Jurnal Sains Informatika Terapan (JSIT)
E-ISSN: 2828-1659, Volume: 5, Issue: 2, Month: July, Year: 2026
Page: 738 - 744, Available online at: https://rcf-indonesia.org/home/

KLASTERISASI K-MEANS UNTUK PENGELOMPOKAN WILAYAH BERISIKO BANJIR DI KOTA PADANG

Teri Ade Putra[1)], Andani Putra Pratama[2)], Pradani Ayu Widya Purnama[3)]
[1,2,3]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia "YPTK" Padang
Corresponding Author: [1] teriadeputra@upiyptk.ac.id

Article Info

Article history:

Received: Jun 14, 2026
Revised: Jun 25, 2026
Accepted: Jul 10, 2026
Published: Jul 30, 2026

Keywords:

K-Means
Klasterisasi
Risiko Banjir
Data Mining
Kota Padang

ABSTRACT

Banjir merupakan salah satu risiko bencana yang berulang kali dihadapi Kota Padang, dipengaruhi oleh kombinasi intensitas curah hujan, kepadatan populasi, histori kejadian banjir, dan elevasi wilayah pada 104 kelurahan yang tersebar di 11 kecamatan. Penelitian ini bertujuan menerapkan algoritma K-Means untuk mengelompokkan kelurahan di Kota Padang berdasarkan tingkat risiko banjir, sekaligus membangun sistem berbasis web bernama SIPEKAT sebagai alat bantu pemetaan dan pengambilan keputusan. Data dihimpun dari empat sumber, yaitu curah hujan dari BMKG Stasiun Teluk Bayur, data kependudukan dari Disdukcapil, histori banjir dari BNPB, dan data elevasi dari dataset geospasial. Seluruh atribut dibersihkan dan dinormalisasi menggunakan metode Min-Max sebelum diproses dengan K-Means. Proses klasterisasi diinisialisasi dari tiga kelurahan representatif sebagai centroid awal dan konvergen pada iterasi kelima, menghasilkan tiga klaster dengan 19, 80, dan 5 anggota kelurahan. Setiap klaster diberi label berdasarkan skor risiko dari rata-rata histori banjir dan elevasi, menghasilkan kategori Risiko Tinggi (19 kelurahan), Risiko Sedang (80 kelurahan), dan Risiko Rendah (5 kelurahan). Hasil clustering divalidasi menggunakan RapidMiner dan menunjukkan kesesuaian dengan perhitungan sistem. Kebaruan penelitian ini terletak pada penggunaan mekanisme pelabelan klaster berbasis skor risiko yang menggabungkan histori kejadian banjir dan elevasi wilayah untuk menghasilkan interpretasi tingkat risiko yang lebih objektif, serta implementasinya ke dalam sistem informasi berbasis web SIPEKAT. Penelitian ini membuktikan bahwa K-Means efektif digunakan untuk mengelompokkan wilayah rawan banjir dan dapat menjadi dasar bagi pemerintah daerah dalam memprioritaskan mitigasi bencana banjir.

1. PENDAHULUAN

Perkembangan teknologi informasi dalam beberapa tahun terakhir berlangsung sangat pesat. Peningkatan pemanfaatan teknologi informasi di Indonesia didorong oleh perluasan infrastruktur digital serta meningkatnya literasi digital masyarakat, sehingga mampu menciptakan efisiensi kerja dan inovasi di berbagai sektor [1]. Transformasi digital berbasis teknologi informasi turut meningkatkan kualitas layanan dan daya saing ekonomi nasional [2]. Dalam konteks pengolahan data, data mining menjadi salah satu teknologi penting karena mampu mengekstraksi pola dan informasi bernilai dari data berjumlah besar untuk mendukung pengambilan keputusan yang lebih akurat, dan penerapannya telah meluas ke berbagai bidang, seperti industri kreatif, bisnis, kesehatan, dan pendidikan [3]–[5].

Salah satu teknik data mining yang paling banyak digunakan untuk pengelompokan (clustering) data adalah algoritma K-Means, karena prosesnya sederhana, efisien, dan mampu menangani data berukuran besar dengan waktu komputasi yang relatif cepat [6]. K-Means telah diterapkan pada berbagai bidang, seperti pendidikan, pemasaran, dan kesehatan, untuk menemukan pola tersembunyi yang mendukung pengambilan keputusan [7]. Meskipun demikian, K-Means memiliki keterbatasan, terutama dalam penentuan jumlah klaster awal dan sensitivitas terhadap nilai centroid awal, sehingga diperlukan evaluasi dan pengujian yang tepat agar hasil klasterisasi lebih optimal [8].

Hasil klasterisasi K-Means dapat memberikan gambaran pola perilaku data yang dijadikan dasar analisis prediktif, khususnya pada data berskala besar [9], sehingga algoritma ini berperan penting sebagai alat pendukung dalam memprediksi informasi di masa depan melalui pemanfaatan pola data historis yang telah dikelompokkan [10].

