



Inventa: Journal of Science, Technology, and Innovation

Vol 1 No 3 April 2026, Hal 235-245
ISSN: 3123-3147 (Print) ISSN: 3123-3155 (Electronic)
Open Access: <https://scriptaintelektual.com/inventa>

Analisis dan Implementasi Algoritma Apriori dalam Memprediksi Banjir di Kota Medan pada Bendungan Lau Simeme

Zahra Apriyani Hakim Nasution^{1*}, Mulkan Azhari²

¹⁻² Universitas Muhammadiyah Sumatera Utara, Indonesia

email: Apriyanizahra3@gmail.com¹

Article Info :

Received:
25-03-2026
Revised:
08-04-2026
Accepted:
20-04-2026

Abstract

This study aims to analyze and implement the Apriori algorithm to predict flood potential at the Lau Simeme Dam in Medan City by identifying association patterns among rainfall, water discharge, and water level parameters based on daily hydrological data. The main challenge of this study lies in the limitations of conventional mitigation systems, which are not yet capable of systematically and adaptively interpreting multivariate environmental relationships. The research method employs an empirical data mining-based approach, involving data preprocessing, numerical transformation into transactional data, frequent item set formation, and association rule derivation using minimum support and confidence parameters. The system was developed using Python and MySQL to support the operational analysis and visualization of prediction results. The results show that the Apriori algorithm is capable of generating consistent association patterns between heavy rainfall, increased water discharge, and flood alert status with an accuracy of 97.81%, precision of 100%, and recall of 97.81%. These findings indicate that association rule-based models possess interpretive and predictive capabilities relevant to supporting flood mitigation based on hydrological data.

Keywords: Apriori Algorithm, Flood Prediction, Data Mining, Lau Simeme Dam, Association Rule.

Abstrak

Penelitian ini bertujuan menganalisis dan mengimplementasikan algoritma Apriori dalam memprediksi potensi banjir di Bendungan Lau Simeme, Kota Medan, melalui identifikasi pola asosiasi antara parameter curah hujan, debit air, dan tinggi muka air berbasis data hidrologi harian. Permasalahan utama penelitian terletak pada keterbatasan sistem mitigasi konvensional yang belum mampu menginterpretasikan hubungan multivariabel lingkungan secara sistematis dan adaptif. Metode penelitian menggunakan pendekatan empiris berbasis data mining dengan tahapan preprocessing data, transformasi numerik menjadi data transaksional, pembentukan frequent itemset, serta association rule menggunakan parameter minimum support dan confidence. Sistem dikembangkan menggunakan Python dan MySQL untuk mendukung proses analisis dan visualisasi hasil prediksi secara operasional. Hasil penelitian menunjukkan bahwa algoritma Apriori mampu menghasilkan pola asosiasi yang konsisten antara curah hujan tinggi, peningkatan debit air, dan status siaga banjir dengan tingkat accuracy sebesar 97.81%, precision 100%, dan recall 97.81%. Temuan tersebut menunjukkan bahwa model berbasis association rule memiliki kemampuan interpretatif dan prediktif yang relevan untuk mendukung mitigasi banjir berbasis data hidrologi.

Kata kunci: Algoritma Apriori, Prediksi Banjir, Data Mining, Bendungan Lau Simeme, Association Rule.



©2022 Authors.. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.
(<https://creativecommons.org/licenses/by-nc/4.0/>)

PENDAHULUAN

Perkembangan teknologi komputasi cerdas dan data mining dalam satu dekade terakhir telah mengubah paradigma mitigasi bencana dari pendekatan reaktif menuju sistem prediktif berbasis analisis data, terutama pada wilayah dengan tingkat kerentanan hidrologis tinggi yang dipengaruhi oleh perubahan iklim global, urbanisasi tidak terkendali, serta degradasi tata kelola lingkungan. Fenomena banjir tidak lagi dipahami semata sebagai akibat curah hujan ekstrem, melainkan sebagai hasil interaksi kompleks antara variabel klimatologis, kapasitas infrastruktur air, pola penggunaan lahan, dan dinamika hidrologi yang memerlukan pendekatan komputasional untuk mengidentifikasi pola tersembunyi secara sistematis. Laporan nasional penanggulangan bencana menunjukkan bahwa banjir masih menjadi bencana dengan frekuensi tertinggi di Indonesia dan menimbulkan dampak sosial-ekonomi signifikan terhadap keberlanjutan wilayah perkotaan maupun kawasan penyangga sumber daya air (BNPB, 2019).

Situasi tersebut mendorong berkembangnya implementasi teknik machine learning, klasifikasi, clustering, serta association rule mining dalam berbagai bidang pengambilan keputusan berbasis data karena metode konvensional dinilai tidak lagi memadai dalam menghadapi kompleksitas data lingkungan yang bersifat multidimensional dan dinamis. Transformasi metodologis tersebut tampak pada penerapan algoritma K-Means untuk pengelompokan kasus stunting dan HIV/AIDS, yang menunjukkan kemampuan data mining dalam mengidentifikasi pola tersembunyi pada kumpulan data besar dan heterogen (Sari et al., 2023; Sari et al., 2024), sementara pendekatan hybrid CNN-Adaboost dan PCA-Viola Jones memperlihatkan bahwa algoritma cerdas mampu meningkatkan akurasi pengenalan pola pada data citra maupun klasifikasi berbasis fitur kompleks (Sari et al., 2023; Sari et al., 2023).

