Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id

E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September

OPEN ACCESS

https://doi.org/10.35870/jpni.v6i3.1609

Sistem *Data Mining* Penentuan Prioritas terhadap Penerima Bantuan Bencana Banjir dengan Metode *Naive Bayes* dan Klusterisasi *K-Means* (Studi Kasus: Wilayah Cengkareng 2025)

Frencis Matheos Sarimole [1], Laily Nurmayanti [2*]

[1,2*] Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

Corresponding Email: matheosfrencis.s@gmail.com [1]

Histori Artikel:
Dikirim 25 Juni 2025; *Diterima dalam bentuk revisi* 10 Juli 2025; *Diterima* 25 Agustus 2025; *Diterbitkan* 10 September 2025.

Abstrak

Penelitian ini merancang sistem pemeringkatan wilayah prioritas penerima bantuan banjir di Jakarta, khususnya Cengkareng, dengan memanfaatkan algoritma K-Means dan Naïve Bayes. Data bersumber dari Satu Data Jakarta tahun 2025, mencakup 158 catatan yang terdiri atas wilayah, kecamatan, kelurahan, rata-rata ketinggian air, jumlah RW, jumlah KK, jumlah jiwa terdampak, dan jumlah kejadian banjir. Analisis dilakukan melalui tahapan pembersihan dan normalisasi data, pengelompokan tingkat risiko menggunakan K-Means ke dalam tiga kategori (tinggi, sedang, rendah), serta klasifikasi prediktif dengan Naïve Bayes. Evaluasi model menggunakan rasio data latih dan uji 70:30, 80:20, dan 90:10 menunjukkan bahwa metode gabungan K-Means dan Naïve Bayes berhasil mencapai akurasi tertinggi sebesar 98,18%, jauh melampaui Naïve Bayes konvensional yang hanya memperoleh akurasi 43,47%. Peningkatan akurasi ini menegaskan efektivitas integrasi kedua algoritma dalam klasifikasi data kompleks. Sistem yang dibangun mempercepat proses penentuan prioritas bantuan, memudahkan verifikasi daftar penerima oleh kader lokal, dan meningkatkan ketepatan distribusi logistik serta evakuasi warga. Simulasi bersama masyarakat dilakukan untuk menguji penerapan sistem di lapangan dan memastikan aksesibilitas informasi risiko banjir secara langsung. Pengembangan selanjutnya diarahkan pada integrasi variabel eksternal seperti data curah hujan real time dan pengujian di wilayah lain.

Kata Kunci: Banjir; Data Mining; Naïve Bayes; K-Means.

Abstract

This research develops a ranking system for flood aid recipients in Jakarta, focusing on Cengkareng, by utilizing K-Means and Naïve Bayes algorithms. Data were obtained from Satu Data Jakarta (2025), comprising 158 records with attributes including region, sub-district, village, average water level, affected RWs, families, individuals, and flood events. The analytical workflow encompasses data cleaning and normalization, risk level clustering using K-Means (three categories: high, medium, low), and predictive classification with Naïve Bayes. Model evaluation at training-testing splits of 70:30, 80:20, and 90:10 reveals that the combined K-Means and Naïve Bayes approach achieves the highest accuracy of 98.18%, significantly outperforming conventional Naïve Bayes which reached only 43.47%. This improvement demonstrates the effectiveness of combining both algorithms for complex data classification. The developed system expedites the prioritization process, facilitates local teams in verifying recipient lists, and enhances the precision of aid distribution and evacuation. Field simulations with community members were conducted to assess the system's practical implementation and ensure direct access to flood risk information. Future development will focus on integrating external variables such as real-time rainfall data and expanding field testing to other regions.

Keyword: Flood; Data Mining; Naïve Bayes; K-Means.

Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

1. Pendahuluan

Republik Indonesia, sebagai negara kepulauan yang terletak di garis khatulistiwa dan berada pada pertemuan tiga lempeng tektonik yaitu lempeng Eurasia, Indo-Australia, dan Pasifik memiliki tingkat kerentanannya yang tinggi terhadap berbagai jenis bencana alam, salah satunya banjir. Sebagai respons terhadap kondisi ini, pemerintah Indonesia telah menetapkan Undang-Undang Nomor 24 Tahun 2007 yang mengatur pembentukan Badan Penanggulangan Bencana Daerah (BPBD) untuk mengelola dan merespons kejadian-kejadian bencana (Angreini & Supratman, 2021). Menurut Badan Nasional Penanggulangan Bencana (BNPB), banjir didefinisikan sebagai fenomena peningkatan volume air yang menggenangi wilayah daratan. Banjir menjadi salah satu bencana alam yang paling sering terjadi di Indonesia, dengan proporsi mencapai 40% dari seluruh kejadian bencana yang tercatat di negara ini. Faktor penyebab banjir meliputi intensitas curah hujan, kemiringan lereng, aliran limpasan sungai, serta perilaku manusia yang kurang peduli terhadap pelestarian lingkungan (Anggraini *et al.*, 2021). Berdasarkan Indeks Risiko Bencana Indonesia (IRBI) tahun 2023, Provinsi DKI Jakarta memiliki skor 61,35, yang menempatkannya dalam kategori risiko sedang. Selain banjir, Jakarta juga menghadapi berbagai ancaman bencana alam lainnya, seperti gempa bumi, tanah longsor, kekeringan, cuaca ekstrem, dan abrasi. Topografi rendah, perubahan iklim, pembangunan yang pesat, serta tingkat kepadatan penduduk yang tinggi, turut memperburuk potensi risiko banjir di ibu kota (Badan Nasional Penanggulangan Bencana, 2023). Mengingat risiko tersebut terus meningkat, diperlukan sistem yang dapat memetakan dan memprioritaskan wilayah yang membutuhkan bantuan banjir dengan lebih objektif dan terukur, sehingga upaya intervensi dapat dilakukan secara tepat waktu dan sesuai sasaran.

