

MODEL PREDIKSI RISIKO BANJIR PADA AREA PLTM GREEN LAHAT MENGGUNAKAN ALGORITMA XGBOOST

Siti Aminah^{*1}, Debi Gusmaliza², Edo Prasetyo³

¹Teknik Informatika, Institut Teknologi Pagar Alam, Pagar Alam, si.aminah701@gmail.com

²Teknik Informatika, Institut Teknologi Pagar Alam, Pagar Alam, debigusmaliza84@gmail.com

³Teknik Informatika, Institut Teknologi Pagar Alam, Pagar Alam, prasetyoedo873@gmail.com

Abstract.

Flooding is one of the major factors that can disrupt the operation of Mini Hydropower Plants (PLTM) due to increased river discharge and potential infrastructure damage. At PLTM GreenLahat, flood monitoring is still conducted conventionally, making early detection of flood risk less effective. This study aims to develop a flood risk prediction model using the Extreme Gradient Boosting (XGBoost) algorithm. The research applies the Knowledge Discovery in Databases (KDD) method, which consists of data selection, preprocessing, transformation, data mining, and evaluation stages. A total of 182 historical records from PLTM Green Lahat were used, including rainfall and transition channel water level data. Flood risk was classified into three categories: low, medium, and high. Model performance was evaluated using Accuracy, Precision, Recall, and F1-Score and RMSE metrics. The experimental results show that the XGBoost model achieved an accuracy of 78.38%, precision of 81%, recall of 78%, F1-score of 79% and RMSE of 0.5927. The findings indicate that XGBoost is capable of classifying flood risk effectively and can be utilized as a decision support tool to improve flood monitoring and mitigation at PLTM Green Lahat.

Keywords: Flood Risk, XGBoost, Machine Learning, KDD, PLTM.

Abstrak

Banjir merupakan salah satu faktor yang dapat mengganggu operasional Pembangkit Listrik Tenaga Minihidro (PLTM) akibat peningkatan debit sungai dan potensi kerusakan infrastruktur. Pada PLTM Green Lahat, pemantauan banjir masih dilakukan secara konvensional sehingga deteksi dini risiko banjir belum optimal. Penelitian ini bertujuan membangun model prediksi risiko banjir menggunakan algoritma *Extreme Gradient Boosting* (XGBoost). Metode yang digunakan adalah *Knowledge Discovery in Databases* (KDD) yang terdiri dari tahapan data *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Dataset yang digunakan sebanyak 182 data historis PLTM Green Lahat yang meliputi data curah hujan dan tinggi saluran transisi. Risiko banjir diklasifikasikan ke dalam tiga kategori yaitu rendah, sedang, dan tinggi. Evaluasi model dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score* dan *RMSE*. Hasil penelitian menunjukkan bahwa model XGBoost memperoleh nilai *accuracy* sebesar 78,38%, *precision* sebesar 81%, *recall* sebesar 78%, *F1-score* sebesar 79% dan *RMSE* sebesar 0,5927. Hasil tersebut menunjukkan bahwa algoritma XGBoost mampu mengklasifikasikan risiko banjir dengan baik dan berpotensi digunakan sebagai sistem pendukung keputusan dalam pemantauan dan mitigasi banjir pada area PLTM Green Lahat.

Kata Kunci: Risiko Banjir, XGBoost, *Machine Learning*, KDD, PLTM.

1. Pendahuluan

Pembangkit Listrik Tenaga Air (PLTA) dan Pembangkit Listrik Tenaga Minihidro (PLTM) merupakan sumber energi terbarukan yang memanfaatkan aliran air untuk menghasilkan energi listrik. Namun, operasional PLTM sangat dipengaruhi oleh kondisi cuaca dan hidrologi. Peningkatan intensitas curah hujan akibat perubahan iklim dapat meningkatkan risiko banjir sungai (*fluvial flooding*) yang berpotensi mengganggu

operasional pembangkit serta menyebabkan kerusakan pada infrastruktur pendukung [1], [2]. Pada PLTM, banjir dapat menyebabkan peningkatan sedimen, gangguan aliran air, hingga penurunan keandalan sistem pembangkitan [3].

