ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

Klasifikasi Tingkat Dampak Banjir Di Provinsi Sumatra Utara Menggunakan Algoritma K-Nearest Neighbor(Knn)

Maria Afri Yani[1*], Dia Alemisa br Sembiiring[2], Sardo Pardingotan Sipayung[3]
[1,2,3]Program Studi Teknik Informatika, Universitas Katolik Santo Thomas, Medan, Indonesia
Email: [1*]mariaafriyani7@gmail.com, [2]diaalemisa@gmail.com, [3]pinsarsiphom@gmail.com
(* Email Corresponding Author: mariaafriyani7@gmail.com)
Dikirim: 30 Oktober 2025 | Diterima: 20 Januari 2026 | Dipublish: 25 Januari 2026

Abstrak

Banjir merupakan salah satu bencana alam yang perlu diwaspadai karena dapat menimbulkan dampak yang signifikan terhadap masyarakat, baik berupa kerusakan infrastruktur maupun korban jiwa. Badan Nasional Penanggulangan Bencana (BNPB) dan Badan Penanggulangan Bencana Daerah (BPBD) berperan dalam menghimpun dan menyediakan data kejadian banjir sebagai dasar pengambilan keputusan dalam penanggulangan bencana. Penelitian ini bertujuan untuk mengklasifikasikan tingkat dampak banjir di Provinsi Sumatera Utara menggunakan metode data mining dengan pendekatan klasifikasi. Dataset yang digunakan merupakan data dampak banjir yang diperoleh dari BNPB dan BPBD Provinsi Sumatera Utara, yang meliputi jumlah rumah rusak, jumlah pengungsi, jumlah korban meninggal, jumlah korban hilang, serta jumlah fasilitas umum yang mengalami kerusakan. Proses data mining dilakukan mengikuti tahapan Knowledge Discovery in Databases (KDD), yang meliputi seleksi data, praproses data, normalisasi, proses klasifikasi, dan evaluasi hasil. Algoritma yang digunakan dalam penelitian ini adalah K-Nearest Neighbor (KNN). Pengolahan dan pengujian data dilakukan menggunakan perangkat lunak RapidMiner. Hasil pengujian menunjukkan bahwa algoritma KNN mampu mengklasifikasikan tingkat dampak banjir ke dalam tiga kelas, yaitu rendah, sedang, dan tinggi, dengan tingkat akurasi terbaik sebesar **89,47%**. Hasil penelitian ini menunjukkan bahwa algoritma KNN cukup efektif digunakan dalam klasifikasi tingkat dampak banjir di Provinsi Sumatera Utara berdasarkan data dampak bencana.

Kata Kunci: Banjir, Dampak Banjir, Klasifikasi, *Data Mining, K-Nearest Neighbor*.

Abstract

*Floods are one of the natural disasters that need to be anticipated because they can cause significant impacts on society, including infrastructure damage and loss of life. The National Disaster Management Agency (BNPB) and the Regional Disaster Management Agency (BPBD) play an important role in collecting and providing flood-related data as a basis for decision-making in disaster management. This study aims to classify the level of flood impact in North Sumatra Province using a data mining approach with classification techniques. The dataset used in this study consists of flood impact data obtained from BNPB and BPBD of North Sumatra Province, including the number of damaged houses, the number of evacuees, the number of fatalities, the number of missing persons, and the number of damaged public facilities. The data mining process follows the Knowledge Discovery in Databases (KDD) stages, which include data selection, preprocessing, normalization, classification, and evaluation. The algorithm applied in this study is the K-Nearest Neighbor (KNN) algorithm. Data processing and testing were conducted using RapidMiner software. The experimental results show that the KNN algorithm is able to classify flood impact levels into three classes, namely low, medium, and high, with the best accuracy of **89.47%**. These results indicate that the KNN algorithm is sufficiently effective for classifying flood impact levels in North Sumatra Province based on disaster impact data.*

Keywords : *flood, flood impact, Classification, Data Mining, K-Nearest Neighbor*

1. PENDAHULUAN

Indonesia merupakan negara yang memiliki kerawanan terhadap jenis bencana alam. Bencana alam ini mengakibatkan banyak kerugian yang berdampak langsung maupun tidak langsung seperti adanya korban jiwa, rusaknya fasilitas dan infrastruktur, hilangnya barang berharga, rusaknya lingkungan hidup, begitupun psikologis para korban bencana. Menurut UU No. 24 Tahun 2011, bencana adalah peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat, yang disebabkan baik oleh faktor alam dan atau faktor non alam maupun faktor manusia, sehingga mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis(Seftiani, 2023)[1]. Dari berbagai jenis bencana alam yang terjadi di Indonesia, banjir merupakan salah satu bencana yang paling sering terjadi. BNPB mencatat bahwa banjir merupakan jenis bencana yang yang sering terjadi di Indonesia. BNPB melaporkan bahwa dari sejumlah kejadian bencana yang tercatat di

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

awal tahun 2025, banjir mendominasi sekitar 80% dari total kejadian bencana, menunjukkan bahwa banjir adalah ancaman utama bencana hidrometeorologi di berbagai wilayah Indonesia.