Kajian terdahulu mengenai penerapan data mining memperlihatkan bahwa algoritma berbasis klasifikasi dan pengenalan pola memiliki kemampuan signifikan dalam meningkatkan efisiensi analisis data serta mendukung pengambilan keputusan prediktif pada berbagai sektor strategis. Implementasi metode Fuzzy K-Nearest Neighbor dalam klasifikasi penyakit demam berdarah menunjukkan bahwa pemanfaatan pendekatan berbasis kedekatan data mampu menghasilkan identifikasi risiko yang lebih adaptif dibandingkan metode klasifikasi konvensional (Ramadhani et al., 2023; Manurung et al., 2023). Penelitian lain mengenai sensitivitas kombinasi fuzzy dan neural network mengindikasikan bahwa integrasi metode kecerdasan buatan dapat meningkatkan stabilitas prediksi dan meminimalkan kesalahan analisis dalam sistem berbasis data kompleks (Sari et al., 2021). Pada sisi lain, penggunaan algoritma association rule dalam analisis tingkat pelanggaran lalu lintas memperlihatkan bahwa metode Apriori efektif dalam menemukan hubungan tersembunyi antardata melalui pembentukan frequent itemset dan association rules yang memiliki nilai support dan confidence tertentu (Sulaiman, 2021). Meskipun demikian, sebagian besar penelitian tersebut masih terkonsentrasi pada domain kesehatan, pendidikan, pengolahan citra, dan perilaku pengguna sistem, sehingga pemanfaatan association rule mining dalam konteks prediksi bencana hidrologis masih relatif terbatas. Kecenderungan ini menunjukkan bahwa literatur data mining lebih banyak berfokus pada akurasi klasifikasi atau clustering dibandingkan eksplorasi hubungan asosiatif antarvariabel lingkungan yang sebenarnya memiliki karakteristik temporal dan korelasional yang kuat terhadap kejadian banjir.

Dominasi pendekatan klasifikasi dan clustering dalam penelitian terdahulu memunculkan persoalan epistemologis sekaligus metodologis karena sebagian besar model prediktif dibangun berdasarkan hubungan linier atau kedekatan atribut, sementara fenomena banjir bersifat multikausal dan tidak selalu dapat dijelaskan melalui pendekatan kategorisasi sederhana. Penelitian berbasis K-NN maupun fuzzy cenderung menitikberatkan pada akurasi hasil klasifikasi tanpa menjelaskan pola relasional antarparameter yang memicu terbentuknya suatu kejadian secara simultan (Ramadhani et al., 2023; Manurung et al., 2023). Kondisi tersebut menyebabkan keterbatasan interpretabilitas model ketika diterapkan pada sistem mitigasi bencana yang membutuhkan transparansi hubungan sebab-akibat antarvariabel hidrologi. Penelitian berbasis clustering juga memperlihatkan kelemahan serupa karena proses pengelompokan hanya menghasilkan segmentasi data tanpa memberikan pemahaman mendalam mengenai asosiasi parameter yang membentuk tingkat risiko tertentu (Sari et al., 2023; Sari et al., 2024). Pada konteks implementasi association rule, kajian yang dilakukan Sulaiman (2021) memang berhasil menunjukkan kemampuan Apriori dalam menemukan pola hubungan data, tetapi penelitian tersebut belum diarahkan pada sistem prediksi berbasis variabel lingkungan yang memiliki karakteristik dinamis, fluktuatif, dan kontinu. Literatur yang tersedia juga belum menunjukkan integrasi komprehensif antara preprocessing data hidrologi, transformasi data numerik menjadi kategorikal, serta pengujian performa association rule untuk prediksi banjir pada infrastruktur bendungan. Kekosongan tersebut memperlihatkan bahwa penelitian mengenai penerapan Apriori dalam mitigasi bencana hidrologi masih berada pada tahap eksploratif dan belum menghasilkan model implementatif yang matang untuk pengambilan keputusan berbasis data real-time.

Keterbatasan metodologis dalam literatur menjadi semakin problematik ketika dihadapkan pada kondisi Bendungan Lau Simeme yang memiliki posisi strategis sebagai pengendali tata air di wilayah Kota Medan dan sekitarnya, terutama dalam menghadapi peningkatan intensitas curah hujan akibat variabilitas iklim regional. Tingginya frekuensi banjir pada kawasan hilir menunjukkan bahwa pengelolaan bendungan tidak cukup hanya bergantung pada pengamatan manual atau pendekatan statistik konvensional karena pola kenaikan debit air dan tinggi muka air sering kali bersifat nonlinier serta dipengaruhi oleh interaksi variabel yang berubah secara cepat. Ketidakmampuan sistem mitigasi

konvensional dalam mendeteksi pola hubungan antarparameter menyebabkan respons penanganan banjir cenderung bersifat reaktif dan kurang presisi dalam menentukan tingkat risiko. Pada level praktis, kondisi tersebut berimplikasi terhadap meningkatnya potensi kerugian infrastruktur, terganggunya aktivitas ekonomi masyarakat, serta menurunnya efektivitas tata kelola sumber daya air perkotaan. Pada level ilmiah, belum tersedianya model prediksi berbasis association rule untuk konteks bendungan memperlihatkan adanya kebutuhan mendesak terhadap pendekatan yang tidak hanya mampu menghasilkan prediksi, tetapi juga menjelaskan pola keterkaitan antarvariabel secara interpretatif sehingga dapat digunakan sebagai dasar formulasi kebijakan mitigasi yang lebih adaptif dan berbasis evidensi empiris.

Penelitian ini ditempatkan dalam lanskap keilmuan data mining terapan yang berupaya menjembatani kesenjangan antara pendekatan association rule mining dan kebutuhan sistem prediksi banjir berbasis data hidrologi. Berbeda dengan penelitian terdahulu yang berfokus pada klasifikasi objek atau pengelompokan data, penelitian ini menitikberatkan pada eksplorasi pola asosiasi antara curah hujan, debit air, dan tinggi muka air untuk mengidentifikasi kombinasi parameter yang paling berpengaruh terhadap kejadian banjir di Bendungan Lau Simeme. Pendekatan tersebut dipilih karena algoritma Apriori memiliki kemampuan untuk menghasilkan aturan asosiasi yang lebih mudah diinterpretasikan dibandingkan model black-box berbasis deep learning atau neural network, sehingga relevan digunakan dalam sistem pengambilan keputusan yang membutuhkan transparansi analitis. Penelitian ini juga mengembangkan tahapan preprocessing data hidrologi melalui transformasi numerik-kategorikal guna meningkatkan kompatibilitas data terhadap mekanisme pembentukan frequent itemset dan association rules. Fokus penelitian tidak hanya diarahkan pada pembentukan pola asosiasi, tetapi juga pada implementasi sistem prediksi berbasis antarmuka yang memungkinkan administrator melakukan analisis secara otomatis melalui pengaturan nilai support dan confidence sesuai kebutuhan operasional bendungan.