Berdasarkan hasil observasi dan wawancara dengan pengelola PLTM Green Lahat, penilaian risiko banjir masih dilakukan secara konvensional melalui pengamatan curah hujan dan tinggi muka air sungai. Pendekatan tersebut belum memanfaatkan data historis secara terintegrasi sehingga potensi banjir belum dapat diprediksi secara kuantitatif dan tepat waktu. Oleh karena itu, diperlukan suatu pendekatan yang mampu memanfaatkan data historis untuk menghasilkan informasi prediktif mengenai tingkat risiko banjir. Prediksi yang dilakukan secara dini dapat membantu pengelola PLTM dalam mengambil langkah mitigasi yang lebih cepat, mengurangi potensi kerusakan infrastruktur, serta menjaga kontinuitas proses pembangkitan listrik. Pemanfaatan teknologi berbasis data juga memungkinkan proses

pemantauan dilakukan secara lebih objektif dan terukur dibandingkan metode konvensional.

Perkembangan teknologi *machine learning* telah banyak dimanfaatkan dalam prediksi banjir. Asnawi et al. [4], menunjukkan bahwa algoritma XGBoost mampu menghasilkan model prediksi banjir dengan tingkat kesalahan yang rendah. Suaif dan Sylvianti Rahayu [5] juga menemukan bahwa XGBoost memiliki performa yang lebih baik dibandingkan ARIMA dan Random Forest. Selain itu, Aufa et al. [6] menunjukkan bahwa XGBoost mampu mengidentifikasi potensi banjir dengan akurasi yang tinggi. Hasil tersebut menunjukkan bahwa XGBoost memiliki potensi yang baik untuk diterapkan pada prediksi risiko banjir.

Meskipun demikian, penelitian sebelumnya masih berfokus pada data hidrometeorologi umum dan belum banyak memanfaatkan data operasional PLTM sebagai variabel prediksi. Padahal, informasi operasional seperti tinggi saluran transisi memiliki hubungan langsung dengan kondisi aliran air pada pembangkit. Oleh karena itu, penelitian ini mengembangkan model prediksi risiko banjir pada area PLTM Green Lahat menggunakan algoritma XGBoost dengan memanfaatkan data curah hujan, tinggi saluran transisi pukul 07.00, dan tinggi saluran transisi pukul 17.00.

Penelitian dilakukan menggunakan metode *Knowledge Discovery in Databases* (KDD) yang meliputi tahapan data selection, preprocessing, transformation, data mining, dan evaluation. Tujuan penelitian ini adalah membangun model prediksi risiko banjir serta mengevaluasi performa algoritma XGBoost sebagai alat bantu dalam pemantauan dan mitigasi risiko banjir pada PLTM Green Lahat.

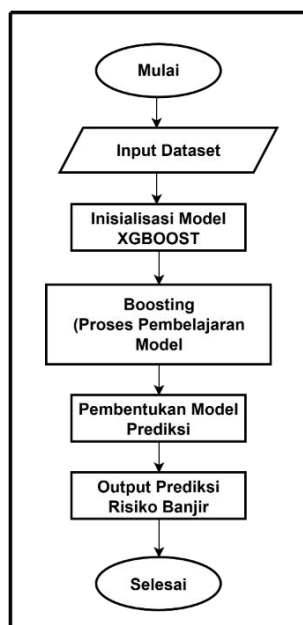
2. Metode

Penelitian ini menggunakan metode *Knowledge Discovery in Databases* (KDD) untuk membangun model prediksi risiko banjir pada area PLTM Green Lahat. Tahapan penelitian meliputi data *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*.

2.1 Dataset Penelitian

Dataset yang digunakan merupakan data historis operasional PLTM Green Lahat sebanyak 182 data pengamatan harian. Variabel yang digunakan terdiri dari curah hujan, tinggi saluran transisi pukul 07.00, dan tinggi saluran transisi pukul 17.00. Variabel tersebut dipilih karena berpengaruh terhadap perubahan debit air dan potensi terjadinya banjir pada area pembangkit.

Data historis tersebut digunakan untuk menggambarkan kondisi hidrologi yang terjadi pada area PLTM Green Lahat selama periode pengamatan. Curah hujan digunakan sebagai indikator kondisi meteorologi, sedangkan tinggi saluran transisi merepresentasikan kondisi aliran air pada sistem pembangkit. Kombinasi ketiga variabel tersebut diharapkan mampu memberikan informasi yang cukup untuk mengidentifikasi tingkat risiko banjir pada area PLTM Green Lahat.