Data Badan Nasional Penanggulangan Bencana (BNPB) menunjukkan bahwa banjir menempati persentase tertinggi dibandingkan jenis bencana lainnya [2]. Banjir sering terjadi akibat curah hujan ekstrem dan perubahan tata guna lahan[3]. Banjir dapat juga terjadi karena debit/volume air yang mengalir pada suatu sungai atau saluran drainase melebihi atau diatas kapasitas pengalirannya. Luapan air biasanya tidak menjadi persoalan bila tidak menimbulkan kerugian, korban meninggal atau luka, tidak merendam permukiman dalam waktu lama, tidak menimbulkan persoalan lain bagi kehidupan sehari-hari. Bila genangan air terjadi cukup tinggi, dalam waktu lama, dan sering maka hal tersebut akan mengganggu kegiatan manusia. Dalam sepuluh tahun terakhir ini, luas area dan frekuensi banjir semakin bertambah dengan kerugian yang makin besar[4].

Banjir merupakan salah satu bencana alam yang terjadi secara musiman di Provinsi Sumatera Utara, khususnya pada saat intensitas curah hujan tinggi Meskipun tidak terjadi sepanjang tahun, banjir yang muncul pada musim hujan dapat menimbulkan dampak yang signifikan terhadap Masyarakat, dampak banjir mencakup kerusakan rumah, fasilitas umum, serta korban jiwa. Pada tahun 2025, beberapa wilayah di Provinsi Sumatera Utara mengalami kejadian banjir dengan tingkat dampak yang berbeda-beda, antara lain Kota Medan, Deli Serdang, Langkat, Asahan, Binjai, Batu Bara, Humbang Hasundutan, Padang Sidempuan, Sibolga, Tebing Tinggi, Mandailing Natal, Nias, Nias Selatan, Nias Utara, Pakpak Bharat, Serdang Bedagai, Tapanuli Selatan, Tapanuli Tengah, dan Tapanuli Utara[5]. Perbedaan tingkat dampak banjir di setiap wilayah tersebut menunjukkan perlunya pengelompokan dampak banjir secara sistematis. Menurut Badan Nasional Penanggulangan Bencana (BNPB) banjir memiliki pengertian yakni meningkatnya volume air dengan menggenangi daratan. Banjir merupakan bencana alam yang sering terjadi, sebanyak 40% dibandingkan dengan bencana alam lainnya. Bencana banjir sendiri diakibatkan oleh beberapa faktor diantaranya curah hujan, kemiringan lereng, limpasan sungai, maupun faktor manusia seperti tidak menjaga lingkungan sekitar. (Anggraini , et al., 2021)[6].

Kondisi tersebut menunjukkan perlunya suatu pendekatan analisis yang dapat mengelompokkan tingkat dampak banjir secara objektif berdasarkan karakteristik data yang tersedia. Salah satu metode yang dapat membantu analisis big data adalah penambangan data (data mining). Penambangan data atau data mining adalah proses mengekstraksi dan menemukan informasi dari databasedan mengkonversinya menjadi informasi yang berguna untuk memunculkan pengetahuan. Melalui data mining juga dapat meningkatkan pengetahuan dan membantu mengembangkan model yang dapat mengungkap koneksi dengan jutaan bahkan miliaran rekaman data[7]. Data mining meninjau berbagai informasi guna mendapat korelasi yang tak terduga dan merangkum informasi melalui metode yang berbeda agar bisa dimengerti dan berguna untuk pemilik informasi[8].

Di antara berbagai teknik yang termasuk dalam data mining, *klasifikasi* merupakan salah satu metode yang sering digunakan untuk mengorganisasikan data ke dalam kelas-kelas tertentu berdasarkan pola-pola yang ditemukan. Klasifikasi merupakan teknik supervised dalam data mining. Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma yang dapat digunakan seperti Decision Tree, Support Vector Machine(SVM), K-Nearest Neighbor(KNN), Neural Network(NN), dan lain –lain[9]. Algoritma pengklasifikasian yang paling banyak digunakan di dalam machine learning adalah K-Nearest Neighbor(KNN). KNN bekerja dengan mencari rentang terpendek antar data yang akan diestimasi dengan tetangganya di data latih dengan tujuan mengklasifikasikan objek baru berdasarkan atribut dan model datapelatihan serta hasilnya akan diklasifikasikan berdasarkan ketepatan dalam penyelesaian pembelajaran[10]. Penerapan teknik klasifikasi dalam pengolahan data bencana semakin mendapat perhatian dalam studi kebencanaan di Indonesia, karena dapat membantu mengekstraksi pola dan keterkaitan antar variabel yang sulit ditangkap secara manual. Klasifikasi memungkinkan data-data numerik seperti jumlah rumah rusak, korban, dan fasilitas terdampak untuk dikelompokkan ke dalam tingkat dampak tertentu secara konsisten dan objektif, sehingga hasilnya dapat dimanfaatkan sebagai acuan dalam perencanaan mitigasi dan strategi respons bencana di tingkat daerah dan nasional (Situmorang & Rahayu, 2024)[11].