Jurnal Sains Informatika Terapan (JSIT)
E-ISSN: 2828-1659, Volume: 5, Issue: 2, Month: July, Year: 2026
Page: 738 - 744, Available online at: https://rcf-indonesia.org/home/

Pada kasus bencana banjir, K-Means dapat mengelompokkan wilayah berdasarkan tingkat kerawanan menggunakan data historis dan parameter lingkungan, seperti curah hujan, ketinggian wilayah, kepadatan penduduk, dan kondisi drainase, sehingga membantu pemetaan risiko secara lebih sistematis [11]. Hasil klasterisasi K-Means dapat digunakan sebagai dasar perencanaan mitigasi bencana karena mampu memberikan gambaran tingkat risiko banjir pada masing-masing klaster wilayah [12], dan penggunaannya dalam pemetaan kerawanan banjir dapat meningkatkan akurasi pengambilan keputusan oleh pemerintah daerah dalam upaya pencegahan dan penanggulangan bencana [13].

Kota Padang, sebagai ibu kota Provinsi Sumatera Barat, memiliki karakteristik wilayah yang kompleks karena terdiri atas kawasan pesisir, dataran rendah, hingga perbukitan yang dilalui beberapa sungai besar, seperti Batang Arau, Batang Kuranji, dan Batang Air Dingin, sehingga menjadikannya rentan terhadap bencana banjir, terutama pada musim hujan dengan intensitas tinggi. Berdasarkan data Badan Penanggulangan Bencana Daerah Provinsi Sumatera Barat, sepanjang tahun 2024 total kerugian sementara akibat bencana alam di provinsi ini mencapai sekitar Rp108,38 miliar, mencakup kerusakan infrastruktur, rumah warga, serta lahan pertanian yang terdampak banjir dan longsor di berbagai kabupaten/kota. Kondisi ini menunjukkan perlunya pemetaan wilayah berisiko banjir yang lebih sistematis dan berbasis data sebagai dasar prioritas mitigasi.

Berbagai penelitian sebelumnya telah memanfaatkan algoritma K-Means untuk melakukan pemetaan wilayah rawan banjir berdasarkan karakteristik lingkungan dan histori kejadian bencana. Namun, sebagian besar penelitian tersebut hanya berfokus pada proses pengelompokan data tanpa memberikan mekanisme pelabelan risiko yang objektif terhadap klaster yang dihasilkan. Selain itu, hasil klasterisasi umumnya masih disajikan dalam bentuk laporan analisis dan belum terintegrasi ke dalam sistem berbasis web yang dapat digunakan sebagai alat bantu pengambilan keputusan oleh pemerintah daerah. Oleh karena itu, diperlukan suatu pendekatan yang tidak hanya mampu melakukan pengelompokan wilayah berisiko banjir, tetapi juga memberikan interpretasi tingkat risiko yang lebih jelas dan mudah dipahami oleh pemangku kepentingan.

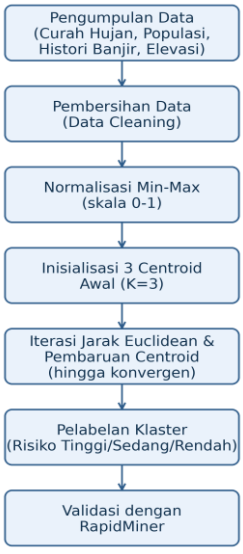
Kebaruan penelitian ini terletak pada penerapan mekanisme pelabelan klaster menggunakan skor risiko yang dihitung berdasarkan kombinasi rata-rata histori kejadian banjir dan elevasi wilayah, serta implementasi hasil klasterisasi ke dalam sistem informasi berbasis web bernama SIPEKAT yang dapat digunakan untuk mendukung proses pemetaan dan penentuan prioritas mitigasi bencana banjir di Kota Padang.

Dalam penelitian ini, histori kejadian banjir dan elevasi wilayah dipilih sebagai dasar pelabelan risiko karena kedua atribut tersebut merupakan indikator yang paling merepresentasikan tingkat kerawanan banjir suatu wilayah. Histori banjir menggambarkan frekuensi kejadian yang pernah terjadi, sedangkan elevasi menunjukkan kerentanan wilayah terhadap genangan akibat posisi topografi yang lebih rendah dibandingkan wilayah sekitarnya.

Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan algoritma K-Means untuk mengelompokkan 104 kelurahan pada 11 kecamatan di Kota Padang ke dalam tiga tingkat risiko banjir berdasarkan empat atribut, yaitu curah hujan, jumlah penduduk, histori kejadian banjir, dan elevasi wilayah. Penelitian ini juga mencakup verifikasi hasil klasterisasi menggunakan perangkat lunak RapidMiner untuk memastikan kesesuaian implementasi algoritma serta pengembangan sistem berbasis web SIPEKAT sebagai alat bantu pemetaan dan pengambilan keputusan mitigasi banjir. Hasil penelitian ini juga diwujudkan dalam bentuk sistem berbasis web bernama SIPEKAT sebagai alat bantu pemetaan dan pengambilan keputusan bagi Pemerintah Kota Padang dalam upaya mitigasi bencana banjir.

2. METODE PENELITIAN

Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan 104 kelurahan pada 11 kecamatan di Kota Padang berdasarkan empat atribut data. Tahapan penelitian meliputi pengumpulan data, pembersihan data (data cleaning), normalisasi Min-Max, inisialisasi centroid, iterasi perhitungan jarak, pembaruan centroid hingga konvergen, pelabelan klaster, serta implementasi dan validasi sistem, sebagaimana digambarkan pada kerangka penelitian pada Gambar 1.