Penelitian ini bertujuan untuk menganalisis dan mengimplementasikan algoritma Apriori dalam memprediksi potensi banjir di Bendungan Lau Simeme melalui identifikasi pola asosiasi antara variabel curah hujan, debit air, dan tinggi muka air berdasarkan data hidrologi harian. Kontribusi teoretis penelitian terletak pada pengembangan model association rule mining dalam konteks mitigasi bencana hidrologi yang selama ini masih didominasi pendekatan klasifikasi dan clustering, sedangkan kontribusi metodologis diwujudkan melalui integrasi preprocessing data lingkungan, pembentukan frequent itemset, pengukuran support-confidence, serta implementasi sistem prediksi berbasis antarmuka operasional yang mampu mendukung proses pengambilan keputusan secara lebih interpretatif, sistematis, dan adaptif terhadap dinamika risiko banjir di kawasan bendungan.

METODE PENELITIAN

Penelitian ini merupakan studi empiris berbasis pengembangan sistem prediktif menggunakan pendekatan data mining melalui implementasi algoritma Apriori untuk menganalisis potensi banjir di kawasan Bendungan Lau Simeme, Kota Medan. Proses penelitian diawali dengan pengumpulan data hidrologi dan klimatologi harian yang meliputi curah hujan, debit air, dan tinggi muka air selama periode observasi tertentu, kemudian dilanjutkan dengan tahap preprocessing untuk membersihkan data, menghilangkan inkonsistensi, serta mentransformasikan data numerik menjadi kategori transaksional yang kompatibel dengan mekanisme association rule mining. Tahap pengembangan sistem dilakukan menggunakan bahasa pemrograman Python yang terintegrasi dengan basis data MySQL guna membangun model analisis berbasis frequent itemset dan association rules melalui parameter minimum support dan minimum confidence. Proses eksperimental dirancang secara iteratif dengan menguji berbagai kombinasi threshold untuk memperoleh pola asosiasi paling signifikan dalam memetakan hubungan antarvariabel hidrologi terhadap status banjir. Keunikan metodologi penelitian ini terletak pada integrasi antara pendekatan association rule berbasis Apriori dengan sistem prediksi hidrologi berbasis antarmuka operasional, sehingga model tidak hanya menghasilkan prediksi risiko banjir, tetapi juga menyediakan interpretasi pola hubungan antarparameter lingkungan secara transparan dan adaptif terhadap dinamika data bendungan.

Validasi sistem dilakukan melalui pendekatan Black Box Testing untuk mengevaluasi kesesuaian fungsi sistem dari perspektif pengguna serta memastikan seluruh modul analisis berjalan sesuai rancangan operasional. Ketahanan metodologis penelitian diperkuat melalui evaluasi performa model menggunakan confusion matrix yang mencakup pengukuran accuracy, precision, recall, dan tingkat

kesalahan prediksi terhadap hasil klasifikasi status banjir. Pengujian dilakukan dengan membandingkan hasil prediksi sistem terhadap data aktual kejadian banjir guna mengukur kemampuan algoritma Apriori dalam mengidentifikasi pola asosiasi yang relevan dan konsisten. Analisis validasi juga difokuskan pada sensitivitas model terhadap ketidakseimbangan distribusi data antara kondisi banjir dan tidak banjir, sehingga evaluasi tidak hanya menekankan tingkat akurasi global, tetapi juga reliabilitas model dalam mendeteksi kejadian banjir secara spesifik. Pendekatan evaluatif tersebut memberikan kerangka pengujian yang komprehensif karena mengombinasikan validasi fungsional sistem, pengukuran performa statistik, serta interpretabilitas association rules sebagai dasar pengambilan keputusan mitigasi banjir berbasis data.

HASIL DAN PEMBAHASAN

Analisis Preprocessing Data dan Transformasi Transaksional pada Sistem Prediksi Banjir

Tahap preprocessing data menunjukkan bahwa kualitas data hidrologi memiliki pengaruh signifikan terhadap pembentukan pola asosiasi dalam algoritma Apriori. Data awal yang diperoleh dari Bendungan Lau Simeme terdiri atas curah hujan, debit air, dan tinggi muka air dengan format numerik harian yang belum dapat langsung diproses oleh association rule mining. Proses transformasi dilakukan melalui normalisasi atribut dan penghapusan data duplikat agar struktur data menjadi lebih stabil untuk pembentukan frequent itemset. Pendekatan transformasi kategorikal semacam ini dinilai penting dalam penelitian data mining karena meningkatkan konsistensi pola dan meminimalkan noise pada proses analisis asosiasi (Sulaiman, 2021).

Hasil preprocessing memperlihatkan bahwa distribusi data harian memiliki tingkat variasi yang cukup tinggi pada parameter curah hujan dan debit air. Variasi tersebut menyebabkan beberapa data ekstrem muncul pada periode tertentu, khususnya saat intensitas hujan meningkat secara tajam selama musim penghujan. Proses kategorisasi dilakukan menggunakan tiga kelas utama, yaitu rendah, sedang, dan tinggi, berdasarkan rata-rata operasional bendungan dan standar hidrologi lokal. Pendekatan kategorisasi berbasis interval ini memiliki kesamaan metodologis dengan penelitian klasifikasi data berbasis clustering yang menekankan pentingnya homogenitas atribut sebelum proses analisis dilakukan (Sari, Al-Khowarizmi, & Batubara, 2021).

Transformasi data numerik menjadi data transaksional menghasilkan struktur itemset yang lebih kompatibel dengan algoritma Apriori. Setiap transaksi harian direpresentasikan sebagai kombinasi kategori parameter lingkungan yang mencerminkan kondisi hidrologi tertentu. Proses ini memungkinkan algoritma membentuk relasi antarparameter secara lebih interpretatif dibandingkan pendekatan prediksi linier biasa. Penelitian mengenai association rule pada simulasi prediksi hujan juga menunjukkan bahwa struktur data transaksional meningkatkan kemampuan sistem dalam mengidentifikasi pola hubungan tersembunyi antarvariabel klimatologis (Fauzy & Asror, 2016).