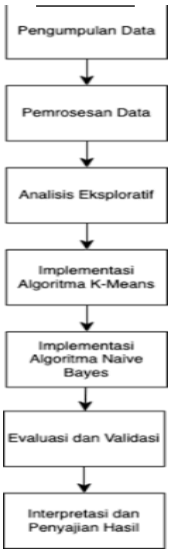
Data mining menawarkan berbagai teknik analisis data yang dapat diterapkan dalam berbagai konteks, seperti klasterisasi, asosiasi, klasifikasi, regresi, dan peramalan (Bui & Bahtiar, 2024). Salah satu metode yang sering digunakan untuk pengelompokan data adalah algoritma K-Means, yang berfokus pada meminimalkan variasi dalam kelompok dan memaksimalkan variasi antar kelompok (Khomsiyah *et al.*, 2021). K-Means terbukti efektif dalam mengidentifikasi pola-pola wilayah rawan banjir berdasarkan karakteristik data. Di sisi lain, Naïve Bayes adalah algoritma klasifikasi yang sederhana namun sangat efisien, dengan kemampuan untuk menangani atribut yang hilang atau tidak lengkap serta tidak memerlukan jumlah data yang besar untuk mencapai hasil yang optimal (Fatonah *et al.*, 2021). Algoritma ini bertujuan untuk mengklasifikasikan objek data ke dalam kategori yang telah ditentukan sebelumnya (Alghifari & Juardi, 2021). Penelitian ini bertujuan untuk mengintegrasikan algoritma K-Means Clustering dan Naïve Bayes Classifier dalam pengembangan sistem pemeringkatan wilayah prioritas penerima bantuan banjir di Jakarta. Kombinasi kedua metode ini diharapkan dapat menghasilkan model yang lebih akurat jika dibandingkan dengan penggunaan Naïve Bayes secara konvensional. Hasil dari pendekatan ini akan dibandingkan dengan metode konvensional untuk menilai keunggulannya dalam klasifikasi wilayah rawan banjir (Nandang Iriadi *et al.*, 2020). Penelitian oleh Martin Saputra (2025) menunjukkan bahwa integrasi K-Means dan Naïve Bayes mampu meningkatkan akurasi prediksi banjir di Jakarta hingga mencapai 97%. Analisis data historis selama sepuluh tahun terakhir digunakan untuk mengelompokkan daerah terdampak dan mengklasifikasikan tingkat ketinggian air. Prediksi yang dihasilkan memberikan gambaran mengenai potensi penurunan dampak banjir di masa depan. Sistem yang dibangun diharapkan dapat memberikan kontribusi yang signifikan dalam mempercepat proses penentuan prioritas bantuan, sehingga mendukung kelancaran distribusi logistik dan evakuasi yang lebih efisien.

2. Metode

Penelitian ini bertujuan untuk mengembangkan sebuah sistem pemeringkatan dan klasifikasi wilayah rawan banjir di Kota Jakarta dengan memanfaatkan integrasi metode *K-Means Clustering* dan *Naïve Bayes Classifier*. Sistem yang dibangun tidak hanya bertujuan untuk mendukung pemerintah

Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

dalam penentuan prioritas penerima bantuan, tetapi juga untuk menyediakan informasi yang dapat diakses dan dipahami oleh masyarakat, sehingga mereka dapat melakukan tindakan antisipasi dan mitigasi secara mandiri. Tahapan metode penelitian ini terdiri dari serangkaian langkah yang dimulai dengan pengumpulan data, dilanjutkan dengan pemrosesan data, analisis menggunakan *K-Means* dan *Naïve Bayes*, serta evaluasi model yang dihasilkan. Proses ini bertujuan untuk memastikan bahwa model yang dikembangkan mampu mengklasifikasikan wilayah rawan banjir secara akurat dan dapat memberikan informasi yang relevan untuk pengambilan keputusan. Berikut ini adalah tahapan penelitian yang diilustrasikan dalam Gambar 1:



Gambar 1. Tahapan metode penelitian

Tahapan metode penelitian dalam studi ini dimulai dengan pengumpulan data dari sumber yang terpercaya, seperti *Satu Data Jakarta* tahun 2025. Data yang dikumpulkan mencakup informasi mengenai wilayah, kecamatan, kelurahan, rata-rata ketinggian air, jumlah RW terdampak, jumlah kepala keluarga (KK) terdampak, jumlah jiwa terdampak, jumlah kejadian banjir, serta faktor-faktor lain yang memengaruhi risiko banjir. Setelah data terkumpul, langkah berikutnya adalah pemrosesan data, yang mencakup pembersihan untuk menghilangkan data yang tidak relevan atau tidak valid, transformasi data sesuai kebutuhan analisis, dan pemilihan fitur untuk memfokuskan analisis. Pemrosesan ini bertujuan memastikan bahwa data yang digunakan memiliki kualitas yang baik, sehingga menghasilkan analisis yang akurat dan dapat diandalkan. Tahap selanjutnya adalah analisis eksploratif, di mana peneliti melakukan analisis awal untuk memahami pola dan hubungan antar variabel yang relevan (Zhang, 2020). Analisis ini berfungsi untuk mengidentifikasi wilayah-wilayah yang rentan terhadap banjir serta faktor-faktor utama yang mempengaruhi kejadian banjir. Setelah itu, algoritma *K-Means* diterapkan untuk mengelompokkan wilayah berdasarkan karakteristik risiko banjir ke dalam beberapa kategori, seperti tinggi, sedang, dan rendah, dengan tujuan menemukan pola spasial yang tidak langsung terlihat dari data mentah. Setelah proses klasterisasi selesai, algoritma *Naïve Bayes* digunakan untuk membangun model klasifikasi yang memprediksi kemungkinan suatu wilayah masuk dalam kategori rawan banjir berdasarkan atribut yang tersedia, sekaligus membantu dalam memahami faktor-faktor penyebab banjir. Model yang dihasilkan kemudian dievaluasi menggunakan metrik seperti akurasi, *performance vector*, dan *weighted mean recall* (Ridwan, 2020). Validasi dilakukan dengan membagi data menjadi data latih dan data uji, serta menguji model pada data baru yang belum pernah dianalisis sebelumnya. Pada tahap interpretasi dan diseminasi hasil, temuan penelitian diinterpretasikan dan disampaikan kepada pemangku kepentingan, seperti pemerintah daerah, LSM, dan masyarakat umum. Selain itu, dilakukan simulasi penggunaan aplikasi sistem yang telah dikembangkan untuk memperlihatkan bagaimana masyarakat dapat mengakses informasi tingkat risiko banjir di wilayah mereka secara langsung melalui aplikasi.

Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

Melalui simulasi ini, diharapkan masyarakat dapat meningkatkan kesiapsiagaan dan melakukan tindakan mitigasi secara mandiri, serta memanfaatkan sistem sebagai acuan dalam pengambilan keputusan terkait penanganan banjir.

