). Prediksi Kejadian Banjir Di Wilayah Kota Bandar Lampung Dengan Metode Artificial Neural Network. *Prosiding*

*Seminar Nasional Darmajaya, Vol. 1.*

<https://jurnal.darmajaya.ac.id/>

Ramadhani, M. A., Amin, M., Ridwan, R., & Tusi, A. (2023). Analisis Tingkat Kerawanan Bencana Banjir di Kota Bandar Lampung Berbasis GIS (Geographic Information System) dan Citra Landsat 8 Oli. *Jurnal Agricultural Biosystem Engineering*, 2(4), 510. <https://doi.org/10.23960/jabe.v2i4.8392>

Syahreza, A., Ningrum, N. K., & Syahrazy, M. A. (2024). Perbandingan Kinerja Model Prediksi Cuaca: Random Forest, Support Vector Regression, dan XGBoost. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 526–534. <https://doi.org/10.29408/edumatic.v8i2.27640>