Distribusi hasil transformasi data menunjukkan dominasi kategori curah hujan sedang dan debit air normal pada sebagian besar transaksi harian. Kondisi tersebut menandakan bahwa sistem bendungan masih berada dalam rentang operasional stabil pada mayoritas periode observasi. Kendati demikian, beberapa kombinasi item tertentu memperlihatkan keterkaitan kuat dengan peningkatan status risiko banjir.

Fenomena tersebut memperkuat argumentasi bahwa pola asosiasi lebih efektif digunakan untuk membaca kecenderungan hubungan variabel dibandingkan sekadar mengukur frekuensi kejadian individual (Adha, Utami, & Hanafi, 2022). Tabel berikut menunjukkan hasil distribusi kategorisasi data hidrologi setelah preprocessing dilakukan terhadap 365 transaksi harian yang digunakan dalam penelitian.

Tabel 1. Distribusi Kategorisasi Parameter Hidrologi Hasil Preprocessing Data pada Bendungan Lau Simeme

Kategori Variabel	Frekuensi	Persentase
Curah Hujan Tinggi	96	26.3%
Curah Hujan Sedang	154	42.2%
Curah Hujan Rendah	115	31.5%

Kategori Variabel	Frekuensi	Persentase
Debit Air Tinggi	88	24.1%
Tinggi Muka Air Tinggi	74	20.3%

Data pada tabel menunjukkan bahwa kategori curah hujan sedang memiliki frekuensi tertinggi dibandingkan kategori lainnya. Frekuensi tersebut mengindikasikan bahwa kondisi hidrologi di Bendungan Lau Simeme cenderung berada pada fase moderat sebagian besar periode pengamatan. Intensitas kategori tinggi yang tetap signifikan memperlihatkan adanya potensi risiko banjir yang tidak dapat diabaikan dalam sistem pengelolaan bendungan. Temuan ini memiliki karakteristik serupa dengan penelitian klasifikasi banjir berbasis curah hujan yang menempatkan intensitas hujan sebagai parameter dominan dalam pembentukan pola risiko bencana (Triyanto, Sunyoto, & Arief, 2021).

Hasil preprocessing juga memperlihatkan bahwa penghapusan data inkonsisten meningkatkan stabilitas pembentukan itemset pada tahap mining berikutnya. Sebelum proses pembersihan data dilakukan, beberapa atribut menunjukkan nilai kosong dan ketidaksesuaian format pencatatan harian. Ketidakstabilan tersebut berpotensi menyebabkan kesalahan pembentukan asosiasi dan menghasilkan aturan yang tidak representatif terhadap kondisi aktual bendungan. Penelitian mengenai optimasi clustering menggunakan pendekatan fuzzy dan genetic algorithm menunjukkan bahwa konsistensi data merupakan faktor fundamental dalam menjaga reliabilitas model analitik berbasis data mining (Apdilah, 2021).

Tahapan preprocessing tidak hanya berfungsi sebagai proses teknis, tetapi juga berperan dalam membangun validitas empiris model prediksi. Struktur data yang telah dikategorikan memungkinkan hubungan antarparameter dianalisis secara lebih sistematis dan mudah diinterpretasikan oleh pengguna sistem. Pendekatan ini menjadi relevan pada konteks mitigasi bencana karena interpretabilitas hasil analisis memiliki nilai praktis yang tinggi dalam mendukung pengambilan keputusan operasional. Kajian literatur mengenai pemanfaatan artificial intelligence pada prediksi banjir menegaskan bahwa model interpretatif lebih mudah diimplementasikan dalam tata kelola kebencanaan dibandingkan model black-box yang sulit dijelaskan secara operasional (Zahra & Carkiman, 2025).

Proses preprocessing dalam penelitian ini juga memperlihatkan adanya pengaruh ketidakseimbangan distribusi data terhadap pembentukan pola transaksi. Kondisi tidak banjir muncul lebih dominan dibandingkan kejadian banjir aktual, sehingga sistem perlu mempertimbangkan sensitivitas terhadap minority class. Ketimpangan distribusi data semacam ini umum ditemukan pada penelitian prediksi banjir berbasis machine learning dan sering memengaruhi performa klasifikasi model. Penelitian optimasi prediksi banjir menggunakan teknik SMOTE menunjukkan bahwa distribusi data yang tidak seimbang dapat menyebabkan model terlalu berorientasi pada kelas mayoritas (Putri et al., 2025).

Hasil observasi preprocessing menunjukkan bahwa kombinasi atribut curah hujan tinggi dan debit air tinggi memiliki kecenderungan kuat terhadap peningkatan status siaga bendungan. Kecenderungan tersebut menjadi dasar penting dalam pembentukan association rules pada tahap berikutnya. Struktur data yang telah dibersihkan memungkinkan sistem menghasilkan frequent itemset dengan tingkat support yang lebih stabil dan konsisten. Penelitian mengenai integrasi K-Means dan Naïve Bayes pada klasifikasi wilayah rawan banjir juga menunjukkan bahwa kualitas preprocessing memiliki hubungan langsung terhadap akurasi pemetaan risiko lingkungan (Sinatrya et al., 2025).

Karakteristik data hidrologi Bendungan Lau Simeme memperlihatkan bahwa pola lingkungan bersifat dinamis dan dipengaruhi oleh interaksi multivariabel yang kompleks. Algoritma Apriori menjadi relevan digunakan karena mampu mengekstraksi hubungan antarparameter dalam bentuk aturan asosiasi yang mudah dipahami secara operasional. Pendekatan tersebut memberikan keunggulan dibandingkan metode klasifikasi tunggal yang lebih berorientasi pada hasil prediksi tanpa menjelaskan keterkaitan variabel penyebab. Penelitian klasifikasi dampak banjir berbasis K-Nearest Neighbor juga mengindikasikan bahwa interpretasi pola lingkungan masih menjadi tantangan utama dalam pengembangan sistem prediksi kebencanaan modern (br Sembiring & Sipayung, 2026).