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini, yang bernama *Data Kejadian Bencana Banjir*, diperoleh dari website *Satu Data Jakarta* melalui tautan https://satudata.jakarta.go.id/open-data/detail?kategori=dataset&page_url=data-kejadian-bencana-banjir&data_no=1. Dataset ini terdiri dari 158 catatan yang mencakup berbagai atribut terkait kejadian banjir di wilayah Jakarta. Atribut-atribut dalam dataset ini akan dianalisis lebih lanjut menggunakan metode *K-Means Clustering* dan *Naïve Bayes Classifier*. Berikut adalah penjelasan mengenai atribut-atribut yang terdapat dalam dataset tersebut:

1) Wilayah: Nama wilayah di Jakarta yang terdampak oleh banjir.
2) Kecamatan: Nama kecamatan di Jakarta yang terdampak oleh banjir.
3) Kelurahan: Nama kelurahan di Jakarta yang terdampak oleh banjir.
4) Jumlah Rata-rata Ketinggian Air (cm): Rata-rata ketinggian air yang menggenangi wilayah, diukur dalam satuan sentimeter.
5) Jumlah RW Terdampak: Jumlah Rukun Warga (RW) yang terdampak oleh banjir.
6) Jumlah KK Terdampak: Jumlah Kepala Keluarga (KK) yang terdampak oleh banjir.
7) Jumlah Jiwa Terdampak: Jumlah individu (jiwa) yang terdampak oleh banjir.
8) Terdampak: Indikator apakah wilayah tersebut terdampak banjir atau tidak.
9) Jumlah Kejadian: Jumlah kejadian banjir yang terjadi di wilayah tersebut.

Tabel berikut menunjukkan sebagian data yang digunakan dalam penelitian ini:

wilayah	kecamatan	kelurahan	jumla	jumla	jumla	jumla	jumlal
JAKARTA SELATAN	KEBAYORAN LAMA	GROGOL SELATAN	50	2	0	0	1
JAKARTA SELATAN	KEBAYORAN LAMA	KEBAYORAN LAMA SELAT	20	1	0	0	1
JAKARTA SELATAN	MAMPANG PRAPATAN	PELA MAMPANG	32.5	1	0	0	1
JAKARTA SELATAN	MAMPANG PRAPATAN	TEGAL PARANG	40	1	0	0	1
JAKARTA SELATAN	MAMPANG PRAPATAN	KUNINGAN BARAT	45	3	0	0	1
JAKARTA SELATAN	PESANGGRAHAN	ULUJAMI	30	1	0	0	1
JAKARTA SELATAN	TEBET	BUKIT DURI	40	2	0	0	1
JAKARTA SELATAN	TEBET	MANGGARAI	50	4	0	0	1
JAKARTA SELATAN	TEBET	KEBON BARU	50	1	0	0	1
JAKARTA SELATAN	PANCORAN	RAWAJATI	70	1	0	0	1
JAKARTA SELATAN	PANCORAN	PENGADEGAN	30	1	0	0	1
JAKARTA TIMUR	KRAMAT JATI	KAMPUNG TENGAH	30	1	0	0	1
JAKARTA TIMUR	KRAMAT JATI	CILILITAN	95	1	0	0	1
JAKARTA TIMUR	KRAMAT JATI	CAWANG	117.5	5	0	0	1
JAKARTA TIMUR	KRAMAT JATI	DUKUH	45	1	0	0	1
JAKARTA TIMUR	KRAMAT JATI	BALEKAMBANG	100	1	0	0	1
JAKARTA TIMUR	JATINEGARA	BIDARA CINA	100	2	0	0	1

Gambar 2. Dataset kejadian Banjir di Wilayah Jakarta

3.1.2 Tahap *Preprocessing*

Tahap *preprocessing* merupakan langkah yang sangat penting dalam analisis data untuk memastikan bahwa data yang digunakan dalam algoritma *K-Means Clustering* dan *Naïve Bayes Classifier* memenuhi kriteria kebersihan, konsistensi, dan kesiapan untuk analisis lebih lanjut (Zai, 2022). Proses preprocessing dalam penelitian ini melibatkan beberapa langkah penting, yaitu pembersihan data, penanganan data hilang, normalisasi data, dan pemisahan data untuk keperluan pelatihan dan pengujian (Chikalkar, 2020).

1) Pembersihan Data

   Langkah pertama dalam preprocessing adalah pembersihan data, yang bertujuan untuk mengidentifikasi dan memperbaiki atau menghapus data yang tidak valid atau mengandung anomali. Dalam dataset yang diperoleh dari *Satu Data Jakarta*, penulis memeriksa adanya duplikasi data, inkonsistensi dalam penamaan wilayah, kecamatan, dan kelurahan, serta kesalahan

Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

penulisan lainnya. Semua data yang tidak memenuhi standar atau yang dianggap tidak relevan kemudian dihapus atau dikoreksi guna memastikan kualitas data yang optimal untuk analisis lebih lanjut.

wilayah	kecamatan	kelurahan	jumlah rata rata k	jumlah rw	jumlah kk	jumlah jiwa	jumlah kejadi
JAKARTA SELATA	KEBAYORAN LAMA	GROGOL SELATAN	50	2	0	0	1
JAKARTA SELATA	KEBAYORAN LAMA	KEBAYORAN LAMA SE	20	1	0	0	1
JAKARTA SELATA	MAMPANG PRAPATA	PELA MAMPANG	32.5	1	0	0	1
JAKARTA SELATA	MAMPANG PRAPATA	TEGAL PARANG	40	1	0	0	1
JAKARTA SELATA	MAMPANG PRAPATA	KUNINGAN BARAT	45	3	0	0	1
JAKARTA SELATA	PESANGGRAHAN	ULUJAMI	30	1	0	0	1
JAKARTA SELATA	TEBET	BUKIT DURI	40	2	0	0	1
JAKARTA SELATA	TEBET	MANGGARAI	50	4	0	0	1
JAKARTA SELATA	TEBET	KEBON BARU	50	1	0	0	1
JAKARTA SELATA	PANCORAN	RAWAJATI	70	1	0	0	1
JAKARTA SELATA	PANCORAN	PENGADEGAN	30	1	0	0	1
JAKARTA TIMUR	KRAMAT JATI	KAMPUNG TENGAH	30	1	0	0	1
JAKARTA TIMUR	KRAMAT JATI	CILILITAN	95	1	0	0	1
JAKARTA TIMUR	KRAMAT JATI	CAWANG	117.5	5	0	0	1

Gambar 3. Data Setelah Dibersihkan

Setelah melalui proses pembersihan, penulis mengidentifikasi, memperbaiki, dan menghapus beberapa data. Semula, data tersebut memiliki 15 atribut, yaitu: triwulan, bulan, wilayah, kecamatan, kelurahan, jumlah rata-rata ketinggian air, jumlah RW terdampak, jumlah KK terdampak, jumlah jiwa terdampak, jumlah kejadian, jumlah korban meninggal, jumlah korban luka, jumlah pengungsi, jumlah tempat pengungsian, dan nilai kerugian. Setelah dibersihkan, data tersebut disederhanakan menjadi 8 atribut, yaitu: wilayah, kecamatan, kelurahan, jumlah rata-rata ketinggian air, jumlah RW terdampak, jumlah KK terdampak, jumlah jiwa terdampak, dan jumlah kejadian.