Gambar 1. Kerangka tahapan penelitian penerapan algoritma K-Means

Jurnal Sains Informatika Terapan (JSIT)
E-ISSN: 2828-1659, Volume: 5, Issue: 2, Month: July, Year: 2026
Page: 738 - 744, Available online at: https://rcf-indonesia.org/home/

K-Means dipilih karena termasuk pendekatan unsupervised learning yang mampu menemukan pola pengelompokan tanpa memerlukan data berlabel [14], [15], berbeda dengan pendekatan supervised learning yang telah banyak diterapkan pada tugas klasifikasi, seperti deteksi risiko penyakit jantung menggunakan K-Nearest Neighbor [16] atau klasifikasi kanker paru-paru [17], maupun reinforcement learning yang berkembang pesat pada sistem pengambilan keputusan otomatis [18]. Kinerja model klasifikasi umumnya dievaluasi menggunakan metrik seperti akurasi, presisi, dan recall [19], sedangkan kualitas hasil clustering K-Means dapat dievaluasi menggunakan pendekatan lain, seperti Metode Elbow [20] atau Silhouette Index dan Davies-Bouldin Index [21]; pendekatan-pendekatan tersebut belum digunakan pada penelitian ini karena jumlah klaster (K=3) ditentukan berdasarkan kebutuhan tiga tingkat risiko yang umum dipahami masyarakat, sehingga menjadi salah satu batasan penelitian yang dibahas lebih lanjut pada bagian pembahasan.

2.1 Pengumpulan dan Pembersihan Data

Data penelitian merupakan data sekunder yang dihimpun dari empat sumber resmi, mencakup seluruh 104 kelurahan pada 11 kecamatan di Kota Padang, sebagaimana dirangkum pada Tabel 1.

Tabel 1. Ringkasan Atribut Data Penelitian

Atribut	Sumber Data	Min	Maks
Curah Hujan (mm)	BMKG Teluk Bayur	1.159,2	1.159,2
Penduduk (jiwa)	Disdukcapil Padang	1.313	36.453
Histori Banjir	BNPB	0	5
Elevasi (mdpl)	Data Geospasial	-3	696

Data curah hujan berbentuk catatan harian dengan simbol khusus ("-" berarti tidak hujan dan dianggap 0 mm, "TTU" berarti tidak terukur dan juga dianggap mendekati 0 mm) yang dijumlahkan menjadi total bulanan; nilai bulan November 2025 (1.159,2 mm) digunakan karena merupakan bulan terjadinya banjir besar yang menjadi objek penelitian, dan nilai ini berlaku seragam untuk seluruh kelurahan karena data BMKG bersifat satu titik observasi kota. Data curah hujan diperoleh dari satu titik pengamatan BMKG Kota Padang sehingga memiliki nilai yang sama pada seluruh kelurahan. Oleh karena itu, atribut ini digunakan sebagai variabel kontekstual yang menggambarkan kondisi meteorologis saat kejadian banjir, bukan sebagai faktor utama pembeda antarwilayah dalam proses klasterisasi. Data kependudukan dari Disdukcapil dipilah terlebih dahulu untuk mengambil baris tingkat kelurahan saja agar tidak terjadi penghitungan ganda dengan tingkat kecamatan dan kota, kemudian divalidasi dengan memastikan jumlah penduduk laki-laki ditambah perempuan sama dengan jumlah total. Data histori banjir dari BNPB dan pemberitaan resmi Pemerintah Kota Padang, yang semula berbentuk narasi kejadian, dikonversi menjadi jumlah kejadian per kelurahan dengan menjumlahkan setiap peristiwa (banjir maupun longsor) yang tercatat menimpa kelurahan tersebut. Data elevasi diperoleh dari titik koordinat representatif tiap kelurahan pada dataset geospasial. Proses pembersihan data ini penting dilakukan karena data mentah yang tidak diproses dengan baik berpotensi menurunkan akurasi model [22], [23], termasuk penanganan nilai yang hilang atau tidak konsisten sebelum data digunakan pada tahap pemodelan [24].

2.2 Normalisasi Data

Keempat atribut yang digunakan memiliki satuan dan skala yang berbeda; misalnya atribut populasi berkisar ribuan hingga puluhan ribu jiwa, sedangkan atribut histori banjir hanya berkisar 0 hingga 5 kejadian. Apabila dibandingkan langsung tanpa penyesuaian skala, atribut dengan rentang nilai lebih besar akan mendominasi proses klasterisasi secara tidak proporsional. Oleh karena itu, seluruh atribut dinormalisasi terlebih dahulu ke rentang 0 sampai 1 menggunakan metode Min-Max Normalization pada Persamaan (1).

$$X' = (X - Xmin) / (Xmaks - Xmin) \quad (1)$$

dengan X' adalah nilai hasil normalisasi, X adalah nilai data asli, Xmin adalah nilai terendah, dan Xmaks adalah nilai tertinggi pada atribut yang sama. Melalui persamaan ini, nilai terendah pada suatu atribut akan menghasilkan angka 0, nilai tertinggi menghasilkan angka 1, dan nilai lainnya berada pada rentang di antaranya.