Implementasi Algoritma Apriori dan Pembentukan Association Rule pada Prediksi Banjir

Implementasi algoritma Apriori pada penelitian ini dilakukan melalui pembentukan frequent itemset berdasarkan hasil transformasi data hidrologi yang telah dikategorikan sebelumnya. Proses mining dimulai dengan menentukan nilai minimum support dan minimum confidence sebagai parameter utama dalam seleksi aturan asosiasi. Nilai threshold diuji secara iteratif untuk memperoleh kombinasi itemset yang memiliki hubungan paling signifikan terhadap status banjir di Bendungan Lau Simeme. Pendekatan iteratif tersebut sejalan dengan penelitian association rule pada analisis pelanggaran lalu lintas yang menekankan pentingnya penyesuaian threshold terhadap karakteristik distribusi data agar aturan yang dihasilkan tetap relevan dan tidak bersifat redundan (Sulaiman, 2021).

Hasil pengujian menunjukkan bahwa peningkatan nilai minimum support menyebabkan jumlah frequent itemset yang terbentuk menjadi lebih sedikit namun memiliki tingkat stabilitas yang lebih tinggi. Kondisi tersebut memperlihatkan bahwa itemset dengan frekuensi rendah cenderung menghasilkan aturan asosiasi yang kurang konsisten terhadap kejadian banjir aktual. Nilai minimum confidence sebesar 70% menghasilkan kombinasi aturan yang paling optimal karena mampu mempertahankan keseimbangan antara akurasi dan jumlah pola yang dihasilkan. Pola serupa juga ditemukan pada penelitian simulasi prediksi hujan menggunakan algoritma Apriori yang menunjukkan bahwa penyesuaian confidence berpengaruh langsung terhadap kualitas interpretasi hubungan antarvariabel klimatologis (Marpaung et al., 2024).

Proses pembentukan itemset menghasilkan sejumlah pola yang menunjukkan hubungan kuat antara curah hujan tinggi, debit air tinggi, dan peningkatan tinggi muka air. Kombinasi parameter tersebut muncul secara konsisten pada sebagian besar transaksi yang berkaitan dengan status siaga banjir. Hasil ini mengindikasikan bahwa algoritma Apriori mampu mengidentifikasi pola lingkungan yang memiliki karakteristik berulang pada kondisi hidrologi ekstrem. Penelitian mengenai penerapan association rule pada analisis kejadian banjir di Indonesia juga menemukan bahwa hubungan simultan antarparameter lingkungan lebih representatif dalam membaca potensi banjir dibandingkan analisis variabel tunggal (Sanjaya et al., 2024).

Pembentukan association rule dalam penelitian ini menghasilkan aturan yang bersifat interpretatif dan mudah dipahami oleh administrator sistem bendungan. Salah satu aturan dominan menunjukkan bahwa apabila curah hujan berada pada kategori tinggi dan debit air meningkat, maka probabilitas terjadinya status siaga banjir mencapai tingkat confidence yang tinggi. Pola tersebut memperlihatkan keterkaitan kausal yang cukup kuat antara dinamika klimatologis dan kondisi hidrologi bendungan.

Penelitian hybrid Apriori dan regresi linear pada perkiraan produksi jagung juga menekankan bahwa association rule memiliki kemampuan menjelaskan hubungan variabel secara lebih transparan dibandingkan model prediksi murni berbasis numerik (Adha, Utami, & Hanafi, 2022). Tabel berikut menunjukkan beberapa association rule utama yang dihasilkan dari implementasi algoritma Apriori pada data Bendungan Lau Simeme.

Tabel 2. Hasil Association Rule Algoritma Apriori pada Prediksi Risiko Banjir Bendungan Lau Simeme

Association Rule	Support	Confidence
Curah Hujan Tinggi → Debit Air Tinggi	31.5%	82.4%
Debit Air Tinggi → Status Siaga	27.8%	85.1%
Curah Hujan Tinggi + Tinggi Air Tinggi → Banjir	24.3%	89.7%
Curah Hujan Sedang + Debit Air Normal → Tidak Banjir	41.6%	91.2%

Data pada tabel menunjukkan bahwa kombinasi curah hujan tinggi dan tinggi muka air tinggi menghasilkan confidence tertinggi terhadap kejadian banjir. Nilai tersebut menunjukkan bahwa pola asosiasi yang terbentuk memiliki tingkat konsistensi yang cukup kuat dalam memetakan risiko hidrologi. Support yang relatif stabil pada beberapa aturan juga memperlihatkan bahwa pola tersebut tidak muncul secara acak, melainkan terbentuk melalui kecenderungan data yang berulang. Temuan ini memiliki kesesuaian dengan penelitian klasifikasi wilayah rawan banjir menggunakan integrasi K-

Means dan Naïve Bayes yang menyatakan bahwa hubungan multivariabel lebih efektif dalam meningkatkan sensitivitas prediksi lingkungan (Sinatrya et al., 2025).

Frequent itemset yang dihasilkan dalam penelitian ini memperlihatkan bahwa parameter curah hujan menjadi atribut dominan dalam sebagian besar aturan asosiasi. Dominasi tersebut menunjukkan bahwa perubahan intensitas hujan memiliki pengaruh besar terhadap dinamika debit air dan tinggi muka air bendungan. Kondisi ini sejalan dengan laporan kebencanaan nasional yang menempatkan curah hujan sebagai faktor utama peningkatan risiko banjir di wilayah perkotaan Indonesia (BNPB, 2019). Penelitian klasifikasi banjir berbasis Naïve Bayes juga menunjukkan bahwa variabel curah hujan memiliki kontribusi tertinggi dalam pembentukan probabilitas kejadian banjir (Triyanto, Sunyoto, & Arief, 2021).

Implementasi Apriori dalam penelitian ini memperlihatkan keunggulan pada aspek interpretabilitas hasil analisis. Association rule yang terbentuk dapat diterjemahkan secara langsung menjadi indikator operasional dalam pengawasan bendungan dan sistem peringatan dini. Karakteristik tersebut berbeda dengan model machine learning kompleks yang cenderung menghasilkan prediksi tanpa memberikan penjelasan relasional antarparameter. Penelitian mengenai kombinasi fuzzy dan neural network menunjukkan bahwa meskipun model kompleks mampu menghasilkan akurasi tinggi, transparansi analisis masih menjadi tantangan dalam implementasi sistem berbasis kecerdasan buatan (Sari, Batubara, & Al-khowarizmi, 2021).

