JURNAL MANAJEMEN TEKNOLOGI INFORMATIKA
ISSN 2988-0645 (*print*) dan 2987-8691 (*online*)
Jl. Veteran No.26B, Purus, Kec. Padang Barat, Kota Padang, Sumatera Barat 25115
Website: https://www.jentik.org | E-mail: admin@jentik.org

Evaluasi Algoritma Decision Tree Random Forest K-Means untuk Klasifikasi Kerusakan Rumah Banjir

Khairil Hamdi[1], Yulia Fatmi[2] , Delpima Suhita[3]
khairilhamdi@jayanusa.ac.id1, yuliafatmi@unp.ac.id2, sdelpima@gmail.com3
[1]Sistem Informasi, STMIK Jaya Nusa Padang
[2]Program Studi Informatika, Departemen Teknik Eloktronika, Fakultas Tenik, Universitas Negeri Padang
[3]Program Studi Teknik Mesin, Fakultas Teknik, Universitas Krisnadwipayana

Informasi Artikel

Diterima : 09-08-2026
Direview : 15-08-2026
Disetujui : 31-08-2026

Kata Kunci

banjir; klasifikasi kerusakan rumah; Decision Tree; Random Forest; K-Means; machine learning

Abstrak

Penilaian tingkat kerusakan rumah pascabanjir diperlukan untuk membantu penentuan prioritas penanganan dan alokasi bantuan secara lebih terukur. Penelitian ini mengevaluasi tiga algoritma, yaitu Decision Tree, Random Forest, dan K-Means, pada dataset simulasi yang terdiri atas 120 rumah terdampak banjir di 19 kabupaten/kota di Sumatera Barat. Setiap data direpresentasikan oleh 11 indikator kondisi bangunan, meliputi kondisi bangunan, atap, rangka atap, kolom atau balok, dinding, lantai, langit-langit, instalasi listrik, serta pintu atau jendela. Kategori target dibagi menjadi rusak ringan, rusak sedang, dan rusak berat berdasarkan persentase kerusakan. Untuk algoritma supervised, data dibagi menjadi 96 data latih dan 24 data uji, sedangkan K-Means memproses seluruh data berdasarkan kemiripan fitur. Hasil menunjukkan bahwa Decision Tree menghasilkan akurasi tertinggi sebesar 79,17%, diikuti Random Forest sebesar 70,83%, dan K-Means sebesar 68,33%. Decision Tree mengklasifikasikan 30 rumah sebagai rusak ringan, 81 rusak sedang, dan 9 rusak berat.

Keywords

Keywords: *flood; housing damage classification; Decision Tree; Random Forest; K-Means; machine learning*

Abstrak

Post-flood housing damage assessment is required to support more measurable decisions on response priorities and aid allocation. This study evaluates three algorithms—Decision Tree, Random Forest, and K-Means—using a simulated dataset of 120 flood-affected houses representing 19 regencies and cities in West Sumatra. Each record contains 11 building-condition indicators covering the standing condition of the structure, roof, roof frame, columns or beams, walls, floors, ceilings, electrical installations, and doors or windows. The target classes consist of minor, moderate, and severe damage based on the reported percentage of damage. For the supervised algorithms, the data were divided into 96 training records and 24 testing records, while K-Means processed all records based on feature similarity. The results show that Decision Tree achieved the highest accuracy of 79.17%, followed by Random Forest at 70.83% and K-Means at 68.33%. Decision Tree classified 30 houses as having minor damage, 81 as moderate damage, and 9 as severe damage.

A. Pendahuluan

Banjir merupakan salah satu bencana yang menimbulkan kerusakan langsung pada bangunan rumah. Kerusakan dapat berada pada tingkat ringan, sedang, hingga berat, sehingga penanganan pascabencana memerlukan mekanisme penilaian yang konsisten. Tanpa pengelompokan berbasis data, prioritas perbaikan dan penyaluran bantuan berisiko tidak sesuai dengan kondisi rumah terdampak [1], [2].

Pemanfaatan machine learning dapat membantu mengubah indikator kondisi bangunan menjadi kategori kerusakan yang lebih mudah digunakan dalam proses pengambilan keputusan. Decision Tree menghasilkan aturan keputusan berbentuk IF-THEN yang dapat ditelusuri, Random Forest menggabungkan banyak pohon keputusan untuk meningkatkan stabilitas model, sedangkan K-Means membentuk kelompok berdasarkan kemiripan data tanpa menggunakan label pada proses pembelajaran [3]–[7].