2.3 Inisialisasi Centroid

Penelitian ini menetapkan jumlah klaster K=3, merepresentasikan tiga tingkat risiko banjir, yaitu Risiko Tinggi, Risiko Sedang, dan Risiko Rendah. Ketiga titik pusat awal (centroid) tidak dipilih secara acak, melainkan diambil dari data kelurahan aktual dengan aturan pemilihan yang berbeda untuk masing-masing centroid: Centroid 1 diambil dari kelurahan dengan nilai histori banjir tertinggi sebagai wakil awal Risiko Tinggi, Centroid 2 diambil dari kelurahan pada posisi median data sebagai wakil awal Risiko Sedang, dan Centroid 3 diambil dari kelurahan dengan nilai elevasi tertinggi sebagai wakil awal Risiko Rendah.

2.4 Algoritma K-Means

Proses klasterisasi K-Means dilakukan secara iteratif melalui tahapan: (1) menghitung jarak setiap kelurahan terhadap masing-masing centroid menggunakan jarak Euclidean pada Persamaan (2); (2) menempatkan setiap kelurahan ke klaster dengan centroid terdekat; (3) memperbarui posisi centroid berdasarkan rata-rata anggota baru pada Persamaan (3); dan (4) mengulang proses tersebut hingga tidak terjadi lagi perubahan posisi centroid maupun anggota klaster (konvergen).

$$Di = \sqrt{[(xi1-c1)^2+(xi2-c2)^2+(xi3-c3)^2+(xi4-c4)^2]} \quad (2)$$

Jurnal Sains Informatika Terapan (JSIT)
E-ISSN: 2828-1659, Volume: 5, Issue: 2, Month: July, Year: 2026
Page: 738 - 744, Available online at: https://rcf-indonesia.org/home/

dengan Di adalah jarak kelurahan ke-i terhadap suatu centroid, xi1 hingga xi4 adalah nilai atribut (curah hujan, populasi, histori banjir, elevasi) kelurahan ke-i yang telah dinormalisasi, dan c1 hingga c4 adalah nilai atribut centroid yang bersangkutan.

$$cj(baru) = (1/nj) \Sigma xi \quad (3)$$

dengan cj(baru) adalah posisi centroid ke-j yang diperbarui dan nj adalah jumlah anggota klaster ke-j.

2.5 Pelabelan Klaster

Karena K-Means hanya mengenali klaster berdasarkan nomor urut tanpa mengetahui makna risikonya, dilakukan tahap pelabelan menggunakan skor risiko pada Persamaan (4).

$$Skor = RataHistoriBanjir - (RataElevasi/1000) \quad (4)$$

Klaster dengan skor tertinggi diberi label "Risiko Tinggi", skor menengah diberi label "Risiko Sedang", dan skor terendah diberi label "Risiko Rendah". Faktor pembagi 1000 digunakan sebagai skema normalisasi sederhana untuk menyetarakan skala elevasi yang memiliki rentang ratusan meter dengan variabel histori banjir yang hanya memiliki rentang 0–5 kejadian. Dengan pendekatan ini, kedua atribut dapat berkontribusi secara proporsional terhadap pembentukan skor risiko tanpa menyebabkan dominasi salah satu atribut.

2.6 Implementasi dan Validasi Sistem

Hasil perhitungan K-Means diimplementasikan ke dalam sistem berbasis web bernama SIPEKAT menggunakan bahasa pemrograman PHP berorientasi objek dan basis data MySQL, yang menyajikan hasil klasterisasi dalam bentuk peta risiko banjir. Untuk memastikan keakuratan implementasi, hasil clustering sistem divalidasi dengan membandingkannya terhadap hasil clustering pada perangkat lunak data mining RapidMiner menggunakan jumlah klaster yang sama, mengikuti pendekatan validasi berbasis RapidMiner yang juga digunakan pada penelitian klasifikasi data lain [25].

3. HASIL DAN PEMBAHASAN

3.1 Normalisasi dan Inisialisasi Centroid

Sebagai ilustrasi penerapan Persamaan (1), atribut populasi memiliki nilai terendah 1.313 jiwa (Kelurahan Belakang Pondok) dan nilai tertinggi 36.453 jiwa (Kelurahan Kuranji). Tabel 2 menunjukkan contoh hasil normalisasi atribut populasi pada 10 kelurahan sampel.

Tabel 2. Contoh Normalisasi Atribut Populasi (10 Kelurahan Sampel)

Kelurahan	Populasi Asli	Ternormalisasi
Lambung Bukit	4.439	0,0890
Surau Gadang	20.102	0,5347
Lubuk Minturun	12.393	0,3153
Pasie Nan Tigo	11.974	0,3034
Indarung	10.716	0,2676
Batu Gadang	9.439	0,2312
Limau Manis	6.089	0,1359
Limau Manis Selatan	10.993	0,2755
Belakang Pondok	1.313	0,0000
Kuranji	36.453	1,0000

Atribut curah hujan menghasilkan nilai normalisasi 0 pada seluruh kelurahan karena nilai minimum dan maksimumnya sama (data berasal dari satu titik observasi kota), sehingga atribut ini secara matematis tidak memengaruhi proses pemisahan klaster meskipun tetap relevan secara kontekstual sebagai faktor pemicu banjir. Berdasarkan aturan pemilihan pada Sub-bab 2.3, diperoleh titik centroid awal sebagaimana dirangkum pada Tabel 3.