Dalam penelitian ini, dilakukan kajian terhadap beberapa penelitian terdahulu yang membahas klasifikasi dan analisis data bencana banjir menggunakan berbagai algoritma data mining. Metode yang digunakan dalam penelitian-penelitian tersebut meliputi Decision Tree, Support Vector Machine (SVM), Naive Bayes, K-Means, serta K-Nearest Neighbor (KNN). Kajian terhadap berbagai algoritma klasifikasi ini dilakukan sebagai bahan perbandingan dan untuk memperoleh gambaran umum mengenai penerapan teknik klasifikasi dalam

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

pengolahan data bencana banjir. Hasil kajian tersebut digunakan sebagai dasar dalam pemilihan algoritma K-Nearest Neighbor (KNN) pada penelitian ini.

penelitian terdahulu yang pernah diteliti oleh Cumel, David Zamri, Rahmaddeni dan Syamsurizal dengan judul *"Perbandingan Metode Data Mining Untuk Prediksi Banjir Dengan Algoritma Naïve Bayes dan KNN*" Berdasarkan hasil evaluasi performansi model klasifikasi yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Nearest Neighbor (kNN) menunjukkan kinerja yang lebih baik dibandingkan dengan algoritma Naïve Bayes. Pengukuran performansi menggunakan nilai akurasi dan error menunjukkan bahwa algoritma kNN dengan nilai k = 5 menghasilkan akurasi sebesar 88,94% dengan tingkat kesalahan (error) sebesar 11,06%. Sementara itu, algoritma Naïve Bayes hanya menghasilkan akurasi sebesar 74,36% dengan error sebesar 25,64%. Hasil perbandingan ini menunjukkan bahwa algoritma kNN lebih efektif dalam melakukan klasifikasi dibandingkan algoritma Naïve Bayes pada data yang digunakan[12].

Penelitian yang dilakukan oleh Pangat Eko Putra, Muhamad Azhri Amrullah, Yahya Hasani Fauzi, Refy Fitriani Saputri dan Lelly Clodia RF[2024] dengan judul *"Penerapan Algoritma C4.5 untuk Klasifikasi Tingkat Korban Banjir di Indonesia"* Penelitian ini berhasil menggunakan algoritma C4.5 untuk mengklasifikasikan tingkat korban banjir di Indonesia berdasarkan data dari Badan Pusat Statistik (BPS) tahun 2023. Dengan mempertimbangkan jumlah korban meninggal, luka-luka, dan terdampak, algoritma ini menghasilkan pohon keputusan yang membagi wilayah ke dalam tiga kategori dampak: Tinggi, Sedang, dan Rendah. Model yang dihasilkan menunjukkan tingkat akurasi sebesar 97,50% dengan margin kesalahan ±7,91%. Hasil analisis menunjukkan bahwa daerah dengan dampak Rendah memiliki sedikit korban meninggal, luka-luka, dan terdampak terbukti efektif dalam memberikan prediksi yang jelas mengenai tingkat dampak banjir di berbagai wilayah[13].

Penelitian yang dilakukan oleh Natzir [2023] dengan judul *"Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naive Bayes, dan Random Forest dalam Prediksi Kejadian Banjir"*. Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN), Naive Bayes, dan Random Forest dalam memprediksi kejadian banjir berdasarkan data historis curah hujan. Evaluasi model dilakukan menggunakan beberapa parameter, yaitu akurasi, presisi, recall, F1-score, dan AUC. Hasil penelitian menunjukkan bahwa algoritma KNN memperoleh akurasi sebesar 91,67%, dengan nilai presisi 100%, recall 85,71%, F1-score 92,31%, dan AUC sebesar 98,57%. Sementara itu, algoritma Naive Bayes menghasilkan akurasi sebesar 87,50%, dan Random Forest menunjukkan kinerja terbaik dengan akurasi 100% pada seluruh metrik evaluasi. Meskipun Random Forest memberikan hasil tertinggi, algoritma KNN tetap menunjukkan performa yang baik dan stabil dalam mengklasifikasikan kejadian banjir berdasarkan data curah hujan historis[14].

Penelitian yang dilakukan oleh Akbar et al. (2024) dengan judul *"Water Level Classification for Detect Flood Disaster Status Using K-Nearest Neighbor (KNN) and Support Vector Machine (SVM)"* membandingkan kinerja algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) dalam klasifikasi ketinggian air sebagai indikator banjir. Hasil penelitian menunjukkan bahwa algoritma SVM menghasilkan akurasi tertinggi sebesar 100%, sedangkan algoritma KNN juga mencapai akurasi 100% pada nilai K = 1 hingga 5. Meskipun SVM menunjukkan performa yang lebih stabil, KNN tetap mampu memberikan hasil klasifikasi yang baik dengan pemilihan nilai K yang tepat[15].

Berdasarkan kajian terhadap penelitian-penelitian terdahulu, dapat disimpulkan bahwa teknik data mining, khususnya metode klasifikasi, telah banyak diterapkan dalam analisis dan prediksi bencana banjir. Berbagai algoritma seperti Decision Tree (C4.5), Naive Bayes, Support Vector Machine (SVM), Random Forest, serta K-Nearest Neighbor (KNN) menunjukkan kinerja yang baik dalam mengolah data banjir, baik berdasarkan data curah hujan, ketinggian muka air, maupun data korban dan dampak banjir. Hasil penelitian terdahulu menunjukkan bahwa algoritma KNN secara konsisten mampu menghasilkan tingkat akurasi yang tinggi dan stabil, terutama ketika nilai parameter K ditentukan secara tepat, sehingga algoritma ini layak digunakan dalam permasalahan klasifikasi bencana banjir.