Laporan awal menggunakan 120 data simulasi rumah terdampak banjir di Sumatera Barat dengan 11 variabel kondisi bangunan. Data tersebut disusun menyerupai pola kejadian banjir, tetapi bukan hasil survei lapangan resmi. Oleh sebab itu, hasil penelitian harus dipahami sebagai studi simulasi dan bukti konsep, bukan sebagai dasar langsung untuk penetapan bantuan aktual [1], [8].

Penelitian ini bertujuan: (1) membandingkan kinerja Decision Tree, Random Forest, dan K-Means; (2) mengidentifikasi fitur kerusakan yang dominan; (3) menganalisis distribusi kategori kerusakan yang dihasilkan setiap algoritma; dan (4) menghitung estimasi kebutuhan bantuan berdasarkan skema nominal yang digunakan dalam laporan. Kontribusi utama penelitian terletak pada penyajian evaluasi komparatif yang menghubungkan hasil klasifikasi dengan kebutuhan prioritas penanganan pascabanjir.

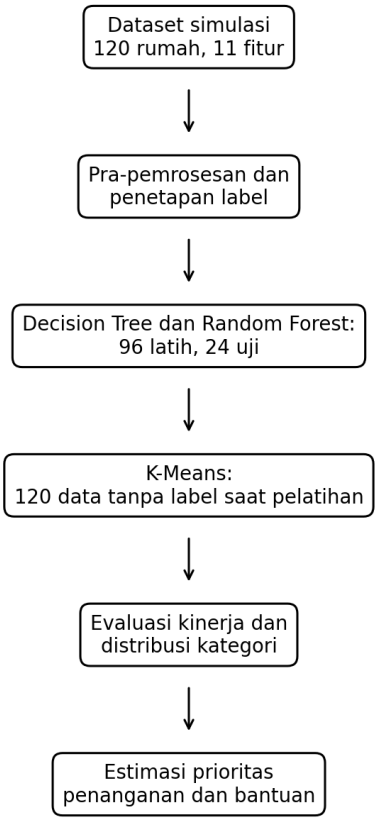
B. Metode Penelitian

Metode penelitian ini dirancang untuk mengevaluasi kemampuan algoritma pembelajaran mesin dalam mengklasifikasikan tingkat kerusakan rumah akibat banjir berdasarkan indikator kondisi fisik bangunan. Pendekatan yang digunakan bersifat kuantitatif-komputasional karena proses analisis dilakukan melalui pengolahan data numerik, pembentukan model, serta perbandingan hasil klasifikasi dari beberapa algoritma. Dengan pendekatan ini, setiap tahapan penelitian diarahkan untuk menghasilkan ukuran kinerja yang dapat dibandingkan secara objektif.

Objek penelitian berupa data simulasi rumah terdampak banjir yang memuat atribut kerusakan pada komponen bangunan, seperti atap, rangka atap, kolom atau balok, dinding, lantai, langit-langit, instalasi listrik, serta pintu atau jendela. Data tersebut digunakan sebagai dasar untuk membangun skenario klasifikasi kerusakan menjadi tiga kategori, yaitu rusak ringan, rusak sedang, dan rusak berat. Tahapan metode disusun mulai dari persiapan data, penentuan variabel, pembentukan kategori kerusakan, pemodelan algoritma, evaluasi kinerja, hingga analisis hasil utuk mendukung prioritas penanganan pascabanjir.

Penelitian menggunakan pendekatan eksperimen komputasional. Tahapan penelitian meliputi penyusunan dataset simulasi, pemeriksaan variabel, pembentukan label kerusakan, pemodelan menggunakan tiga algoritma, evaluasi

hasil, analisis fitur, serta estimasi bantuan. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian komparasi algoritma

1. Dataset dan Variabel

Dataset terdiri atas 120 rumah terdampak banjir yang mewakili ruang lingkup 19 kabupaten/kota di Sumatera Barat. Seluruh data merupakan data dummy atau simulasi. Variabel target adalah persentase kerusakan, sedangkan variabel prediktor terdiri atas 11 kondisi fisik rumah sebagaimana dirangkum berikut.