Tabel 3. Titik Acuan Awal (Centroid) Setelah Normalisasi

Centroid	CH	Pop.	H.Banjir	Elevasi
C1	0,0000	0,6049	1,0000	0,0100
C2	0,0000	0,2005	0,6000	0,0172
C3	0,0000	0,2312	0,0000	1,0000

Kondisi ini menunjukkan bahwa atribut curah hujan tidak memberikan kontribusi matematis terhadap proses pemisahan klaster karena seluruh objek memiliki nilai yang sama. Namun demikian, atribut tersebut tetap dipertahankan dalam penelitian karena secara konseptual curah hujan merupakan faktor utama pemicu terjadinya banjir. Dengan demikian, atribut ini masih memiliki nilai informatif dalam menjelaskan konteks kejadian banjir meskipun tidak memengaruhi proses pembentukan klaster secara langsung.

3.2 Proses Iterasi K-Means

Proses iterasi dilakukan terhadap seluruh 104 kelurahan pada setiap putaran, dengan menghitung jarak setiap kelurahan ke tiga centroid menggunakan Persamaan (2), mengelompokkan kelurahan ke klaster terdekat, kemudian memperbarui posisi centroid menggunakan Persamaan (3). Ringkasan jumlah anggota klaster dan besar pergeseran centroid pada setiap iterasi ditunjukkan pada Tabel 4.

Tabel 4. Ringkasan Iterasi K-Means hingga Konvergen

Iterasi	C1	C2	C3	Pergeseran
1	4	95	5	1,10662
2	12	87	5	0,31382
3	18	81	5	0,18396
4	19	80	5	0,02351
5	19	80	5	0,00000

Pada iterasi pertama, terbentuk susunan sementara 4 kelurahan pada Klaster 1, 95 kelurahan pada Klaster

Jurnal Sains Informatika Terapan (JSIT)
E-ISSN: 2828-1659, Volume: 5, Issue: 2, Month: July, Year: 2026
Page: 738 - 744, Available online at: https://rcf-indonesia.org/home/

2, dan 5 kelurahan pada Klaster 3, dengan pergeseran centroid yang masih besar (1,10662). Pembaruan posisi centroid pada iterasi berikutnya menyebabkan sejumlah kelurahan berpindah klaster, sehingga susunan berubah menjadi 12–87–5 pada iterasi kedua dan 18–81–5 pada iterasi ketiga, dengan pergeseran yang terus mengecil (0,31382 dan 0,18396). Pada iterasi keempat susunan menjadi 19–80–5 dengan pergeseran 0,02351, dan pada iterasi kelima diperoleh susunan yang identik (19–80–5) dengan pergeseran 0,00000, yang menandakan proses telah mencapai kondisi konvergen. Dengan demikian, hasil pengelompokan pada iterasi kelima ditetapkan sebagai hasil akhir klasterisasi.

3.3 Karakteristik dan Pelabelan Klaster

Setelah proses iterasi konvergen, diperoleh tiga klaster akhir dengan karakteristik rata-rata atribut sebagaimana disajikan pada Tabel 5.

Tabel 5. Karakteristik Rata-Rata Setiap Klaster

Klaster	n	Rt.Pop.	Rt.HB	Rt.Elv
1	19	10.166	3,53	38,1
2	80	9.088	0,19	27,2
3	5	8.335	0,40	426,4

Pelabelan klaster dilakukan menggunakan skor risiko pada Persamaan (4), dengan hasil perhitungan dan label yang diperoleh ditunjukkan pada Tabel 6.

Tabel 6. Skor dan Label Risiko Setiap Klaster

Klaster	Skor	Label Risiko
1	3,4882	Risiko Tinggi
2	0,1603	Risiko Sedang
3	-0,0264	Risiko Rendah

Klaster 1 memperoleh skor tertinggi (3,4882) sehingga diberi label Risiko Tinggi, beranggotakan 19 kelurahan, di antaranya Dadok Tunggul Hitam, Lubuk Minturun, Batipuh Panjang, Pasie Nan Tigo, Koto Panjang Ikua Koto, Aie Pacah, Surau Gadang, Gunung Pangilun, dan Ulak Karang Selatan. Klaster 3 memperoleh skor terendah dan bernilai negatif (-0,0264) karena rata-rata elevasinya jauh lebih tinggi dibandingkan rata-rata histori banjirnya, sehingga diberi label Risiko Rendah, beranggotakan 5 kelurahan, yaitu Indarung, Batu Gadang, Limau Manis, Limau Manis Selatan, dan Lambung Bukit — seluruhnya berada di kawasan perbukitan pada Kecamatan Lubuk Kilangan dan Pauh. Klaster 2 dengan skor di antara keduanya (0,1603) diberi label Risiko Sedang, beranggotakan 80 kelurahan lainnya yang tersebar di seluruh Kota Padang.