2) Normalisasi Data
Normalisasi data dilakukan untuk memastikan bahwa setiap atribut berada dalam skala yang serupa, sehingga menghindari dominasi atribut tertentu dalam analisis (Sirichanya & Kraisak, 2021). Dalam penelitian ini, atribut numerik seperti *jumlah rata-rata ketinggian air*, *jumlah RW terdampak*, *jumlah KK terdampak*, dan *jumlah jiwa terdampak* dinormalisasi menggunakan metode *Min-Max Scaling*. Metode ini mengubah nilai atribut ke dalam rentang [0, 1], sehingga memastikan setiap atribut memiliki pengaruh yang seimbang dalam analisis. Proses normalisasi ini diharapkan dapat meningkatkan kinerja algoritma *K-Means Clustering* dan *Naïve Bayes Classifier*.

3) Pemisahan Data untuk Pelatihan dan Pengujian
Setelah melalui tahap pembersihan dan normalisasi, data kemudian dibagi menjadi dua subset: data untuk pelatihan dan data untuk pengujian. Data pelatihan digunakan untuk mengembangkan model algoritma, sementara data pengujian berfungsi untuk menilai kinerja model tersebut. Pembagian data ini dilakukan dengan tiga rasio pembagian yang berbeda, yaitu 70% untuk data pelatihan dan 30% untuk data pengujian, 80% untuk data pelatihan dan 20% untuk data pengujian, serta 90% untuk data pelatihan dan 10% untuk data pengujian.

3.1.3 Implementasi Gabungan Algoritma K-Means dan Naïve Bayes dengan RapidMiner

Setelah pemahaman yang mendalam tentang data tercapai, langkah selanjutnya dalam penelitian ini adalah implementasi algoritma *K-Means*. Algoritma ini digunakan untuk mengelompokkan wilayah berdasarkan karakteristik yang serupa terkait dengan risiko banjir. Tujuan dari penggunaan algoritma ini adalah untuk mengidentifikasi pola-pola spasial yang mungkin tidak terlihat langsung dari data mentah. Sebelum melakukan pengolahan data di *RapidMiner*, langkah pertama adalah mengimpor dataset yang akan digunakan. Setelah dataset diimpor, penting untuk memastikan bahwa tipe data pada setiap atribut sudah sesuai. Langkah berikutnya adalah dengan menggunakan fitur *drag and drop* untuk memasukkan dataset ke dalam alur kerja *RapidMiner* yang akan diproses.

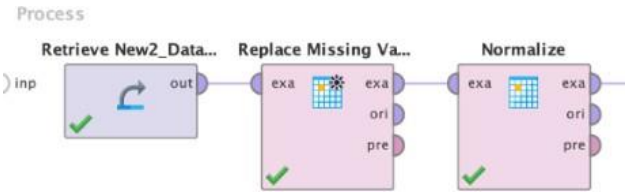
Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

Row No.	wilayah	kecamatan	kelurahan	jumlah_rat...	jumlah_rw_...	jumlah_kk_...	jumlah_jiw...	jumlah_kej...
1	JAKARTA SE...	KEBAYORAN...	GROGOL SE...	50	2	0	0	1
2	JAKARTA SE...	KEBAYORAN...	KEBAYORAN...	20	1	0	0	1
3	JAKARTA SE...	MAMPANG P...	PELA MAMP...	32.500	1	0	0	1
4	JAKARTA SE...	MAMPANG P...	TEGAL PAR...	40	1	0	0	1
5	JAKARTA SE...	MAMPANG P...	KUNINGAN ...	45	3	0	0	1
6	JAKARTA SE...	PESANGGRA...	ULUJAMI	30	1	0	0	1
7	JAKARTA SE...	TEBET	BUKIT DURI	40	2	0	0	1
8	JAKARTA SE...	TEBET	MANGGARAI	50	4	0	0	1
9	JAKARTA SE...	TEBET	KEBON BARU	50	1	0	0	1
10	JAKARTA SE...	PANCORAN	RAWAJATI	70	1	0	0	1
11	JAKARTA SE...	PANCORAN	PENGADEGAN	30	1	0	0	1
12	JAKARTA TI...	KRAMAT JATI	KAMPUNG T...	30	1	0	0	1
13	JAKARTA TI...	KRAMAT JATI	CILILITAN	95	1	0	0	1

ExampleSet (153 examples,0 special attributes,8 regular attributes)

Gambar 4. Dataset yang telah diimport pada rapidminer

Selanjutnya akan menggunakan operator *replace missing value* untuk menghilangkan missing value dan perlu untuk menambahkan operator *Normalize* untuk menormalisasikan *dataset* yang sebelumnya diimport sebagai berikut.



Gambar 5. Menambahkan Operator Normalize

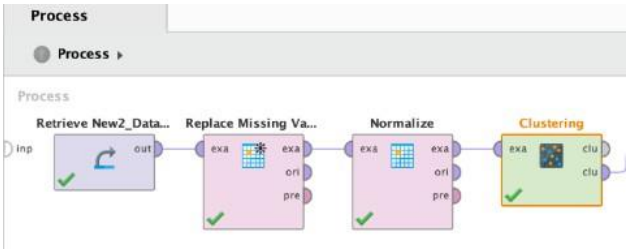
Diperoleh output sebagai berikut:

Row No.	jumlah_rat...	jumlah_rw_...	jumlah_kk_...	jumlah_jiw...	jumlah_kej...	wilayah	kecamatan	kelurahan
1	0.206	0.502	-0.154	-0.138	0.221	JAKARTA SE...	KEBAYORAN...	GROGOL SE...
2	-0.968	-0.412	-0.154	-0.138	0.221	JAKARTA SE...	KEBAYORAN...	KEBAYORAN...
3	-0.479	-0.412	-0.154	-0.138	0.221	JAKARTA SE...	MAMPANG P...	PELA MAMP...
4	-0.185	-0.412	-0.154	-0.138	0.221	JAKARTA SE...	MAMPANG P...	TEGAL PAR...
5	0.010	1.416	-0.154	-0.138	0.221	JAKARTA SE...	MAMPANG P...	KUNINGAN ...
6	-0.576	-0.412	-0.154	-0.138	0.221	JAKARTA SE...	PESANGGRA...	ULUJAMI
7	-0.185	0.502	-0.154	-0.138	0.221	JAKARTA SE...	TEBET	BUKIT DURI
8	0.206	2.330	-0.154	-0.138	0.221	JAKARTA SE...	TEBET	MANGGARAI
9	0.206	-0.412	-0.154	-0.138	0.221	JAKARTA SE...	TEBET	KEBON BARU
10	0.988	-0.412	-0.154	-0.138	0.221	JAKARTA SE...	PANCORAN	RAWAJATI
11	-0.576	-0.412	-0.154	-0.138	0.221	JAKARTA SE...	PANCORAN	PENGADEGAN
12	-0.576	-0.412	-0.154	-0.138	0.221	JAKARTA TI...	KRAMAT JATI	KAMPUNG T...
13	1.966	-0.412	-0.154	-0.138	0.221	JAKARTA TI...	KRAMAT JATI	CILILITAN

ExampleSet (153 examples,0 special attributes,8 regular attributes)

Gambar 6. Hasil Normalisasi Data

Berikutnya menambahkan operator *K-Means Clustering* untuk mengklasterisasikan data kejadian bajir di DKI Jakarta sebagai berikut.

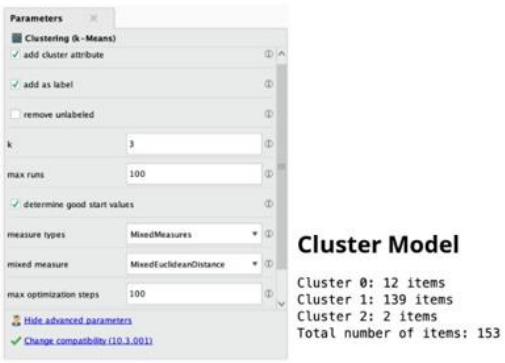


Gambar 7. Penambahan Operator K-Means Clustering

Adapun pengaturan parameter pada operator *K-Means Clustering* menggunakan k=3 yakni membagi dataset menjadi 3 kategori yakni tinggi, rendah, sedang (Nigam & Rajavat, 2020). Pengkategorian tersebut didasarkan pada ketinggian air sebagai indikator utama. Kategori 'tinggi'

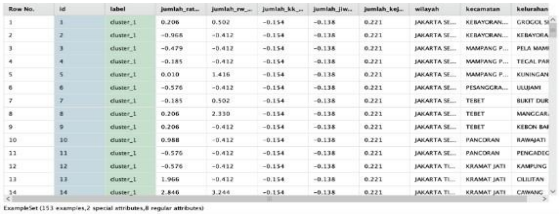
Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

menunjukkan area yang memerlukan prioritas evakuasi segera, sementara kategori 'sedang' mengindikasikan kebutuhan untuk bersiap-siap menghadapi potensi evakuasi. Kategori 'rendah' menunjukkan area yang belum memerlukan tindakan evakuasi dalam waktu dekat, namun tetap perlu dipantau. Selanjutnya *max runs*=100 yakni maksimal melakukan iterasi sebanyak 100 kali. *Measure types* yang digunakan adalah *mixed Measured* karena jenis tipe data yang ada pada tiap atribut beragam sehingga digunakan pengukuran campur. *Mixed measure* yang digunakan adalah *Mixed Euclidean Distance*, dengan pengaturan parameter sebagai berikut.



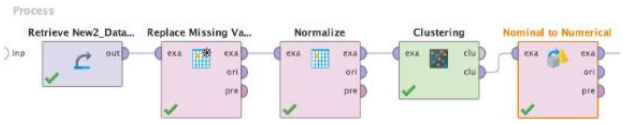
Gambar 8. Setting Parameter pada operator K-Means Clustering dan Hasil Cluster Model K-Means

Berdasarkan hasil clustering untuk pengelompokkan data banjir berdasarkan pengkategorian ketinggian air sebagai indikator utama diperoleh *model cluster* 0 yang berarti "sedang" sebanyak 12 *items*, *cluster* 1 yang berarti "tinggi" sebanyak 139 *items,* dan juga cluster 2 yang berarti "rendah" diperoleh sebanyak 2 *items* dengan total *items* sebanyak 153.

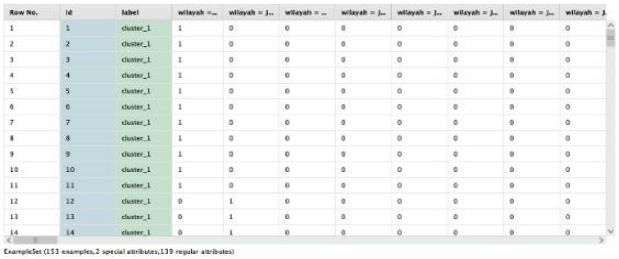


Gambar 9. Dataset Hasil Clustering K-Means

Selanjutnya perlu untuk menambahkan operator *Nominal to Numerical* untuk mengubah atribut yang bernilai nominal menjadi bernilai numerical sebagai berikut.



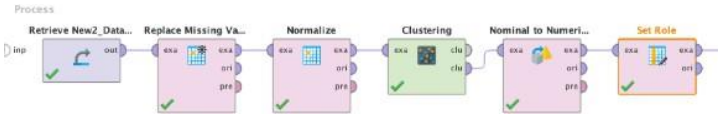
Gambar 10. Penambahan Operator *Nominal to Numerical*



Gambar 11. Dataset Hasil Penambahan Operator *Nominal To Numerical*

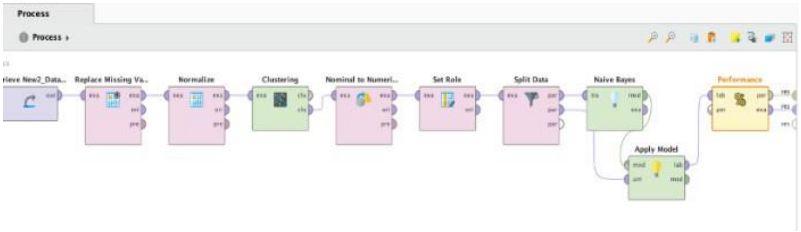
Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

Sebelum melakukan klasifikasi *naïve bayes* maka perlu menambahkan operator *set role* untuk mengatur atribut label sebagai label.