Ringkasan 11 variabel yang digunakan adalah sebagai berikut:

1. Bangunan berdiri: Kondisi bangunan: berdiri, sebagian berdiri, atau tidak berdiri.
2. Atap bocor/lepas: Tingkat kerusakan atau terlepasnya bagian atap.
3. Rangka atap patah: Keberadaan kerusakan atau patah pada rangka atap.
4. Kolom/balok retak: Retakan pada struktur kolom atau balok.
5. Kolom/balok patah: Kerusakan berat pada struktur utama.
6. Dinding retak: Keberadaan retakan pada dinding.
7. Dinding roboh: Kondisi keruntuhan dinding rumah.
8. Lantai rusak: Kerusakan pada bagian lantai.
9. Langit-langit lepas: Kerusakan atau terlepasnya plafon.
10. Instalasi listrik rusak: Tingkat kerusakan instalasi listrik.
11. Pintu/jendela rusak: Kerusakan pada pintu dan jendela.

2. Pembentukan Kategori Kerusakan

Kategori kerusakan mengikuti batas yang digunakan dalam laporan: rusak ringan untuk persentase kerusakan ≤30%, rusak sedang untuk nilai di atas 30% sampai 70%, dan rusak berat untuk nilai >70%. Laporan menyebutkan bahwa penyusunan data mengacu pada Peraturan Menteri PUPR Nomor 29/PRT/M/2018. Karena dokumen sumber tidak memuat uraian teknis pembobotan setiap indikator, artikel ini mempertahankan kategori akhir yang tersedia tanpa merekonstruksi bobot yang tidak terdokumentasi.

Tabel 1. Kategori kerusakan dan asumsi bantuan (*diturunkan dari perhitungan laporan)

Kategori	Batas	Bantuan/rumah*
Rusak ringan	≤30%	Rp10.000.000
Rusak sedang	>30%–70%	Rp25.000.000
Rusak berat	>70%	Rp50.000.000

3. Pemodelan Algoritma

Decision Tree. Model membagi data secara bertahap berdasarkan fitur yang paling informatif hingga menghasilkan node daun sebagai keputusan akhir. Keunggulan utamanya adalah interpretabilitas karena jalur keputusan dapat diterjemahkan menjadi aturan IF-THEN [4].

Random Forest. Model membangun sejumlah pohon keputusan dari sampel dan subset fitur yang berbeda, kemudian menggabungkan hasil prediksi. Pendekatan ensemble ini umumnya mengurangi ketergantungan pada satu struktur pohon, meskipun kinerjanya tetap dipengaruhi ukuran dan karakteristik data [5].

K-Means. Algoritma membagi data ke dalam K kelompok berdasarkan kedekatan terhadap centroid. K-Means tidak menggunakan label ketika membentuk klaster, sehingga evaluasi terhadap kategori kerusakan memerlukan pemetaan klaster ke label setelah proses klasterisasi [6]. Prosedur pemetaan tersebut tidak dijelaskan dalam laporan sumber dan dicatat sebagai keterbatasan.