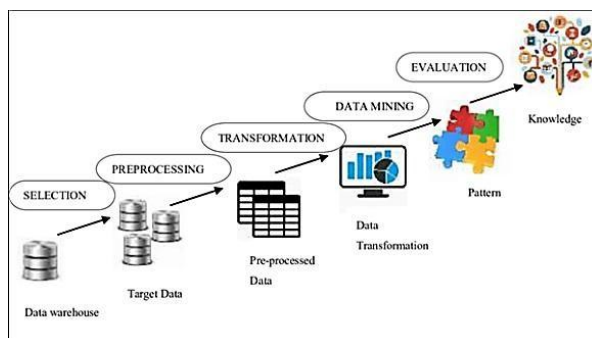


Gambar 1. Flowchart Alur Metode Algoritma Xgboost

2.2 Tahapan Penelitian

Tahapan penelitian menggunakan metode KDD yang terdiri dari:

1. *Data Selection*, yaitu memilih variabel yang digunakan dalam penelitian.
2. *Data Preprocessing*, yaitu membersihkan data dari nilai yang tidak valid dan memastikankualitas data.
3. *Data Transformation*, yaitu mengubah data ke dalam format yang sesuai untuk prosesklasifikasi serta membentuk kategori risiko banjir.
4. *Data Mining*, yaitu membangun model prediksi menggunakan algoritma XGBoost.
5. *Evaluation*, yaitu mengukur performa model menggunakan beberapa metrik evaluasi.



Gambar 2. Tahapan Penelitian Menggunakan Metode KDD

2.3 Pembentukan Kategori Risiko Banjir

Data diklasifikasikan ke dalam tiga kategori risiko banjir berdasarkan tinggi muka air yangdigunakan pada PLTM Green Lahat.

Tabel 1. Kategori Risiko Banjir

Kategori	Kriteria
Rendah	0 cm
Sedang	1–39 cm

Tinggi ≥ 40 cm

2.4 Implementasi Algoritma XGBoost

Algoritma XGBoost digunakan untuk membangun model klasifikasi risiko banjir berdasarkan variabel curah hujan, tinggi saluran transisi pukul 07.00, dan tinggi saluran transisi pukul 17.00. Dataset dibagi menjadi data latih dan data uji, kemudian model dilatih untuk mempelajari pola hubungan antarvariabel sehingga dapat memprediksi kategori risiko banjir.

2.5 Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix. Berdasarkan confusion matrix tersebut dihitung nilai *Accuracy*, *Precision*, *Recall*, *F1-Score* untuk mengetahui tingkat performa model dalam mengklasifikasikan risiko banjir. RMSE sebagai metrik tambahan untuk mengukur rata-rata kesalahan prediksi model secara numerik."

3. Hasil dan Pembahasan

3.1. Penyajian Hasil

3.1.1. Hasil Pembentukan Kategori Risiko Banjir

Berdasarkan proses transformasi data, seluruh data pengamatan diklasifikasikan ke dalam tiga kategori risiko banjir, yaitu risiko rendah, sedang, dan tinggi. Klasifikasi dilakukan berdasarkan tinggi muka air yang digunakan pada operasional PLTM Green Lahat. Hasil kategorisasi ini digunakan sebagai label kelas pada proses pelatihan model XGBoost.

3.1.2. Hasil Klasifikasi Menggunakan XGBoost

Model XGBoost dibangun menggunakan tiga variabel masukan yaitu curah hujan, tinggi saluran transisi pukul 07.00, dan tinggi saluran transisi pukul 17.00. Setelah proses pelatihan dan pengujian dilakukan, diperoleh *confusion matrix* untuk masing-masing kategori risiko banjir sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Hasil *Confusion Matrix*

XGBoost	TP	FN	FP	TN
Risiko Rendah	18	2	2	15
Risiko Sedang	9	4	2	22
Risiko Tinggi	2	2	4	29

Berdasarkan *confusion matrix* tersebut, diperoleh nilai evaluasi model sebagaimana ditunjukkan pada Tabel 3.