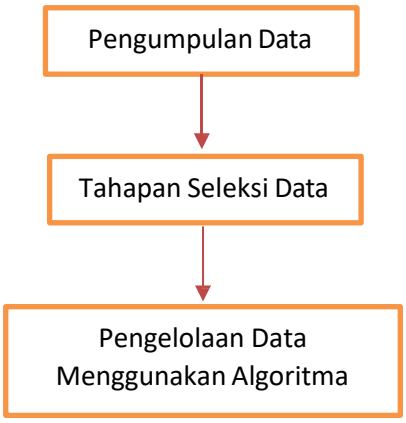
Penelitian ini bertujuan untuk menerapkan algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan tingkat dampak banjir di Provinsi Sumatera Utara menggunakan data dampak banjir yang bersumber dari Badan Nasional Penanggulangan Bencana (BNPB) dan Badan Penanggulangan Bencana Daerah (BPBD). Data yang digunakan meliputi jumlah rumah rusak, jumlah pengungsi, jumlah korban meninggal, jumlah korban hilang, serta jumlah fasilitas umum yang mengalami kerusakan. Melalui penelitian ini diharapkan dapat dihasilkan model klasifikasi tingkat dampak banjir yang mampu mengelompokkan wilayah ke dalam kelas rendah, sedang, dan tinggi, sehingga dapat menjadi bahan pendukung pengambilan keputusan dalam upaya mitigasi dan penanggulangan bencana banjir di Provinsi Sumatera Utara.

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan metode pada penelitian ini ditampilkan pada gambar 1.



Gambar 1. Alur Penelitian

Tahapan pada gambar 1, dapat dijelaskan sebagai berikut:

2.2 Pengumpulan Data

Tahap pertama yaitu proses pengumpulan data dari setiap parameter yang dibutuhkan. Pengumpulan data harus memenuhi beberapa prinsip yaitu mengumpulkan data selengkap-lengkapnya, mempertimbangkan ketepatan data dan kebenaran data[16]. Tahap awal penelitian adalah pengumpulan data dampak banjir. Data yang digunakan merupakan data sekunder yang diperoleh dari BNPB dan BPBD Provinsi Sumatera Utara pada tahun 2025, saat beberapa wilayah di Sumatera Utara mengalami kejadian banjir. Wilayah yang menjadi objek penelitian meliputi Kota Medan, Deli Serdang, Langkat, Asahan, Binjai, Batubara, Humbang Hasundutan, Padang Sidempuan, Sibolga, Tebing Tinggi, Mandailing Natal, Nias, Nias Selatan, Nias Utara, Pakpak Bharat, Serdang Bedagai, Tapanuli Selatan, Tapanuli Tengah, dan Tapanuli Utara. Data diproses secara manual di Microsoft Excel.

2.3 Tahapan Seleksi Data

Dalam tahap seleksi data, data yang telah dikumpulkan dipilih untuk mengidentifikasi atribut-atribut yang penting bagi penelitian[17]. Selain itu tahapan ini dilakukan sebelum tahap pemrosesan data. Pada tahap ini, data yang digunakan tadi akan dilakukan seleksi untuk mencari data atribut yang penting dalam penelitian. Data yang telah diseleksi menggunakan metode Software RapidMiner yang dapat dilihat pada Tabel 1.

Tabel 1. Atribut Terpilih

No	Atribut Terpilih	Keterangan
1	Kota/Kabupaten	Atribut
2	Total Rumah Rusak	Atribut
3	Jumlah Pengungsi	Atribut
4	Meninggal	Atribut
5	Hilang	Atribut
6	Fasilitas Rusak	Atribut
7	Kelas Dampak	Label

2.4 Pengolahan Data

Sebelum memasuki tahap pengolahan data, dilakukan terlebih dahulu proses pra-pemrosesan data. Tahap ini bertujuan untuk memperbaiki data, seperti menangani nilai yang hilang *(missing values)* atau kesalahan

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

pengetikan[18]. Pada tahap ini data yang telah diseleksi sesuai dengan keperluan penelitian diproses menggunakan Altair AI Studio Versi 2026.0.2 . Dataset lengkap dapat diakses melalui (https://www.bpbd.go.id). Berikut dataset dampak banjir yang dapat dilihat pada Gambar 2.

No	Kabupaten/Kota	Total Rumah Rusak	Jumlah Pengungsi	Meninggal	Hilang	Fasilitas Rusak	Kelas Dampak
1	Kota Medan	424	0	12	0	319	Tinggi
2	Deli Serdang	113	0	17	0	215	Tinggi
3	Langkat	3,420	2.3	16	0	298	Tinggi
4	Asahan	0	0	0	0	12	Rendah
5	Binjai	37	0	0	0	35	Sedang
6	batubara	42	0	0	0	0	Sedang
7	humbang hasudutan	330	855	10	1	11	Tinggi
8	padang sidempuan	330	0	1	0	22	Tinggi
9	sibolga	646	0	55	0	8	Tinggi
10	tebing tinggi	35	0	0	0	22	Sedang
12	mandailing natal	38	0	0	0	84	Sedang
13	nias	15	8	2	0	3	Tinggi
14	nias selatan	65	0	1	0	27	Tinggi
15	nias utara	0	0	0	0	0	Rendah
16	pak pak bharat	17	0	2	0	1	Tinggi
17	serdang begadai	24	0	0	0	44	Rendah
18	Tapanuli Selatan	2,461	4,200	93	4	91	Tinggi
19	Tapanuli Tengah	8,926	5000	130	34	38	Tinggi
20	Tapanuli utara	772	1	36	2	24	Tinggi