4. Pembagian Data dan Evaluasi

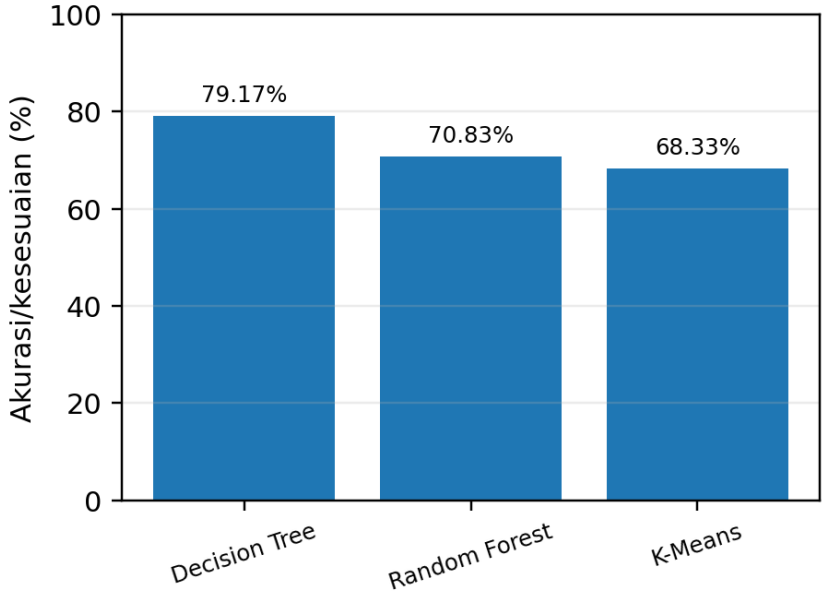
Decision Tree dan Random Forest menggunakan pembagian 80:20, yaitu 96 data latih dan 24 data uji. K-Means memproses seluruh 120 data karena termasuk algoritma unsupervised. Kinerja utama dinilai menggunakan akurasi atau nilai kesesuaian yang dilaporkan, distribusi kategori kerusakan, waktu proses pada tabel perbandingan akhir, serta kemampuan model menghasilkan tiga kategori kerusakan.

Akurasi dihitung sebagai proporsi jumlah prediksi benar terhadap seluruh data uji. Untuk K-Means, istilah akurasi harus dipahami secara hati-hati karena kualitas nilai tersebut bergantung pada cara pemetaan klaster ke label rujukan. Total bantuan dihitung menggunakan persamaan: Total bantuan = (Rp10.000.000 × n ringan) + (Rp25.000.000 × n sedang) + (Rp50.000.000 × n berat)

C. Hasil dan Pembahasan

1. Perbandingan Kinerja Algoritma

Decision Tree memperoleh nilai tertinggi sebesar 79,17%. Random Forest berada pada urutan kedua dengan 70,83%, sedangkan K-Means menghasilkan 68,33%. Perbandingan nilai tersebut ditampilkan pada Gambar 2. Karena K-Means tidak belajar dari label, nilainya tidak sepenuhnya setara secara metodologis dengan akurasi dua algoritma supervised.



Gambar 2. Perbandingan akurasi atau kesesuaian model

Ringkasan hasil pada Tabel 2 disusun untuk memperlihatkan perbandingan keluaran utama dari ketiga algoritma yang diuji, yaitu nilai akurasi atau kesesuaian model, distribusi jumlah rumah pada setiap kategori kerusakan, serta estimasi kebutuhan bantuan yang dihasilkan. Penyajian dalam bentuk tabel memudahkan identifikasi algoritma yang paling sesuai digunakan sebagai dasar simulasi pengambilan keputusan, sekaligus menunjukkan perbedaan karakteristik antara metode supervised dan unsupervised dalam mengelompokkan tingkat kerusakan rumah akibat banjir.

Tabel 2. Ringkasan hasil tiga algoritma

Algoritma	Akurasi	R/S/B	Estimasi bantuan
Decision Tree	79,17%	30/81/9	Rp2.775.000.000
Random Forest	70,83%	33/80/7	Rp2.680.000.000
K-Means	68,33%	61/59/0	Rp2.085.000.000

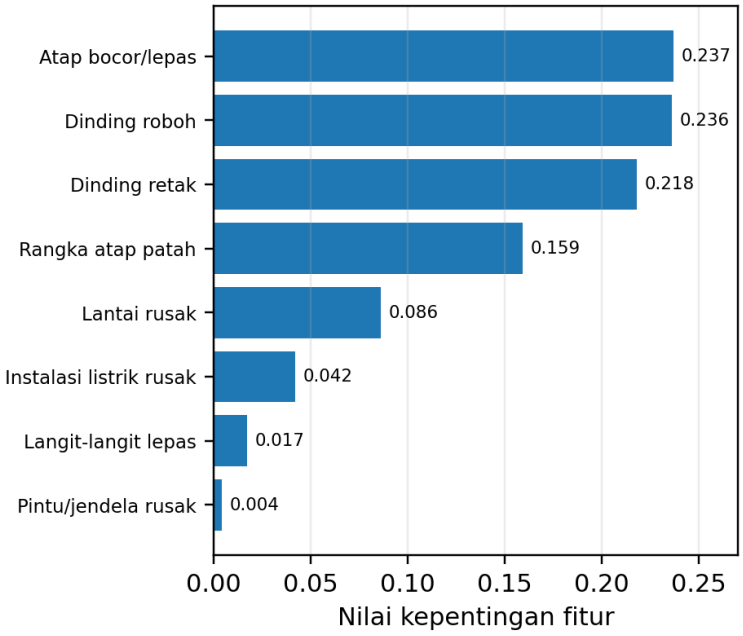
Catatan konsistensi: pada bagian Random Forest, laporan memuat angka 85 rumah rusak sedang pada satu tabel, tetapi grafik, uraian hasil, dan tabel