Gambar 12. Penambahan operator Set Role

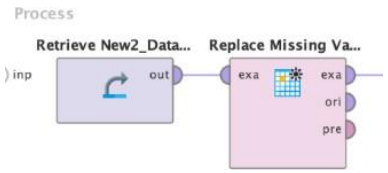
Langkah berikutnya adalah menambahkan operator pembagi data untuk memisahkan data menjadi dua subset, yaitu data pelatihan dan data pengujian. Data pelatihan digunakan untuk mengembangkan model algoritma, sedangkan data pengujian berfungsi untuk menilai performa model. Pembagian data ini dilakukan dalam tiga skenario proporsi, yaitu 70% data untuk pelatihan dan 30% untuk pengujian, 80% untuk pelatihan dan 20% untuk pengujian, serta 90% untuk pelatihan dan 10% untuk pengujian. Berikutnya menambahkan operator *Naïve Bayes*, *Apply Model* dan *performance* untuk mengklasifikasikan *dataset* pada data latih dan juga data uji sebagai berikut:



Gambar 13. Proses Keseluruhan *K-Means Clustering* dan *Naïve Bayes*

3.1.4 Implementasi Algoritma *Naïve Bayes* Konvensional dengan Rapidminer

Selain *K-Means*, penelitian juga melibatkan implementasi algoritma *Naïve Bayes*. Algoritma ini digunakan untuk membangun model klasifikasi yang dapat memprediksi wilayah-wilayah yang rentan terhadap banjir berdasarkan atribut-atribut yang diberikan. *Naïve Bayes* digunakan untuk memahami faktor-faktor yang berkontribusi terhadap risiko banjir dan memperkirakan probabilitas kejadian banjir di wilayah-wilayah tertentu (Erick *et al.*, 2024). Sebelum melakukan pemrosesan data maka perlu untuk import dataset yang akan digunakan. Kemudian *drag and drop* pada *process* sebagai berikut:

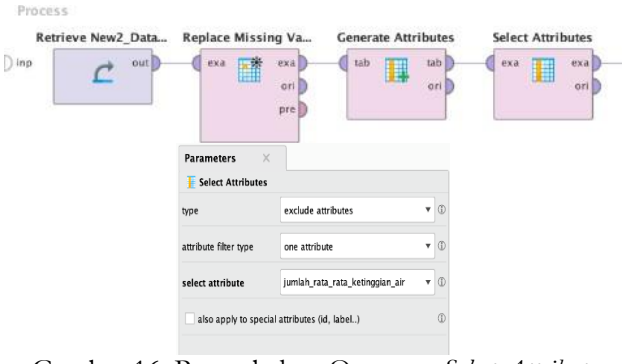


Gambar 14. Penambahan Operator *Replace Missing Value*

Selanjutnya menambahkan operator *replace missing value* untuk menghilangkan *missing value*. Kemudian perlu untuk menambahkan operator *generate attributes* serta menambahkan *expression*. Selain itu perlu menambahkan operator *select attribute* serta *exclude attribute* jumlah rata-rata ketinggian air sebagai berikut:

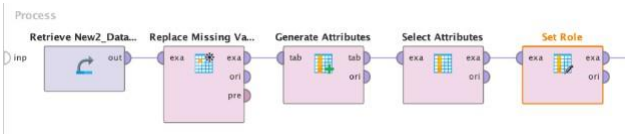
Gambar 15. Penambahan Operator *Generate Attribute*

Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609



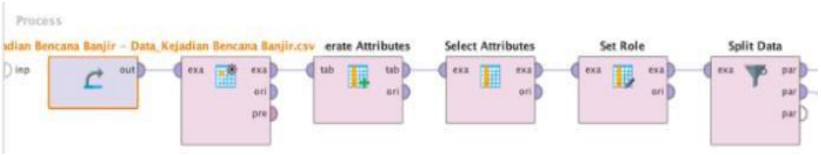
Gambar 16. Penambahan Operator *Select Attribute*

Selanjutnya menambahkan operator *set role* untuk *setting attribute* label sebagai label untuk pemrosesan klasifikasi *Naïve Bayes* sebagai berikut:



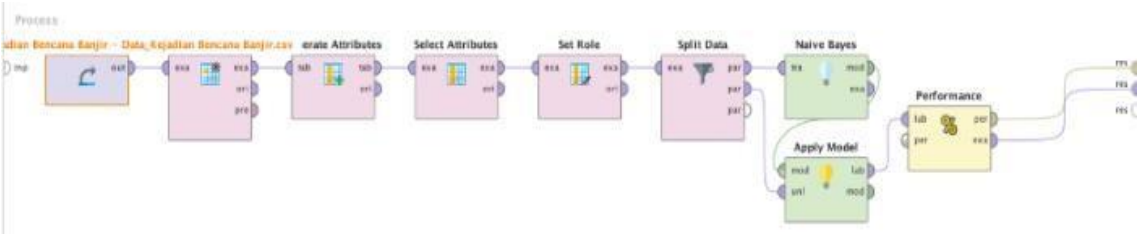
Gambar 17. Penambahan Operator Set Role

Selanjutnya menambahkan operator *set role* untuk *setting attribute* label sebagai label untuk pemrosesan klasifikasi *Naïve Bayes* Kemudian menambahkan operator *Split Data* untuk membagi data menjadi 2 sub bagian yakni data uji dan dan data latih yang terbagi menjadi 3 skenario proporsi yakni 70% untuk pelatihan dan 30% untuk pengujian, 80% untuk pelatihan dan 20% untuk pengujian, 90% untuk pelatihan dan 10% untuk pengujian. Hal tersebut dilakukan untuk memastikan bahwa model yang dihasilkan memiliki generalisasi yang baik terhadap data baru sebagai berikut.



Gambar 18. Penambahan Operator *Split Data* dan *Setting Split Data*

Langkah terakhir adalah menambahkan operator performance untuk mengetahui kinerja naïve bayes dalam mengklasifikasikan *dataset* daerah rawan banjir Kota Jakarta tersebut.