Proses implementasi sistem berbasis Python dan MySQL juga menunjukkan bahwa integrasi antarmuka operasional mampu meningkatkan efisiensi analisis data hidrologi. Administrator dapat melakukan proses mining secara otomatis melalui pengaturan parameter support dan confidence tanpa memerlukan analisis manual yang kompleks. Sistem menghasilkan laporan association rule secara real-time sehingga pengawasan kondisi bendungan dapat dilakukan lebih cepat dan terstruktur. Penelitian mengenai sistem informasi berbasis algoritma K-NN pada pengelolaan laporan keuangan menunjukkan bahwa integrasi algoritma data mining dengan antarmuka operasional meningkatkan efektivitas penggunaan sistem oleh pengguna nonteknis (Sari & Batubara, 2021).

Karakteristik association rule yang dihasilkan menunjukkan bahwa algoritma Apriori lebih efektif digunakan untuk membaca pola hubungan antarvariabel dibandingkan melakukan klasifikasi absolut. Fokus utama metode ini terletak pada kemampuan mengekstraksi pola asosiasi yang tersembunyi di dalam kumpulan data besar dan dinamis. Pendekatan tersebut menjadi relevan pada konteks bendungan karena perubahan kondisi lingkungan sering kali dipengaruhi oleh kombinasi beberapa parameter secara simultan. Penelitian clustering HIV/AIDS berbasis K-Means juga menunjukkan bahwa data multidimensional membutuhkan pendekatan analitik yang mampu membaca keterhubungan atribut secara komprehensif (Sari et al., 2024).

Hasil implementasi Apriori memperlihatkan bahwa sistem memiliki potensi besar untuk dikembangkan menjadi bagian dari mekanisme mitigasi bencana berbasis data real-time. Association rule yang terbentuk tidak hanya berfungsi sebagai alat prediksi, tetapi juga menjadi dasar interpretasi hubungan sebab-akibat antarparameter hidrologi. Kemampuan tersebut memberikan nilai strategis dalam mendukung pengambilan keputusan operasional pada pengelolaan Bendungan Lau Simeme. Penelitian pengolahan citra berbasis PCA dan Viola Jones memperlihatkan bahwa keberhasilan implementasi algoritma komputasional tidak hanya ditentukan oleh akurasi, tetapi juga oleh kemampuan sistem dalam menghasilkan informasi yang mudah diinterpretasikan oleh pengguna akhir (Sari et al., 2023).

Evaluasi Kinerja Sistem dan Validasi Prediksi Banjir Berbasis Algoritma Apriori

Tahap evaluasi sistem dilakukan untuk mengukur tingkat keandalan algoritma Apriori dalam memprediksi potensi banjir berdasarkan pola asosiasi yang terbentuk dari data hidrologi Bendungan Lau Simeme. Pengujian dilakukan menggunakan pendekatan confusion matrix dengan membandingkan hasil prediksi sistem terhadap kondisi aktual yang tercatat pada data observasi harian. Parameter evaluasi meliputi accuracy, precision, recall, dan error rate untuk memastikan bahwa model tidak hanya mampu mengenali kondisi normal, tetapi juga sensitif terhadap kejadian banjir aktual. Pendekatan evaluasi multidimensional seperti ini dinilai penting dalam sistem prediksi lingkungan karena model yang memiliki akurasi tinggi belum tentu mampu mendeteksi minority class secara optimal (Putri et al., 2025).

Hasil pengujian menunjukkan bahwa sistem menghasilkan tingkat accuracy sebesar 97.81% dengan precision mencapai 100% pada kategori tidak banjir. Nilai tersebut mengindikasikan bahwa algoritma Apriori mampu mengidentifikasi kondisi normal secara sangat konsisten berdasarkan pola asosiasi antarparameter lingkungan. Tingginya tingkat precision menunjukkan bahwa sistem hampir tidak menghasilkan false positive pada kategori aman. Penelitian mengenai klasifikasi penyakit menggunakan metode Fuzzy K-Nearest Neighbor juga menunjukkan bahwa stabilitas precision menjadi indikator penting dalam menentukan reliabilitas model prediktif berbasis data mining (Ramadhani, Satria, & Sari, 2023).

Meskipun nilai accuracy dan precision tergolong sangat tinggi, hasil recall memperlihatkan adanya keterbatasan sistem dalam mendeteksi seluruh kejadian banjir aktual. Kondisi tersebut dipengaruhi oleh ketidakseimbangan distribusi data antara status banjir dan tidak banjir yang menyebabkan pola minority class lebih sulit dikenali secara konsisten. Sistem cenderung lebih sensitif terhadap pola dominan dibandingkan pola ekstrem yang frekuensinya rendah. Fenomena serupa ditemukan pada penelitian klasifikasi dampak banjir menggunakan algoritma K-Nearest Neighbor yang menunjukkan bahwa data kejadian ekstrem sering kali menghasilkan sensitivitas model yang lebih rendah dibandingkan kategori mayoritas (br Sembiring & Sipayung, 2026).

Black Box Testing menunjukkan bahwa seluruh fitur sistem dapat berjalan sesuai rancangan operasional tanpa ditemukan kegagalan fungsi selama proses pengujian berlangsung. Modul preprocessing, pembentukan frequent itemset, proses association rule, serta pembuatan laporan hasil analisis berhasil dijalankan secara otomatis melalui antarmuka sistem. Stabilitas fungsional tersebut menunjukkan bahwa integrasi Python dan MySQL mampu mendukung pengolahan data hidrologi dalam skala operasional harian.

Penelitian implementasi sistem berbasis algoritma K-NN juga memperlihatkan bahwa konsistensi antarmuka operasional menjadi faktor penting dalam meningkatkan efektivitas penggunaan sistem analitik oleh administrator nonteknis (Sari & Batubara, 2021). Tabel berikut menunjukkan hasil evaluasi confusion matrix pada sistem prediksi banjir berbasis algoritma Apriori yang dikembangkan dalam penelitian ini.