perbandingan akhir menggunakan 80 rumah. Artikel ini menggunakan 80 karena jumlah 33 + 80 + 7 konsisten dengan total 120 data dan estimasi bantuan Rp2.680.000.000. Laporan juga menampilkan variasi waktu proses pada beberapa bagian; oleh karena itu, waktu komputasi tidak dijadikan dasar utama penentuan model terbaik.

2. Hasil Decision Tree

Decision Tree mengklasifikasikan 30 rumah sebagai rusak ringan, 81 rumah sebagai rusak sedang, dan 9 rumah sebagai rusak berat. Kategori rusak sedang mendominasi sebesar 67,5%. Model ini dinilai paling sesuai pada dataset simulasi karena menghasilkan akurasi tertinggi serta menyediakan struktur pohon yang dapat ditelusuri.

Analisis kepentingan fitur menunjukkan bahwa atap bocor atau lepas memiliki nilai 0,237, dinding roboh 0,236, dan dinding retak 0,218. Ketiga fitur tersebut menjadi indikator utama pada pohon yang terbentuk. Rangka atap patah memiliki nilai 0,159 dan lantai rusak 0,086. Sementara itu, bangunan berdiri, kolom atau balok retak, dan kolom atau balok patah dilaporkan bernilai 0,000 pada model ini, yang berarti tidak digunakan dalam pembentukan pohon pada konfigurasi dan sampel tersebut.



Gambar 3. Kepentingan fitur pada Decision Tree

Dominasi fitur atap dan dinding menunjukkan bahwa komponen yang mudah diamati dapat berperan besar dalam klasifikasi. Namun, nilai kepentingan nol tidak dapat langsung diartikan bahwa suatu fitur tidak penting secara struktural. Kondisi tersebut hanya menunjukkan bahwa fitur tersebut tidak dipilih oleh model pada pembagian data dan parameter yang digunakan.

3. Hasil Random Forest

Random Forest menghasilkan 33 rumah rusak ringan, 80 rumah rusak sedang, dan 7 rumah rusak berat, dengan akurasi 70,83%. Hasil tersebut memperlihatkan pola distribusi yang mendekati Decision Tree, tetapi dengan jumlah rumah rusak berat yang sedikit lebih rendah. Pada dataset yang relatif kecil, pembentukan

banyak pohon belum memberikan peningkatan akurasi dibandingkan satu pohon keputusan.

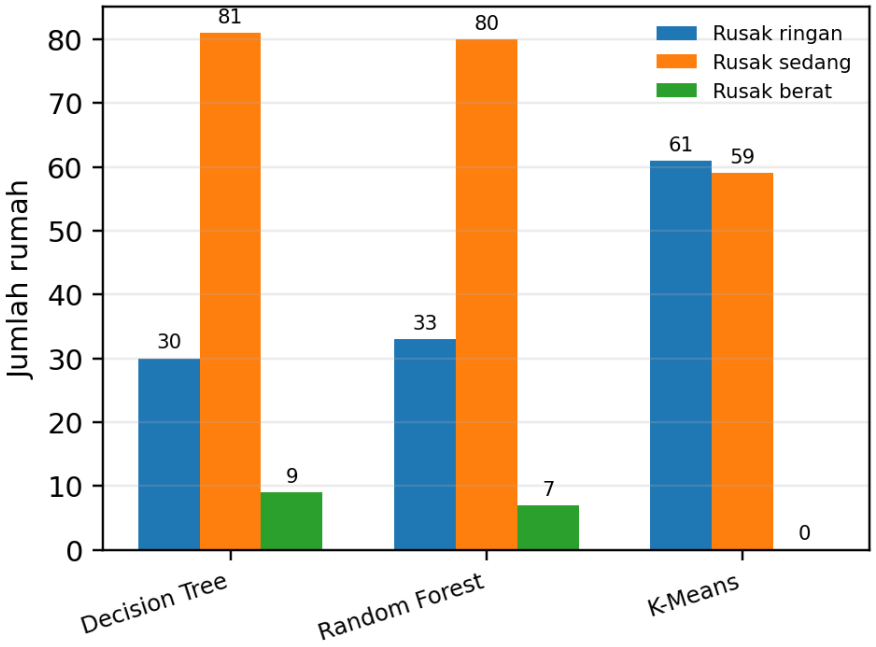
Estimasi bantuan berdasarkan hasil Random Forest adalah Rp2.680.000.000. Nilai ini lebih rendah Rp95.000.000 dibandingkan Decision Tree karena terdapat dua rumah lebih sedikit pada kategori rusak berat dan perubahan kecil pada kategori lainnya. Perbedaan ini memperlihatkan bahwa kesalahan klasifikasi dapat berpengaruh langsung pada estimasi kebutuhan anggaran.