Gambar 19. Proses Lengkap Klasifikasi *Naïve Bayes* Konvensional

3.1.5 Evaluasi dan Validasi

Tahap berikutnya meliputi evaluasi dan validasi model yang dikembangkan dengan menggunakan algoritma *K-Means* dan *Naïve Bayes*. Evaluasi bertujuan untuk menilai performa model

Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

dalam mengklasifikasikan daerah-daerah yang rawan banjir. Metrik evaluasi yang digunakan mencakup akurasi dan *weighted mean recall* (Riyanto *et al.*, 2023). karena dalam banyak kasus, dataset memiliki kelas yang tidak seimbang (*imbalanced class*). *Weighted mean recall* memberikan bobot yang berbeda pada setiap kelas berdasarkan jumlah sampelnya sementara Dalam konteks integrasi *K-Means* dan *Naive Bayes*, akurasi penting untuk melihat seberapa baik model mengklasifikasikan data setelah proses *clustering*.Validasi dilakukan untuk memastikan bahwa model yang dibangun dapat digeneralisasi dengan baik pada data baru yang belum pernah dilihat sebelumnya, hasil *performance* dapat dilihat pada tabel berikut:

Tabel 1. Hasil Perbandingan Performance

	Akurasi			Weighted Mean Recall			Mean	
	70:30	80:20	90:10	70:30	80:20	90:10	Akurasi	WMR
Penggabungan K-Means dan Naïve Bayes	97,87%	96.67%	100%	66.67%	65.48%	66.67%	98.18%	66.27%
Naïve Bayes Konvensional	40,43%	43.33%	46.67%	56.04%	50.44%	36.67%	43.47%	47.72%

Berdasarkan Tabel 1, terlihat bahwa metode "Penggabungan *K-Means* dan *Naïve Bayes*" mencapai tingkat akurasi tertinggi, yaitu 98.18%. Hasil ini merupakan yang tertinggi dibandingkan dengan metode "*Naïve Bayes* Konvensional" yang hanya mencapai akurasi 43.47%. Perbandingan ini menunjukkan bahwa penggabungan metode *K-Means* dan *Naïve Bayes* secara signifikan meningkatkan akurasi dalam klasifikasi data. Dengan demikian, dapat disimpulkan bahwa metode "Penggabungan *K-Means* dan *Naïve Bayes*" sangat efektif dalam meningkatkan akurasi klasifikasi. Hasil akurasi sebesar 98.18% menunjukkan bahwa metode ini memiliki kinerja yang sangat baik dan dapat diandalkan untuk tugas-tugas klasifikasi data yang kompleks. Penelitian ini memberikan bukti kuat tentang potensi penggabungan metode dalam meningkatkan kinerja data mining.

Gambar 20. Simulasi bersama Warga Terkait Diadakannya sistem penentu kelayakan penerima bantuan sosial bencana banjir

3.2 Pembahasan

Pengumpulan data menjadi langkah awal dalam analisis banjir Jakarta. Dataset "Data Kejadian Bencana Banjir" dari Satu Data Jakarta memuat 158 catatan kejadian dengan berbagai atribut seperti wilayah, ketinggian air, jumlah RW, KK, dan jiwa terdampak. Pemilihan atribut mengacu pada penelitian Anggraini (2021) dan Fatonah (2021) yang menekankan nilai data spasial dan demografis untuk pemetaan daerah rawan banjir. Data terstruktur menjadi syarat utama agar analisis menggunakan K-Means dan Naïve Bayes berjalan optimal, sebagaimana dikemukakan Nandang Iriadi (2020). *Preprocessing* data memastikan kualitas analisis. Pembersihan data mengatasi masalah duplikasi, inkonsistensi penamaan, dan kesalahan penulisan, menghasilkan data yang valid (Zai, 2022; Chikalkar, 2020). Penelitian menyederhanakan data dari 15 menjadi 8 atribut utama sesuai kebutuhan analisis risiko banjir. Normalisasi dengan Min-Max Scaling mengubah nilai numerik ke rentang [0,1], mencegah dominasi atribut tertentu (Sirichanya & Kraisak, 2021). Pemisahan data pelatihan dan pengujian dalam beberapa rasio (70:30, 80:20, 90:10) bertujuan menguji kemampuan

E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

model terhadap data baru, sesuai rekomendasi Riyanto (2023). Algoritma *K-Means* digunakan untuk mengelompokkan wilayah berdasarkan tingkat risiko banjir dengan parameter k=3 (tinggi, sedang, rendah). Penentuan kategori didasarkan pada ketinggian air sebagai indikator utama, mengikuti pendekatan Khomsiyah (2021) dan Effendi (2024). Kategori 'tinggi' menandai wilayah prioritas evakuasi, 'sedang' untuk area yang perlu bersiap menghadapi potensi evakuasi, dan 'rendah' untuk area dalam tahap pemantauan. Mixed Euclidean Distance dipilih sebagai pengukuran jarak untuk mengakomodasi keragaman tipe data pada atribut, seperti dijelaskan Nigam & Rajavat (2020). Setelah klasterisasi, data dikonversi dari nominal ke numerik untuk meningkatkan kompatibilitas dengan algoritma Naïve Bayes. Penetapan label memungkinkan model melakukan klasifikasi secara terstruktur. Pembagian data dalam tiga skenario proporsi memastikan model memiliki kemampuan generalisasi yang baik terhadap data baru.