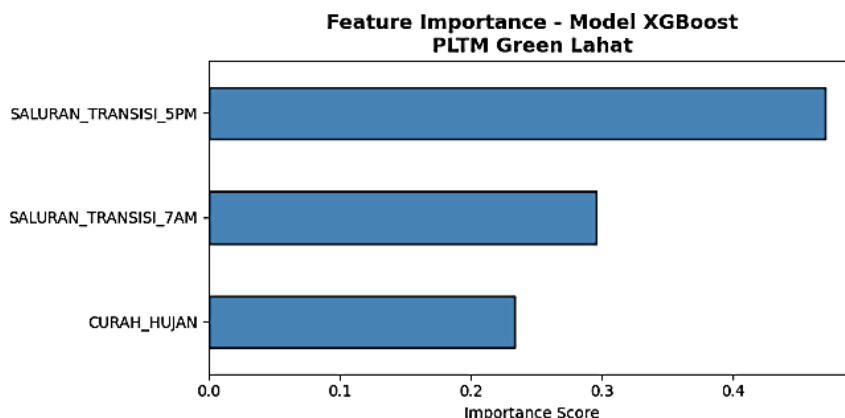
Tabel 3. Hasil Evaluasi Model XGBoost

XGBoost	Precision	Recall	F1-Score	Accuracy	RMSE
Risiko Rendah	90%	90%	90%	-	-
Risiko Sedang	82%	69%	75%	-	-
Risiko Tinggi	33%	50%	40%	-	-
Rata-rata	81%	78%	79%	78,38%	0,5927

Hasil evaluasi menunjukkan bahwa model XGBoost mampu mengklasifikasikan risiko banjir dengan tingkat akurasi sebesar 78,38%. Selain itu, model menghasilkan nilai *precision* sebesar 81%, *recall* sebesar 78%, *F1-score* sebesar 79% dan RMSE 0,5927.

3.1.3. Hasil *Feature Importance*

Analisis *feature importance* dilakukan untuk mengetahui tingkat kontribusi masing-masing variabel terhadap proses klasifikasi risiko banjir menggunakan algoritma XGBoost. Hasil analisis ditunjukkan pada gambar 3.



Gambar 3. Hasil *Feature Importance* Algoritma XGBoost

Berdasarkan Gambar 3, variabel Tinggi Saluran Transisi pukul 17.00 memiliki nilai *importance* tertinggi sebesar 47,04%, diikuti oleh Tinggi Saluran Transisi pukul 07.00 sebesar 29,62%, dan Curah Hujan sebesar 23,33%. Hasil tersebut menunjukkan bahwa tinggi saluran transisi pada pukul 17.00 memberikan kontribusi terbesar dalam proses klasifikasi risiko banjir dibandingkan variabel lainnya.

3.2. Pembahasan

Hasil penelitian menunjukkan bahwa algoritma XGBoost mampu membangun model klasifikasi risiko banjir dengan nilai accuracy sebesar 78,38% dan RMSE sebesar 0,5927 yang mengindikasikan tingkat kesalahan prediksi yang masih dapat diterima. Nilai tersebut menunjukkan bahwa sebagian besar data pengujian dapat diklasifikasikan dengan benar ke dalam kategori risiko yang sesuai. Selain itu, nilai *precision* sebesar 81% menunjukkan bahwa prediksi yang dihasilkan model memiliki tingkat ketepatan yang cukup baik, sedangkan nilai *recall* sebesar 78% menunjukkan bahwa model mampu mengenali sebagian besar kejadian risiko banjir yang terdapat pada data pengujian.

Berdasarkan hasil *feature importance*, variabel tinggi saluran transisi pukul 17.00 memiliki kontribusi terbesar terhadap proses klasifikasi dengan nilai sebesar 47,04%. Hal ini menunjukkan bahwa kondisi tinggi air pada sore hari lebih merepresentasikan potensi peningkatan debit air dibandingkan variabel lainnya. Sementara itu, tinggi saluran transisi pukul 07.00 dan curah hujan juga memberikan kontribusi yang cukup signifikan dalam membantu model membedakan tingkat risiko banjir.

Meskipun model menunjukkan performa yang cukup baik secara keseluruhan, hasil *confusion matrix* menunjukkan bahwa kategori risiko tinggi masih memiliki tingkat klasifikasi yang lebih rendah dibandingkan kategori lainnya. Kondisi ini disebabkan oleh jumlah data risiko tinggi yang relatif sedikit dibandingkan kategori risiko rendah dan sedang. Ketidakseimbangan data menyebabkan model memiliki jumlah contoh yang lebih terbatas untuk mempelajari karakteristik risiko tinggi sehingga memengaruhi kemampuan klasifikasi pada kategori tersebut.