4. **Hasil K-Means**

K-Means membentuk 61 data rusak ringan dan 59 data rusak sedang, tetapi tidak menghasilkan data rusak berat. Nilai kesesuaian yang dilaporkan adalah 68,33%. Kegagalan membentuk kategori berat menunjukkan bahwa pola jarak pada fitur belum cukup memisahkan kelompok kerusakan berat, atau prosedur pemetaan klaster ke tiga kategori belum sesuai.

Estimasi bantuan K-Means sebesar Rp2.085.000.000 merupakan nilai terendah. Nilai tersebut bukan berarti kebutuhan bantuan yang sebenarnya lebih kecil, melainkan konsekuensi langsung dari tidak adanya rumah yang ditempatkan dalam kategori rusak berat. Karena itu, K-Means lebih tepat diposisikan sebagai alat eksplorasi pola atau pembanding, bukan sebagai model utama untuk penetapan kategori bantuan pada konfigurasi data ini.

5. **Distribusi Kerusakan dan Implikasi Penanganan**



Gambar 4. Distribusi kategori kerusakan menurut algoritma

Ketiga algoritma memberikan distribusi yang berbeda. Decision Tree dan Random Forest sama-sama menempatkan sebagian besar rumah pada kategori sedang, sedangkan K-Means membagi data hampir seimbang antara kategori ringan dan sedang. Perbedaan tersebut penting karena setiap kategori dikaitkan dengan nominal bantuan yang berbeda.

Berdasarkan hasil Decision Tree, 9 rumah rusak berat diprioritaskan untuk penanganan awal dengan estimasi kedatangan bantuan 1–3 hari dan target penyelesaian perbaikan sekitar tiga bulan. Sebanyak 81 rumah rusak sedang

diproyeksikan menerima penanganan dalam 1–2 minggu dengan waktu perbaikan sekitar dua bulan, sedangkan 30 rumah rusak ringan dapat ditangani dalam 2–4 minggu dengan waktu perbaikan sekitar satu bulan. Jadwal ini berasal dari saran tindakan pada laporan dan masih memerlukan penyesuaian terhadap kondisi lapangan, akses wilayah, ketersediaan material, serta kapasitas pelaksana.

Tabel 3. Prioritas penanganan berdasarkan hasil Decision Tree

Kategori	Jumlah	Respons	Target perbaikan
Berat	9	1–3 hari	3 bulan
Sedang	81	1–2 minggu	2 bulan
Ringan	30	2–4 minggu	1 bulan

6. **Keterbatasan Penelitian**

Pertama, data yang digunakan adalah data simulasi sehingga belum mencerminkan variasi kondisi bangunan, kualitas survei, dan karakteristik wilayah secara utuh. Kedua, ukuran dataset hanya 120 data dengan 24 data uji, sehingga perubahan beberapa prediksi dapat menghasilkan perubahan persentase yang cukup besar. Ketiga, laporan tidak mendokumentasikan teknik pengkodean setiap fitur, penanganan data hilang, parameter model, random state, validasi silang, confusion matrix, precision, recall, dan F1-score.