3.4 Validasi dan Implementasi Sistem

Hasil verifikasi menunjukkan bahwa seluruh keluaran sistem SIPEKAT, meliputi jumlah iterasi, jumlah anggota tiap klaster, susunan kelurahan pada kategori Risiko Tinggi dan Risiko Rendah, serta nilai skor pelabelan, identik dengan hasil perhitungan manual, sehingga tidak ditemukan kesalahan logika maupun pembulatan yang signifikan pada implementasinya. Proses clustering pada RapidMiner dengan jumlah klaster yang sama juga menghasilkan pengelompokan yang konsisten dengan hasil sistem, sehingga implementasi algoritma K-Means pada penelitian ini dinyatakan valid.

Selain validasi menggunakan RapidMiner, kualitas hasil klasterisasi dapat dievaluasi menggunakan ukuran internal clustering seperti Silhouette Index dan Davies-Bouldin Index (DBI). Pada penelitian ini evaluasi tersebut belum dilakukan karena fokus penelitian diarahkan pada implementasi sistem dan interpretasi hasil klasterisasi untuk kebutuhan mitigasi bencana. Meskipun demikian, penggunaan metrik evaluasi clustering tersebut direkomendasikan pada penelitian selanjutnya untuk memperoleh ukuran kuantitatif mengenai tingkat kualitas pemisahan antar klaster yang dihasilkan.

3.5 Pembahasan

Hasil penelitian ini sejalan dengan temuan penelitian sebelumnya yang menunjukkan bahwa algoritma K-Means efektif digunakan untuk memetakan wilayah rawan banjir berdasarkan kesamaan pola data hidrologi [11], serta dapat dijadikan dasar objektif dalam perencanaan mitigasi bencana [12]. Konsistensi hasil antara sistem berbasis web dan RapidMiner juga memperkuat temuan bahwa K-Means dapat diterapkan secara andal pada berbagai domain permasalahan, sebagaimana ditunjukkan pada penelitian segmentasi performa pembalap [26], segmentasi pelanggan retail [27], serta perbandingannya dengan algoritma hierarchical clustering [28] dan DBSCAN [29] pada domain lain.

Dibandingkan dengan penelitian Rahman dan Fauzi [11] serta Utami dan Prasetya [12], penelitian ini memiliki beberapa perbedaan. Penelitian sebelumnya berfokus pada proses identifikasi dan pemetaan wilayah rawan banjir menggunakan algoritma K-Means, sedangkan penelitian ini menambahkan mekanisme pelabelan risiko berbasis skor objektif yang dihitung dari karakteristik masing-masing klaster. Selain itu, hasil penelitian tidak hanya disajikan dalam bentuk analisis data, tetapi juga diimplementasikan ke dalam sistem berbasis web SIPEKAT yang dapat digunakan sebagai sarana visualisasi dan pendukung pengambilan keputusan oleh pemerintah daerah. Dengan demikian, penelitian ini memberikan kontribusi tambahan pada aspek interpretasi hasil klasterisasi dan implementasi praktisnya.

Meskipun demikian, penelitian ini memiliki sejumlah keterbatasan. Pertama, jumlah klaster (K=3) ditentukan secara manual berdasarkan kebutuhan tiga tingkat risiko yang umum dipahami masyarakat, bukan melalui evaluasi kuantitatif seperti Metode Elbow [20]

Jurnal Sains Informatika Terapan (JSIT)
E-ISSN: 2828-1659, Volume: 5, Issue: 2, Month: July, Year: 2026
Page: 738 - 744, Available online at: https://rcf-indonesia.org/home/

atau Silhouette Index dan Davies-Bouldin Index [21], sehingga penentuan jumlah klaster yang lebih optimal masih perlu dikaji pada penelitian selanjutnya, sebagaimana juga disarankan pada penelitian klasterisasi kesiapan siswa dalam pemilihan jurusan [30]. Kedua, atribut curah hujan yang digunakan bersifat seragam untuk seluruh kelurahan karena bersumber dari satu titik observasi BMKG, sehingga atribut ini belum memberikan variansi yang berpengaruh terhadap pemisahan klaster meskipun tetap relevan secara kontekstual. Integrasi data curah hujan per kelurahan secara real-time pada penelitian mendatang berpotensi meningkatkan akurasi dan relevansi hasil klasterisasi risiko banjir di Kota Padang. Keterbatasan lain dalam penelitian ini adalah belum digunakannya atribut hidrologi lain seperti kapasitas drainase, penggunaan lahan, jarak terhadap sungai, dan kemiringan lereng yang berpotensi memengaruhi tingkat kerawanan banjir. Penambahan atribut-atribut tersebut pada penelitian selanjutnya diharapkan dapat menghasilkan model klasterisasi yang lebih representatif terhadap kondisi nyata di lapangan.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan beberapa hal sebagai berikut. Pertama, algoritma K-Means dapat diterapkan untuk mengelompokkan 104 kelurahan pada 11 kecamatan di Kota Padang berdasarkan empat atribut, yaitu curah hujan, jumlah penduduk, histori kejadian banjir, dan elevasi wilayah, melalui tahapan normalisasi Min-Max, inisialisasi tiga titik centroid awal, perhitungan jarak Euclidean, serta pembaruan centroid secara iteratif hingga mencapai kondisi konvergen pada iterasi kelima. Kedua, hasil klasterisasi menghasilkan tiga kategori risiko banjir, yaitu Risiko Tinggi (19 kelurahan), Risiko Sedang (80 kelurahan), dan Risiko Rendah (5 kelurahan), yang diperoleh melalui pelabelan klaster berbasis skor risiko dari rata-rata histori banjir dan elevasi tiap klaster. Ketiga, sistem berbasis web SIPEKAT yang dibangun menggunakan PHP dan MySQL terbukti mampu mereplikasi hasil perhitungan manual algoritma K-Means secara akurat, dan hasilnya konsisten dengan validasi menggunakan RapidMiner. Dengan demikian, algoritma K-Means terbukti efektif digunakan untuk mengelompokkan wilayah rawan banjir di Kota Padang dan dapat dijadikan dasar objektif bagi pemerintah daerah dalam memprioritaskan upaya mitigasi bencana banjir. Kebaruan penelitian ini terletak pada penerapan skema pelabelan risiko berbasis skor klaster dan implementasi hasil klasterisasi ke dalam sistem SIPEKAT yang memungkinkan hasil analisis dimanfaatkan secara langsung dalam proses pengambilan keputusan terkait mitigasi bencana banjir.