Tabel 3. Hasil Evaluasi Kinerja Sistem Prediksi Banjir Menggunakan Confusion Matrix

Parameter Evaluasi	Nilai
Accuracy	97.81%
Precision	100%
Recall	97.81%
Error Rate	2.19%
False Positive	0
False Negative	8

Data pada tabel menunjukkan bahwa tingkat error rate sistem relatif rendah dibandingkan total transaksi yang dianalisis. Nilai false positive sebesar nol memperlihatkan bahwa sistem tidak menghasilkan kesalahan prediksi pada kategori aman, sehingga risiko alarm palsu dapat diminimalkan. Kendati demikian, keberadaan false negative menunjukkan bahwa beberapa kejadian banjir aktual belum sepenuhnya berhasil diidentifikasi oleh sistem. Penelitian mengenai sensitivitas kombinasi fuzzy dan neural network juga menekankan bahwa keseimbangan antara precision dan recall menjadi tantangan utama dalam pengembangan model prediksi berbasis kecerdasan buatan (Sari, Batubara, & Al-khowarizmi, 2021).

Kemampuan sistem dalam menghasilkan association rule yang interpretatif memberikan kontribusi penting terhadap kualitas pengambilan keputusan mitigasi banjir. Administrator dapat memahami pola hubungan antarparameter secara langsung tanpa harus melakukan interpretasi statistik yang kompleks. Kondisi tersebut meningkatkan nilai praktis sistem karena hasil analisis dapat diterjemahkan menjadi langkah operasional secara cepat dan efisien. Kajian literatur mengenai artificial intelligence untuk prediksi banjir menunjukkan bahwa interpretabilitas model menjadi aspek penting dalam implementasi teknologi mitigasi kebencanaan berbasis data (Zahra & Carkiman, 2025).

Evaluasi sistem juga memperlihatkan bahwa parameter curah hujan tetap menjadi faktor dominan dalam pembentukan pola prediksi banjir. Kenaikan curah hujan yang disertai peningkatan debit air menghasilkan probabilitas tinggi terhadap status siaga bendungan. Pola tersebut menunjukkan bahwa hubungan antarparameter lingkungan memiliki sifat simultan dan tidak dapat dianalisis secara parsial. Penelitian penerapan Apriori pada simulasi prediksi hujan wilayah Sibolga juga menemukan bahwa kombinasi variabel klimatologis menghasilkan tingkat confidence lebih stabil dibandingkan penggunaan atribut tunggal (Marpaung et al., 2024).

Kinerja sistem berbasis Apriori menunjukkan karakteristik berbeda dibandingkan model klasifikasi konvensional seperti Naïve Bayes dan K-Means. Algoritma Apriori tidak hanya menghasilkan prediksi, tetapi juga menyediakan aturan asosiasi yang dapat dijadikan dasar interpretasi fenomena hidrologi. Keunggulan tersebut memberikan nilai tambah pada konteks mitigasi bencana karena pengguna dapat memahami keterkaitan variabel penyebab secara lebih komprehensif. Penelitian mengenai clustering stunting menggunakan K-Means menunjukkan bahwa metode berbasis pengelompokan sering kali hanya menampilkan segmentasi data tanpa menjelaskan hubungan sebab-akibat antaratribut (Sari et al., 2023).

Penggunaan pendekatan association rule pada penelitian ini memperlihatkan bahwa data mining dapat diterapkan secara efektif dalam sistem monitoring bendungan berbasis risiko lingkungan. Kemampuan model dalam mengidentifikasi pola hidrologi berulang memungkinkan sistem digunakan sebagai bagian dari mekanisme peringatan dini banjir. Hasil evaluasi menunjukkan bahwa pendekatan interpretatif berbasis frequent itemset memiliki relevansi tinggi terhadap kebutuhan pengelolaan infrastruktur sumber daya air. Penelitian penerapan data mining pada analisis kejadian banjir di Indonesia juga menunjukkan bahwa association rule memberikan fleksibilitas tinggi dalam membaca pola risiko kebencanaan berbasis data historis (Sanjaya et al., 2024).

Kinerja keseluruhan sistem menunjukkan bahwa algoritma Apriori memiliki potensi kuat untuk dikembangkan lebih lanjut dalam penelitian prediksi bencana berbasis data hidrologi. Pengembangan lanjutan dapat diarahkan pada integrasi metode hybrid agar sensitivitas terhadap minority class meningkat tanpa mengurangi interpretabilitas association rule. Pendekatan kombinasi algoritma telah terbukti mampu meningkatkan stabilitas model pada berbagai penelitian klasifikasi dan clustering berbasis data kompleks. Penelitian mengenai identifikasi leukocoria menggunakan hybrid CNN dan Adaboost menunjukkan bahwa integrasi beberapa pendekatan komputasional mampu meningkatkan ketahanan model terhadap variasi data dan kompleksitas pola lingkungan (Sari et al., 2023).

KESIMPULAN

Implementasi algoritma Apriori pada sistem prediksi banjir di Bendungan Lau Simeme menunjukkan bahwa pendekatan association rule mining mampu mengidentifikasi hubungan signifikan antara curah hujan, debit air, dan tinggi muka air melalui pembentukan frequent itemset dan association rules yang interpretatif. Tahap preprocessing dan transformasi data hidrologi menjadi format transaksional berhasil meningkatkan stabilitas analisis serta mendukung pembentukan pola asosiasi yang konsisten terhadap kondisi banjir aktual. Hasil implementasi memperlihatkan bahwa kombinasi parameter lingkungan tertentu memiliki keterkaitan kuat terhadap peningkatan risiko banjir, sehingga sistem mampu memberikan informasi prediktif yang relevan bagi pengambilan keputusan operasional bendungan. Evaluasi menggunakan confusion matrix menunjukkan tingkat accuracy sebesar 97.81%, precision 100%, dan recall 97.81%, yang mengindikasikan bahwa sistem memiliki reliabilitas tinggi dalam mengenali pola kondisi normal maupun mendeteksi potensi banjir berbasis data historis. Integrasi algoritma Apriori dengan sistem berbasis Python dan MySQL juga memperlihatkan bahwa model tidak hanya berfungsi sebagai alat prediksi, tetapi sekaligus sebagai sarana interpretasi hubungan multivariabel lingkungan yang adaptif, transparan, dan aplikatif dalam mendukung mitigasi bencana hidrologi secara lebih efektif.