Hasil penelitian ini menunjukkan bahwa kombinasi data curah hujan dan tinggi saluran transisi mampu digunakan untuk membangun model prediksi risiko banjir pada area PLTM Green Lahat. Pemanfaatan data operasional pembangkit sebagai variabel prediksi

memberikan informasi tambahan yang dapat membantu model dalam mengenali pola risiko

banjir secara lebih baik. Dengan demikian, model yang dihasilkan berpotensi digunakan sebagai sistem pendukung keputusan untuk membantu operator dalam melakukan pemantauan kondisi sungai dan mitigasi risiko banjir secara lebih cepat dan efektif.

4. Kesimpulan

Penelitian ini berhasil membangun model prediksi risiko banjir pada area PLTM Green Lahat menggunakan algoritma XGBoost dengan memanfaatkan variabel curah hujan, tinggi saluran transisi pukul 07.00, dan tinggi saluran transisi pukul 17.00. Model yang dibangun mampu mengklasifikasikan risiko banjir ke dalam kategori rendah, sedang, dan tinggi berdasarkan data historis operasional pembangkit.

Hasil penelitian menunjukkan bahwa kombinasi data hidrologi dan data operasional pembangkit dapat digunakan untuk mendukung prediksi risiko banjir. Analisis feature importance menunjukkan bahwa variabel tinggi saluran transisi pukul 17.00 memberikan kontribusi terbesar dalam proses klasifikasi, sehingga menjadi salah satu indikator penting dalam pemantauan kondisi banjir pada area PLTM Green Lahat.

Penelitian ini memberikan kontribusi dalam pemanfaatan algoritma XGBoost untuk klasifikasi risiko banjir pada lingkungan PLTM dengan menggabungkan data hidrologi dan data operasional pembangkit. Untuk penelitian selanjutnya, jumlah data dan variabel penelitian dapat ditambah serta dilakukan perbandingan dengan algoritma *machine learning* lainnya guna meningkatkan performa model prediksi.

5. Referensi

- [1]. Z. KILIÇ, "Exploring the Link between Climate Change and Flood Risk: A Bibliometric Perspective," *International Journal of Environment and Climate Change*, vol. 15, no. 6, pp. 222–232, Jun. 2025, doi: 10.9734/IJECC/2025/V15I64885.
- [2]. E. T. H. Simamora, P. Siagian, and Rahmiani, "Analisis Pengaruh Ketidakstabilan Debit Air Dan Curah Hujan Terhadap Daya Pada Pembangkit Listrik Tenaga Mini Hidro (PLTM) Karo Bumi Energi," *Jurnal Serambi Engineering*, vol. 9, no. 3, pp. 10048–10055, Jul. 2024, Accessed: Dec. 20, 2025. [Online]. Available: <https://jse.serambimekkah.id/index.php/jse/article/view/383>
- [3]. N. Shrestha and Arasu Sibi, "Nepal floods show need for resilient infrastructure amid climate change | AP News," APNews. Accessed: Dec. 20, 2025. [Online]. Available: <https://apnews.com/article/nepal-floods-china-glaciers-recovery-climate-change-disasters-monsoons-5cd9dad3eb7a2e94b96951ab25e9617c>
- [4]. F. M. Asnawi, H. Bisono, A. M. Megantara, and Kusri, "Aplikasi Prediksi Banjir Menggunakan Algoritma XGBoost Berbasis Website," *Journal of Economic, Management, Accounting and Technology*, 2024.
- [5]. A. Suaif and E. Sylvianti Rahayu, "Analisis Faktor dan Pola Kejadian Banjir di Bandar Lampung Menggunakan Arima, Random Forest, dan XGBOOST," 2025.
- [6]. R. A. Afa, E. Altirika, A. Pramudyantoro, Y. B. Pratama, and Z. Wahyuzi, "Pengenalan Pola Kasus Potensi Banjir di Pangkalpinang Dengan Algoritma Random Forest dan Xgboost Menggunakan Google Earth Engine," *Journal of Information Technology and Society*, vol. 3, no. 1, pp. 1–6, Jul. 2025, doi: 10.35438/JITS.V3I1.1415.