Keempat, prosedur pemetaan klaster K-Means ke label kerusakan tidak dijelaskan, sehingga nilai 68,33% tidak dapat direplikasi secara penuh. Kelima, laporan memuat Naive Bayes pada tabel perbandingan akhir, tetapi tidak menyediakan uraian metode dan hasil yang setara dengan tiga algoritma lainnya. Oleh sebab itu, Naive Bayes tidak dimasukkan dalam analisis artikel ini. Keenam, terdapat inkonsistensi jumlah hasil Random Forest dan waktu proses antarbagian yang harus diperbaiki pada eksperimen ulang

D. Simpulan

Pada dataset simulasi 120 rumah terdampak banjir dengan 11 variabel kondisi bangunan, Decision Tree memberikan hasil terbaik dengan akurasi 79,17%. Model menghasilkan 30 rumah rusak ringan, 81 rusak sedang, dan 9 rusak berat. Fitur paling dominan adalah atap bocor atau lepas, dinding roboh, dan dinding retak. Random Forest menghasilkan akurasi 70,83%, sedangkan K-Means menghasilkan nilai kesesuaian 68,33% dan gagal membentuk kategori rusak berat.

Berdasarkan skema bantuan dalam laporan, keluaran Decision Tree menghasilkan estimasi kebutuhan Rp2.775.000.000 dan dapat digunakan sebagai dasar simulasi prioritas penanganan. Meskipun Decision Tree menjadi model paling sesuai pada eksperimen ini, hasil belum dapat digunakan langsung untuk keputusan operasional karena dataset bersifat dummy. Penelitian lanjutan perlu menggunakan data lapangan terverifikasi, mendokumentasikan pembobotan dan pengkodean fitur, menerapkan validasi silang, menambahkan confusion matrix serta metrik per kelas, menangani ketidakseimbangan kelas, dan menguji kestabilan model pada wilayah serta periode bencana yang berbeda

E. Ucapan Terima Kasih

Artikel ini disusun dari laporan tugas mata kuliah Perancangan dan Analisis Algoritma pada Program Studi Informatika, Departemen Elektronika, Fakultas Teknik, Universitas Negeri Padang. Penulis dapat menambahkan nama dosen pengampu, pembimbing, atau pihak lain yang berkontribusi setelah memperoleh persetujuan yang bersangkutan.

F. Referensi

[1]. Kementerian Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia, *Peraturan Menteri PUPR Nomor 29/PRT/M/2018 tentang Standar Teknis Standar Pelayanan Minimal Pekerjaan Umum dan Perumahan Rakyat*. Jakarta, Indonesia, 2018.

[2]. Badan Nasional Penanggulangan Bencana, *Peraturan Kepala BNPB Nomor 2 Tahun 2012 tentang Pedoman Umum Pengkajian Risiko Bencana*. Jakarta, Indonesia, 2012.

[3]. E. Rahmawati, C. E. Widodo, and S. Koesuma, "Machine learning for post-disaster building damage classification and rehabilitation recommendation: A review," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 15, no. 1, 2026, doi: 10.32736/sisfokom.v15i01.2532.

[4]. J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.

[5]. L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[6]. J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Mathematical Statistics and Probability*, vol. 1, Berkeley, CA, USA: University of California Press, 1967, pp. 281–297.

[7]. Pedregosa *et al*., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[8]. S. Hadi, *Sistem E-MTQ Terintegrasi* [Software], 2021. [Online]. Available: https://e-mtq.com/

[9]. Anisha, K. Malavika, K. Erfana, K. Dheeraj, A. Subahan, S. Raj, S. Mangalathu, and R. Davis, "Flood damage assessment using machine-learning methods," 2022.

[10]. S. Al Shafian and D. Hu, "Integrating machine learning and remote sensing in disaster management: A decadal review of post-disaster building damage assessment," *Buildings*, vol. 14, no. 8, Art. no. 2344, 2024, doi: 10.3390/buildings14082344.

[11]. C.-F. Liu, L. Huang, K. Yin, S. Brody, and A. Mostafavi, "FloodDamageCast: Building flood damage machine learning and data augmentation," *arXiv preprint arXiv:2405.14232*, 2024. [Online]. Available: https://arxiv.org/abs/2405.14232

[12]. S. Hadi, "Penerapan IoT pada smart farming," 2025.