Penelitian selanjutnya disarankan menggunakan metode penentuan jumlah klaster yang lebih objektif, seperti Metode Elbow atau Silhouette Index, serta mengintegrasikan data curah hujan per wilayah secara real-time untuk meningkatkan akurasi hasil klasterisasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer, Universitas Putra Indonesia “YPTK” Padang, serta kepada BMKG Stasiun Meteorologi Maritim Teluk Bayur, Disdukcapil Kota Padang, dan BNPB Kota Padang atas data yang telah disediakan untuk mendukung penelitian ini.

DAFTAR PUSTAKA

[1] A. Nugroho and D. R. Putri, “Perkembangan teknologi informasi dan dampaknya terhadap masyarakat di Indonesia,” Jurnal Teknologi Informasi dan Komunikasi, vol. 8, no. 2, pp. 45–56, 2021.

[2] M. P. Sari and R. A. Wibowo, “Transformasi digital dan peran teknologi informasi dalam pembangunan ekonomi Indonesia,” Jurnal Sistem Informasi Nasional, vol. 12, no. 1, pp. 23–34, 2023.

[3] E. Prasetyo and F. Nugraha, “Penerapan data mining dalam pengolahan data untuk mendukung pengambilan keputusan,” Jurnal Informatika dan Komputer, vol. 5, no. 1, pp. 11–20, 2020.

[4] L. Hakim, A. Peryanto, D. Susanto, and Y. F. Widodo, “Meninjau peranan text mining sebagai alat strategis dalam industri kreatif melalui sajian webinar,” Jurnal Pengabdian Masyarakat (PIMAS), vol. 4, no. 3, pp. 225–233, 2025, doi: 10.35960/pimas.v1i2.1868.

[5] E. Saputri, “Teknik dan aplikasi data mining di Indonesia: Tinjauan literatur satu dekade (2015–2024),” IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi, vol. 4, no. 2, pp. 138–149, 2025, doi: 10.24246/itexplore.v4i2.2025.pp138-149.

[6] E. Prasetyo and F. Nugraha, “Analisis penerapan algoritma K-Means dalam pengelompokan data,” Jurnal Informatika dan Komputer, vol. 5, no. 2, pp. 34–42, 2020.

[7] A. Ramadhan and H. Susanto, “Implementasi algoritma K-Means untuk clustering data pada berbagai bidang,” Jurnal Sistem Informasi, vol. 9, no. 1, pp. 15–25, 2021.

[8] D. Lestari and R. Wahyudi, “Evaluasi algoritma K-Means dalam data mining dan permasalahannya,” Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 10, no. 3, pp. 201–210, 2023.

[9] R. Hidayat and A. Pramudita, “Analisis klasterisasi data menggunakan algoritma K-Means untuk mendukung pengambilan keputusan,” Jurnal Ilmiah Teknologi Informasi, vol. 7, no. 2, pp. 101–110, 2021.

[10] S. Yuliana and B. Setiawan, “Penerapan algoritma K-Means dalam analisis pola data sebagai dasar prediksi informasi,” Jurnal Informatika dan Sistem Informasi, vol. 13, no. 1, pp. 55–64, 2024.

[11] F. Rahman and A. Fauzi, “Penerapan algoritma K-Means untuk pemetaan daerah rawan banjir,” Jurnal Geoinformatika Indonesia, vol. 5, no. 2, pp. 67–76, 2020.

[12] D. R. Utami and E. Prasetya, “Analisis klasterisasi wilayah rawan banjir menggunakan metode K-Means,” Jurnal Sistem Informasi Geografis, vol. 11, no. 1, pp. 45–54, 2022.

[13] R. Maulana and S. Hartono, “Pemanfaatan data mining algoritma K-Means dalam mitigasi bencana banjir,” Jurnal Teknologi Informasi dan Kebencanaan, vol. 4, no. 1, pp. 1–10, 2024.