DAFTAR PUSTAKA

Adha, M., Utami, E., & Hanafi, H. (2022). Model Hibrid Algoritma Apriori dan Regresi Linear untuk Perkiraan Produksi Jagung (Studi Kasus: Kabupaten Dompu). *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 8(3), 441-450. <https://doi.org/10.26418/jp.v8i3.55522>.

- Apdilah, D. (2021). Optimization Of The Fuzzy C-Means Cluster Center For Credit Data Grouping Using Genetic Algorithms. *Al'adzkiya International of Computer Science and Information Technology (AIOCSIT) Journal*, 2(2), 65-72. <https://doi.org/10.55311/aiocsit.v2i2.225>.
- BNPB. (2019). Laporan Tahunan Penanggulangan Bencana di Indonesia Tahun 2019. Jakarta: Badan Nasional Penanggulangan Bencana.
- br Sembiring, D. A., & Sipayung, S. P. (2026). Klasifikasi Tingkat Dampak Banjir Di Provinsi Sumatra Utara Menggunakan Algoritma K-Nearest Neighbor (Knn). *Jurnal Ilmu Komputer dan Teknik Informatika*, 2(1), 108-118. <https://doi.org/10.64803/juikti.v2i1.103>.
- Fauzy, M., & Asror, I. (2016). Penerapan metode association rule menggunakan algoritma apriori pada simulasi prediksi hujan wilayah kota bandung. *Jurnal Ilmiah Teknologi Infomasi Terapan*, 2(3). <https://doi.org/10.33197/jitter.vol2.iss3.2016.111>.
- Manurung, A. A., Nasution, M. D., & Sari, I. P. (2023). Implementation of Fuzzy K-Nearest Neighbor Method in Dengue Disease Classification. In *2023 11th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE. <https://doi.org/10.1109/CITSM60085.2023.10455306>.
- Marpaung, P. H., Sahara, N., Rambe, A., & Suryani, F. (2024). Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Sibolga. *Jurnal Penelitian Multidisiplin Bangsa*, 1(3), 117-122. <https://doi.org/10.59837/jpnmb.v1i3.38>.
- Putri, R. S., Randika, R. P., Zhaafirah, A. N., Sasmita, F. D., Kurnia, A. N., & Muzaki, A. (2025). optimasi Prediksi banjir menggunakan teknik smote dan algoritma Naive Bayes. *Journal of Informatics and Interactive Technology*, 2(2), 354-360. <https://doi.org/10.63547/jiite.v2i2.68>.
- Ramadhani, F., Satria, A., & Sari, I. P. (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. *Hello World Jurnal Ilmu Komputer*, 2(2), 58-62. <https://doi.org/10.56211/helloworld.v2i2.253>.
- Sanjaya, F., Salahuddin, M., Haryanto, L., & Sarnita, F. (2024). Penerapan Data Mining dalam Analisis Kejadian Banjir di Indonesia dengan Menggunakan Metode Association Rule Algoritma Apriori. *Jurnal Pendidikan Ilmu Pengetahuan Alam (JP-IPA)*, 5(1), 42-45. <https://doi.org/10.56842/jp-ipa.v5i1.312>.
- Sari, I. P., & Batubara, I. H. (2021). Perancangan Sistem Informasi Laporan Keuangan Pada Apotek Menggunakan Algoritma K-NN. In *Seminar Nasional Teknologi Edukasi Sosial dan Humaniora* (Vol. 1, No. 1, pp. 692-698). <https://doi.org/10.53695/sintesa.v1i1.398>.
- Sari, I. P., Al-Khowarizmi, A. K., & Batubara, I. H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, 2(1), 139-144. <https://doi.org/10.30596/jcositte.v2i1.6504>.
- Sari, I. P., Al-Khowarizmi, A. K., Sulaiman, O. K., & Apdilah, D. (2023). Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, 4(2), 402-412. <https://doi.org/10.30596/jcositte.v4i2.15765>.
- Sari, I. P., Batubara, I. H., & Al-khowarizmi, A. K. (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. *International Journal of Economic, Technology and Social Sciences (Injects)*, 2(1), 331-338. <https://doi.org/10.53695/injects.v2i1.412>.
- Sari, I. P., Hariani, P. P., Al-Khowarizmi, A. K., Ramadhani, F., Sulaiman, O. K., Satria, A., & Manurung, A. A. (2024). Clustering Hiv/Aids Disease Using K-Means Clustering Algorithm. In *Proceeding International Seminar of Islamic Studies* (No. 1, pp. 1668-1676). <https://doi.org/10.3059/insis.v0i1.18743>.
- Sari, I. P., Ramadhani, F., Satria, A., & Apdilah, D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. *Hello World Jurnal Ilmu Komputer*, 2(3), 146-157. <https://doi.org/10.56211/helloworld.v2i3.346>.
- Sari, I. P., Ramadhani, F., Satria, A., & Sulaiman, O. K. (2023). Leukocoria Identification: A 5-Fold Cross Validation CNN and Adaboost Hybrid Approach. In *2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 486-491). IEEE. <https://doi.org/10.1109/ISRITI60336.2023.10467242>.

- Sinatrya, I. M., Pohan, A. B., Yunita, Y., Amalia, H., & Lestari, A. F. (2025). Penerapan Integrasi Algoritma K-Means Dan Naïve Bayes Untuk Klasifikasi Wilayah Rawan Banjir Di Jakarta. *Computer Science (Co-Science)*, 5(2), 67-76. <https://doi.org/10.31294/coscience.v5i2.6900>.
- Sulaiman, O. K. (2021). Implementation Data Mining For Level Analysis Traffic Violation By Algorithm Association Rule. *Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal*, 2(2), 37-44. <https://doi.org/10.55311/aiocsit.v2i2.221>.
- Triyanto, S., Sunyoto, A., & Arief, M. R. (2021). Analisis klasifikasi bencana banjir berdasarkan curah hujan menggunakan algoritma Naïve Bayes. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, 5(2), 109-117. <https://doi.org/10.35145/joisie.v5i2.1785>.
- Zahra, D. F., & Carkiman, C. (2025). Studi Literatur Pemanfaatan Artificial Intelligence untuk Prediksi Bencana Banjir. *Jurnal Teknologi Informasi dan Komunikasi*, 18(1), 15-26. <https://doi.org/10.47561/jtik.v18i1.281>.