Gambar 2. Dataset Dampak Banjir Sumatra Utara

Berdasarkan Tabel 2 (Dataset Dampak Banjir di Provinsi Sumatera Utara), keterangan atribut yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Kabupaten/Kota
   Atribut yang mengidentifikasi wilayah administratif kabupaten atau kota di Provinsi Sumatera Utara yang terdampak banjir.
b. Total Rumah Rusak
   Atribut yang menunjukkan jumlah rumah penduduk yang mengalami kerusakan akibat bencana banjir pada masing-masing kabupaten/kota.
c. Total Jumlah Pengungsi
   Atribut yang menunjukkan jumlah penduduk yang harus mengungsi akibat terdampak bencana banjir di setiap wilayah kabupaten/kota.
d. Meninggal
   Atribut yang menunjukkan jumlah korban jiwa yang meninggal dunia akibat bencana banjir di masing-masing wilayah.
e. Hilang
   Atribut yang menunjukkan jumlah korban yang dinyatakan hilang akibat bencana banjir pada setiap kabupaten/kota.
f. Fasilitas Rusak
   Atribut yang menunjukkan jumlah fasilitas umum yang mengalami kerusakan akibat bencana banjir, seperti fasilitas pendidikan, kesehatan, dan infrastruktur lainnya.
g. Kelas Dampak
   Atribut label (target) yang menunjukkan tingkat dampak banjir pada setiap wilayah, yang diklasifikasikan ke dalam tiga kelas, yaitu Tinggi, Sedang, dan Rendah.

Atribut Kelas Dampak digunakan sebagai label dalam proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN), sedangkan atribut lainnya berperan sebagai atribut prediktor.

Tabel 2. Nilai Atribut Dataset Korban Banjir

No	Atribut	Keterangan
1	Kabupaten/Kota	Kota Medan, Deli Serdang, Langkat, Asahan, Binjai, Batubara, dst
2	Total Rumah Rusak	0, 15, 35, 113, 330, 424, 646, 772, 2.461, 8.926, dst

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

3	Jumlah Pengungsi	0, 1, 8, 855, 1.200, 4.200, 5.000, dst
4	Meninggal	0, 1, 2, 10, 12, 16, 36, 55, 93, 130, dst
5	Hilang	0, 1, 2, 4, 34, dst
6	Fasilitas Rusak	0, 1, 3, 8, 11, 22, 35, 84, 215, 298, 319, dst
7	Kelas Dampak	Tinggi, Sedang, Rendah

3. HASIL DAN PEMBAHASAN

3.1. Analisis Masalah

Bagian ini membahas hasil klasifikasi tingkat dampak banjir di kabupaten/kota Provinsi Sumatera Utara menggunakan algoritma K-Nearest Neighbor (KNN). Analisis dilakukan dengan mengikuti alur penelitian yang telah dirancang sebelumnya, dimulai dari pemahaman permasalahan, pengolahan data, hingga evaluasi hasil klasifikasi. Fokus utama pembahasan adalah bagaimana algoritma KNN mampu mengelompokkan wilayah terdampak banjir berdasarkan karakteristik data dampak yang digunakan, serta bagaimana variasi nilai parameter k memengaruhi hasil klasifikasi yang diperoleh.

Data yang dianalisis merupakan data dampak banjir yang bersumber dari BNPB dan BPBD Provinsi Sumatera Utara tahun 2025. Dataset ini mencakup beberapa atribut utama, yaitu jumlah rumah rusak, jumlah pengungsi, jumlah korban meninggal, jumlah korban hilang, serta jumlah fasilitas umum yang mengalami kerusakan. Atribut-atribut tersebut dipilih karena secara langsung merepresentasikan tingkat keparahan dampak banjir di setiap wilayah. Karakteristik data yang digunakan menunjukkan adanya variasi nilai yang cukup besar antar wilayah, mulai dari wilayah dengan dampak relatif rendah hingga wilayah dengan dampak yang sangat tinggi. Kondisi ini menunjukkan bahwa data memiliki pola yang beragam dan kompleks, sehingga memerlukan metode klasifikasi yang mampu menangani perbedaan karakteristik tersebut secara objektif.

3.2. Penerapan Metode K-Nearest Neighbor(KNN)

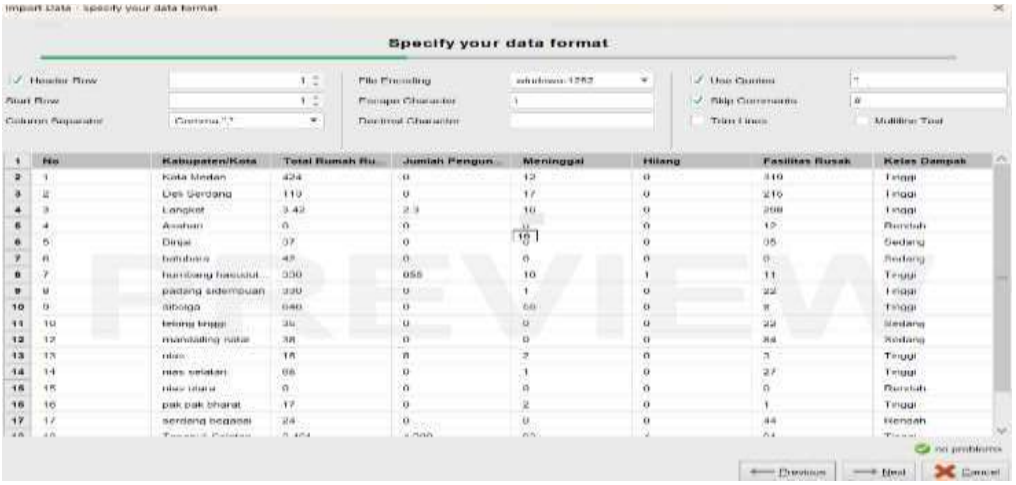
Sebelum proses klasifikasi dilakukan dataset terlebih dahulu melalui tahap pra-pemrosesan. Tahapan ini meliputi penanganan nilai hilang serta normalisasi data untuk memastikan setiap atribut berada pada skala yang sebanding. Proses pra-pemrosesan ini sangat penting karena algoritma KNN bekerja berdasarkan perhitungan jarak antar data, sehingga perbedaan skala antar atribut dapat memengaruhi hasil klasifikasi. Setelah proses pra-pemrosesan selesai, atribut *Kelas Dampak* ditetapkan sebagai label, sedangkan atribut lainnya berperan sebagai atribut prediktor dalam proses klasifikasi.