Sebagai pembanding, algoritma *Naïve Bayes* konvensional juga diterapkan untuk klasifikasi wilayah rawan banjir tanpa proses klasterisasi awal. Proses mengikuti tahapan preprocessing serupa, meliputi penanganan *missing value*, *generate attribute*, dan penetapan label. Model diuji pada tiga rasio pembagian data untuk menilai performa klasifikasi, mengacu pada metode Alghifari & Juardi (2021) dan Erick (2024). Evaluasi model menggunakan metrik akurasi dan *weighted mean recall*. *Weighted mean recall* dipilih karena relevansinya untuk dataset dengan kelas tidak seimbang, memberikan bobot berbeda pada tiap kelas sesuai jumlah sampel (Riyanto, 2023). Hasil evaluasi menunjukkan metode gabungan *K-Means* dan *Naïve Bayes* mencapai akurasi 98,18%, jauh melampaui *Naïve Bayes* konvensional yang hanya mencapai 43,47%. Peningkatan akurasi sejalan dengan temuan Martin Saputra (2025) dan Effendi (2024) yang menyatakan bahwa integrasi metode data mining dapat meningkatkan kinerja klasifikasi secara signifikan. Temuan tersebut membuktikan bahwa penggabungan *K-Means* dan *Naïve Bayes* sangat efektif untuk klasifikasi data kompleks, khususnya dalam penentuan wilayah prioritas penerima bantuan banjir. Akurasi tinggi yang dicapai menunjukkan pendekatan tersebut dapat diandalkan untuk mendukung pengambilan keputusan berbasis data dalam penanganan bencana. Penerapan sistem penentu kelayakan penerima bantuan sosial banjir diuji melalui simulasi bersama masyarakat. Simulasi bertujuan memastikan sistem tidak hanya unggul secara teknis, tetapi juga mudah diakses dan dipahami warga. Keterlibatan masyarakat memperkuat validasi eksternal dan mempercepat verifikasi daftar penerima bantuan oleh kader lokal (Angreini & Supratman, 2021). Langkah tersebut selaras dengan upaya BNPB (2023) meningkatkan partisipasi publik dalam penanggulangan bencana. Akurasi tinggi menunjukkan potensi integrasi K-Means dan Naïve Bayes dalam pemetaan risiko bencana. Pengembangan masa depan dapat diarahkan pada integrasi variabel eksternal seperti curah hujan *real time*, serta uji coba di wilayah lain untuk menguji generalisasi model (Bui & Bahtiar, 2024; Zhang, 2020).

4. Kesimpulan

Hasil dari penelitian ini menunjukkan bahwa penggabungan algoritma *K-Means Clustering* dan *Naïve Bayes Classifie*r memiliki performa yang lebih unggul dibandingkan dengan penggunaan algoritma *Naïve Bayes* konvensional dalam mengklasifikasikan wilayah rawan banjir di kota Jakarta. Dalam hal rata-rata akurasi, model yang menggunakan kombinasi kedua algoritma ini berhasil mencapai tingkat rata-rata akurasi sebesar 98.18% pada rasio split data 70:30, 80:20, 90:10 yang menunjukkan kemampuan model dalam mengklasifikasikan data dengan benar. Selain itu, nilai *Weighted Mean Recall* yang diperoleh sebesar 66.67% menunjukkan bahwa model memiliki tingkat keberhasilan yang baik dalam mendeteksi kelas-kelas banjir yang lebih jarang terjadi.

Jurnal Pengabdian Nasional (JPN) Indonesia
https://journal.stmiki.ac.id
E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

5. Daftar Pustaka

Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan dan Minuman Menggunakan Metode Algoritma Naïve Bayes: Studi Kasus: Makan Barbeque Sepuasnya. *Jurnal Ilmiah Informatika*, *9*(02), 75-81. https://doi.org/10.33884/jif.v9i02.3755.

Anggraini, N., Pangaribuan, B., Siregar, A. P., Sintampalam, G., Muhammad, A., Damanik, M. R. S., & Rahmadi, M. T. (2021). Analisis pemetaan daerah rawan banjir di kota medan tahun 2020. *Jurnal Samudra Geografi*, *4*(2), 27-33. https://doi.org/10.33059/jsg.v4i2.3851.

Angreini, S., & Supratman, E. (2021). Visualisasi Data Lokasi Rawan Bencana Di Provinsi Sumatera Selatan Menggunakan Tableau. *Jurnal Nasional Ilmu Komputer*, *2*(2), 135-147.

Bui, M. A., & Bahtiar, A. (2024). Implementasi metode algoritma K-Means Clustering untuk mengelompokkan transaksi penjualan barang di Toko Arino. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1451-1456.

Burhaein, E., Fadjeri, A., & Widiyono, I. P. (2024). Application of naive bayes algorithm for physical fitness level classification. *International Journal of Disabilities Sports and Health Sciences*, *7*(1), 178-187.

Effendi, M. M., & Siswandi, A. (2024). Analysis Prediksi Wilayah Rawan Banjir dengan Algoritma K-Means. *Journal of Information System Research (JOSH)*, *5*(2), 697-703.

Fatonah, N. S., Buana, M., Selatan, J. M., Kembangan, K., Barat, J., Khusus, D., ... & Com, N. (2021). Penerapan Deteksi Bencana Banjir Menggunakan Metode Machine Learning. *vol*, *10*, 119-126.

Iriadi, N., & Priatno, A. I. (2020). Penerapan Data Mining dengan Rapid Miner.

Khomsiyah, J., Ramdhani, A., Damayanti, A. F., & Rohman, D. (2021). Penerapan Algoritma K-means Clustering untuk Pengelompokan Wilayah Rawan Banjir. *JURNAL ILMIAH BETRIK: Besemah Teknologi Informasi dan Komputer*, *12*(3), 249-253.

Learning, M. M. M. Penerapan Deteksi Bencana Banjir Menggunakan Metode Machine Learning.

Nigam, N., & Rajavat, A. (2020). A Systematic Literature Review of Data Classification Techniques. *International Journal of Computer Applications*, *177*(44), 41.

Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, *4*(1), 15-21. https://doi.org/10.47970/siskom-kb.v4i1.169.

Riyanto, S., Imas, S. S., Djatna, T., & Atikah, T. D. (2023). Comparative analysis using various performance metrics in imbalanced data for multi-class text classification. *International Journal of Advanced Computer Science and Applications*, *14*(6).

Saputra, M. (2025). FLOOD PREDICTION WITH NAIVE BAYES METHOD. *Technovasia: Journal of Technology & Computer Research in Innovation, Science, and Applications*, *1*(1), 10-17.

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

E - ISSN: 2723-7060 | P - ISSN: 2776-8066
Vol. 6 No. 3 (2025) | September
OPEN ACCESS
https://doi.org/10.35870/jpni.v6i3.1609

Jakarta. *Computer Science (CO-SCIENCE)*, *5*(2), 67-76. https://doi.org/10.31294/coscience.v5i2.6900.

Sirichanya, C., & Kraisak, K. (2021). Semantic data mining in the information age: A systematic review. *International Journal of Intelligent Systems*, *36*(8), 3880-3916. https://doi.org/10.1002/int.22443.

Yunus, A. Y., Ahmad, S. N., Latief, R., Mulfiyanti, D., Badrun, B., Syarif, M., ... & Gusty, S. (2024). *Bencana alam dan manajemen risiko bencana*. Tohar Media.

Zai, C. (2022). Implementasi data mining sebagai pengolahan data. *Jurnal Portal Data*, *2*(3).

Zhang, X. (2020). Research on data mining algorithm based on pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, *34*(06), 2059015. https://doi.org/10.1142/S0218001420590156.