[14] R. S. Nurhalizah, R. Ardianto, and Purwono, “Analisis supervised dan unsupervised learning pada machine learning: Systematic literature review,” Jurnal Ilmu Komputer dan Informatika (JIKI), vol. 4, no. 1, pp. 61–72, 2024, doi: 10.54082/jiki.168.

[15] H. Abijono, P. Santoso, and N. L. Anggreini, “Supervised learning and unsupervised learning algorithm in data processing,”

G-Tech: Jurnal Teknologi Terapan, vol. 4, no. 2, pp. 315–318, 2021, doi: 10.33379/gtech.v4i2.635.

[16] A. Dari and I. Fajri, "Penerapan algoritma K-Nearest Neighbor (KNN) untuk klasifikasi resiko penyakit jantung," Journal of Information System Research (JOSH), vol. 6, no. 1, pp. 428–436, 2024, doi: 10.47065/josh.v6i1.6038.

[17] Z. Hadiansyah, Z. Rozikin, and M. Fatchan, "Implementasi algoritma K-Nearest Neighbor dalam klasifikasi penyakit kanker paru-paru," Journal of Computer System and Informatics (JoSYC), vol. 6, no. 1, pp. 96–106, 2024, doi: 10.47065/josyc.v6i1.6195.

[18] I. Permata Sari, W. Kurnia, and N. Hendrastuty, "Deep reinforcement learning: Perkembangan dan aplikasi terkini," Jurnal Teknologi Pintar, vol. 4, no. 1, 2024.

[19] A. F. Masruriyah, H. Y. Novita, C. E. Sukmawati, A. R. Ramadhan, S. N. N. Arif, and B. A. Dermawan, "Pengukuran kinerja model klasifikasi dengan data oversampling pada algoritma supervised learning untuk penyakit jantung," Computer Science (CO-SCIENCE), vol. 4, no. 1, pp. 62–70, 2024, doi: 10.31294/coscience.v4i1.2389.

[20] N. A. Maori and E. Evanita, "Metode Elbow dalam optimasi jumlah cluster pada K-Means clustering," Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.

[21] M. R. Syahkur, D. Hartama, and S. Solikhun, "Evaluasi jumlah cluster pada algoritma K-Means++ menggunakan Silhouette dan Elbow dengan validasi nilai DBI dalam mengelompokkan gizi balita," JST (Jurnal Sains dan Teknologi), vol. 13, no. 3, pp. 487–496, 2024, doi: 10.23887/jstundiksha.v13i3.86419.

[22] A. A. A. Daniswara and I. K. D. Nuryana, "Data preprocessing pola pada penilaian mahasiswa program profesi guru," JINACS: Journal of Informatics and Computer Science, vol. 5, no. 1, pp. 97–100, 2023, doi: 10.26740/jinacs.v5n01.p97-100.

[23] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing data dan klasifikasi untuk prediksi kinerja akademik siswa," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.

[24] M. R. A. Prasetyo, A. M. Priyatno, and Nurhaeni, "Penanganan imputasi missing values pada data time series dengan menggunakan metode data mining," Jurnal Informasi dan Teknologi, vol. 5, no. 2, pp. 52–62, 2023, doi: 10.37034/jidt.v5i2.324.

[25] A. Sifaunajah and R. D. Wahyuningtyas, "The classification of prospective students with RapidMiner," NEWTON: Networking and Information Technology, vol. 2, no. 2, pp. 72–78, 2022, doi: 10.32764/newton.v2i2.2681.

[26] M. Sahira, A. Salsabila, S. Salsabila, A. N. Putri, K. D. Tania, and W. K. Sari, "Penerapan metode K-Means clustering untuk segmentasi performa pembalap F1 season 2024," Infomatek, vol. 27, no. 1, pp. 113–122, 2025, doi: 10.23969/infomatek.v27i1.24297.

[27] R. D. Nugraha, D. D. Adelia, and D. Rivaldi, "Segmentasi pelanggan retail berbasis perilaku menggunakan algoritma K-Means clustering," Digital Transformation Technology (Digitech), vol. 5, no. 2, pp. 141–148, 2025, doi: 10.47709/digitech.v5i2.6340.

[28] A. A. R. Mulyana, A. R. Juwita, A. M. Siregar, and A. Fauzi, "Penerapan algoritma K-means clustering dan hierarchical clustering dalam mengelompokkan data pengangguran di Karawang," Jurnal Algoritma, vol. 21, no. 2, pp. 209–220, 2024.

[29] M. R. P. Pratama, M. I. Fieldi, M. S. Albani, M. Al Fachrozi, F. R. Aderiyana, K. D. Tania, and A. Meiriza, "Perbandingan algoritma K-Means, K-Medoid, dan DBSCAN untuk clustering kualitas hidup Indonesia dalam perspektif knowledge management dan data discovery," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 9, no. 4, pp. 5903–5910, 2025, doi: 10.36040/jati.v9i4.13916.

[30] R. Revina, F. Helmiah, and A. K. Syahputra, "Penerapan K-Means clustering kesiapan siswa SMA dalam pemilihan jurusan kuliah," Jurnal Algoritma, vol. 23, no. 1, pp. 595–604, 2026.