Proses pemodelan dan pengujian algoritma KNN dilakukan menggunakan perangkat lunak RapidMiner Studio. Dataset dibagi menjadi data latih dan data uji dengan rasio 70:30 menggunakan metode *Split Validation*. Pembagian ini bertujuan untuk menguji kemampuan model dalam mengklasifikasikan data baru yang belum pernah dilatih sebelumnya. Pada penelitian ini digunakan dua variasi nilai parameter, yaitu k = 3 dan k = 5, untuk melihat pengaruh jumlah tetangga terdekat terhadap performa model klasifikasi.

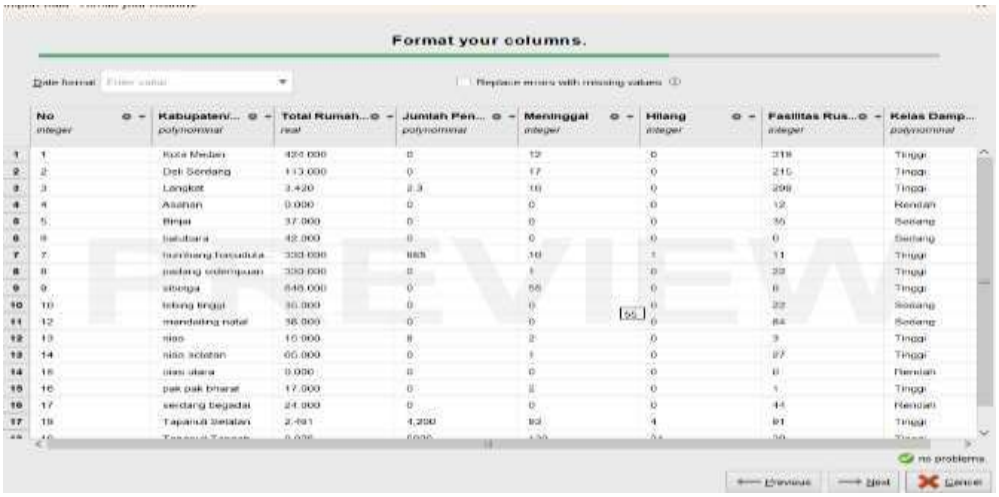
Alur pemodelan pada RapidMiner dimulai dari proses impor dataset, pengaturan format data, serta penentuan atribut label, sebagaimana ditunjukkan pada Gambar 3 dan Gambar 4. Selanjutnya, dataset diproses melalui operator pra-pemrosesan, dilanjutkan dengan pembagian data latih dan data uji. Proses klasifikasi dilakukan menggunakan algoritma KNN dengan nilai k yang telah ditentukan, kemudian hasil klasifikasi dievaluasi menggunakan metrik akurasi dan confusion matrix. Tampilan alur proses pengujian model ditunjukkan pada Gambar 3 sampai gambar 5, yang memperlihatkan hubungan antar tahapan pemodelan hingga menghasilkan output klasifikasi.

a. Pada tahap ini dilakukan pengaturan format data (*Specify your data format*) saat proses impor dataset ke dalam RapidMiner. Pada tahap ini ditentukan baris header, pemisah kolom, serta pemilihan variabel yang digunakan dalam penelitian. Proses ini bertujuan memastikan data telah sesuai sebelum dilakukan pra-pemrosesan lebih lanjut, sebagaimana ditunjukkan pada Gambar 3 dan Gambar 4.

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103



Gambar 3. Tahap Specify Your Data Format pada proses impor data



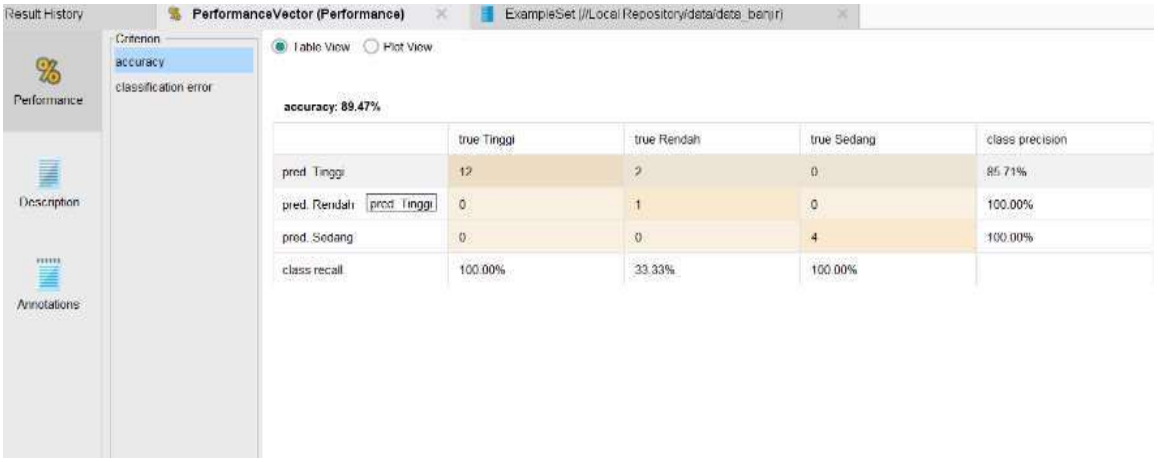
Gambar 4. Pemilihan Variabel sebagai Label

b. Tampilan proses pengujian model pada penelitian ini ditunjukkan pada Gambar 4, yang memperlihatkan alur pemodelan menggunakan metode *Split Data* guna penerapan algoritma KNN.

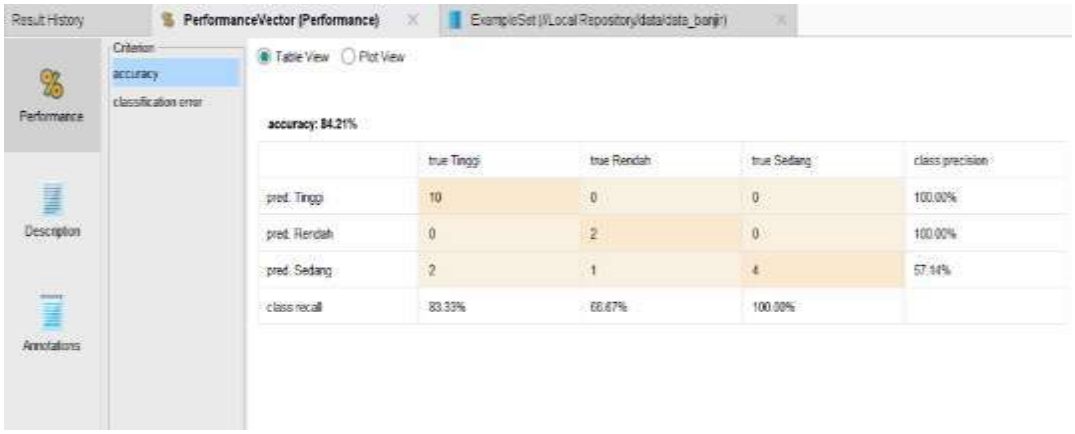
Gambar 5. Hasil keputusan pada *views result*

Proses dimulai dari pengambilan dataset, dilanjutkan dengan pra-pemrosesan data berupa penanganan nilai hilang dan normalisasi. Atribut Kelas Dampak ditetapkan sebagai label, kemudian dataset dibagi menjadi data latih dan data uji dengan rasio 70:30. Proses klasifikasi dilakukan menggunakan algoritma KNN dengan nilai $k = 3$ dan $k = 5$, dan evaluasi model dilakukan menggunakan metrik akurasi dan confusion matrix. Setelah seluruh tahapan pemodelan selesai dijalankan, RapidMiner akan memproses data dan menampilkan hasil klasifikasi pada tampilan *Result*. Pada penelitian ini dilakukan evaluasi performance pada setiap proses klasifikasi[19]. Hasil yang ditampilkan berupa nilai evaluasi kinerja model K-Nearest Neighbor (KNN), seperti akurasi, confusion matrix, serta nilai precision dan recall, sebagaimana ditunjukkan pada Gambar 6 dan gambar 7 dengan nilai k yang berbeda yai k=3, k=5.

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103



Gambar 6. Hasil Pengujian K=3



Gambar 7. Hasil Pengujian K=5

Nilai tetangga terdekat (K) dihitung berdasarkan jarak terpendek antara sampel uji dan sampel latih[20]. Hasil penelitian ini diperoleh dari proses pengujian algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan tingkat dampak banjir di Provinsi Sumatera Utara ke dalam tiga kelas, yaitu Tinggi, Sedang, dan Rendah. Pengujian dilakukan menggunakan dua variasi nilai parameter, yaitu k = 3 dan k = 5, dengan tujuan untuk mengetahui perbedaan hasil klasifikasi yang dihasilkan oleh masing-masing nilai k. Berdasarkan hasil pengujian menggunakan nilai k = 3, diperoleh tingkat akurasi sebesar 89,47% dengan classification error sebesar 10,53%. Hasil confusion matrix menunjukkan bahwa sebagian besar data uji dapat diklasifikasikan dengan benar ke dalam kelas Tinggi, Sedang, dan Rendah. Pada pengujian ini, model mampu mengelompokkan data kelas Sedang dan Rendah dengan sangat baik, meskipun masih terdapat beberapa kesalahan klasifikasi pada kelas Tinggi. Selanjutnya, hasil pengujian menggunakan nilai k = 5 menunjukkan tingkat akurasi sebesar 84% dengan classification error sebesar 15,79%. Dibandingkan dengan nilai k = 3, hasil klasifikasi pada k = 5 mengalami penurunan akurasi serta peningkatan tingkat kesalahan klasifikasi. Hal ini menunjukkan adanya perbedaan kinerja model KNN terhadap variasi nilai parameter k yang digunakan. Secara umum, hasil klasifikasi menunjukkan bahwa algoritma KNN mampu mengelompokkan tingkat dampak banjir berdasarkan data yang digunakan, dengan perbedaan performa yang dipengaruhi oleh pemilihan nilai parameter k.

Selain melihat nilai akurasi dan classification error, hasil klasifikasi juga dapat dianalisis lebih lanjut berdasarkan karakteristik masing-masing kelas dampak banjir. Kelas Tinggi umumnya merepresentasikan wilayah dengan jumlah rumah rusak, pengungsi, korban meninggal, serta fasilitas rusak yang relatif besar dibandingkan wilayah lain. Pada hasil pengujian, beberapa kesalahan klasifikasi masih ditemukan pada kelas Tinggi, yang menunjukkan bahwa karakteristik data pada kelas ini memiliki kedekatan nilai dengan kelas Sedang pada beberapa atribut tertentu. Kondisi tersebut dapat terjadi karena adanya wilayah yang memiliki jumlah pengungsi dan rumah rusak cukup tinggi, tetapi jumlah korban jiwa dan fasilitas rusak relatif rendah, sehingga secara jarak numerik masih dianggap dekat dengan kelas Sedang oleh algoritma KNN.

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

Pada kelas Sedang, hasil klasifikasi menunjukkan tingkat ketepatan yang lebih baik dibandingkan kelas Tinggi. Hal ini disebabkan oleh karakteristik data pada kelas Sedang yang cenderung lebih konsisten dan memiliki pola nilai atribut yang berada di tengah antara kelas Tinggi dan Rendah. Algoritma KNN mampu mengenali pola kedekatan data pada kelas ini dengan cukup baik, terutama pada penggunaan nilai k = 3, di mana mayoritas data uji berhasil diklasifikasikan sesuai dengan label sebenarnya. Kondisi ini menunjukkan bahwa algoritma KNN bekerja optimal ketika data memiliki distribusi yang relatif seimbang dan tidak terlalu ekstrem pada satu atribut tertentu.

Sementara itu, kelas Rendah menunjukkan hasil klasifikasi yang paling stabil. Wilayah yang termasuk dalam kelas Rendah umumnya memiliki nilai atribut yang kecil atau bahkan nol pada beberapa indikator dampak banjir, seperti korban meninggal, korban hilang, dan fasilitas rusak. Pola data yang relatif homogen pada kelas Rendah memudahkan algoritma KNN dalam menentukan kedekatan antar data, sehingga tingkat kesalahan klasifikasi pada kelas ini cenderung lebih rendah. Hal ini memperkuat temuan bahwa algoritma KNN sangat efektif dalam mengklasifikasikan data dengan karakteristik yang jelas dan tidak tumpang tindih dengan kelas lain.

Perbedaan hasil klasifikasi antara penggunaan nilai k = 3 dan k = 5 juga memberikan gambaran penting mengenai pengaruh jumlah tetangga terdekat terhadap kinerja model. Pada nilai k = 3, keputusan klasifikasi lebih dipengaruhi oleh tetangga terdekat yang benar-benar memiliki kemiripan tinggi dengan data uji. Hal ini membuat model lebih sensitif terhadap pola lokal pada data, sehingga mampu menghasilkan akurasi yang lebih tinggi. Sebaliknya, pada nilai k = 5, keputusan klasifikasi melibatkan lebih banyak tetangga, yang dalam beberapa kasus justru memasukkan data dari kelas lain yang memiliki jarak relatif dekat, sehingga menurunkan ketepatan klasifikasi.

Hasil ini menunjukkan bahwa pemilihan nilai k yang terlalu besar tidak selalu memberikan hasil yang lebih baik, terutama pada dataset dengan jumlah data yang terbatas dan distribusi kelas yang tidak sepenuhnya seimbang. Dalam konteks penelitian ini, nilai k = 3 dianggap lebih optimal karena mampu menjaga keseimbangan antara sensitivitas terhadap pola lokal dan stabilitas hasil klasifikasi. Temuan ini sejalan dengan karakteristik algoritma KNN yang sangat bergantung pada struktur data dan parameter k yang digunakan.

Jika dikaitkan dengan penelitian terdahulu, hasil yang diperoleh dalam penelitian ini menunjukkan konsistensi dengan beberapa penelitian sebelumnya yang menyatakan bahwa algoritma KNN memiliki performa yang baik dalam klasifikasi data bencana banjir. Beberapa penelitian terdahulu melaporkan bahwa KNN mampu menghasilkan akurasi yang tinggi pada data dampak banjir maupun data hidrologi, terutama ketika nilai k dipilih secara tepat. Meskipun pada beberapa studi lain algoritma seperti Random Forest atau SVM menunjukkan performa yang lebih tinggi, KNN tetap menjadi metode yang relevan karena kesederhanaan, interpretabilitas, dan kemampuannya dalam menangani data numerik multivariat.

Dalam penelitian ini, fokus utama bukanlah untuk mencari algoritma dengan akurasi tertinggi secara absolut, melainkan untuk mengevaluasi kemampuan KNN dalam mengklasifikasikan tingkat dampak banjir berdasarkan data dampak nyata di lapangan. Hasil yang diperoleh menunjukkan bahwa KNN mampu memberikan klasifikasi yang cukup akurat dan stabil, sehingga dapat digunakan sebagai pendekatan awal dalam analisis dampak bencana banjir di tingkat regional. Hal ini menjadi penting karena informasi mengenai tingkat dampak banjir sangat dibutuhkan oleh instansi terkait dalam proses perencanaan mitigasi dan penanggulangan bencana.

Implikasi praktis dari hasil penelitian ini adalah bahwa hasil klasifikasi dapat digunakan sebagai dasar dalam pemetaan wilayah rawan dampak banjir. Wilayah yang termasuk dalam kelas Tinggi dapat menjadi prioritas utama dalam alokasi sumber daya, seperti bantuan logistik, perbaikan infrastruktur, serta penyusunan rencana evakuasi. Wilayah dengan kelas Sedang dapat difokuskan pada upaya penguatan kesiapsiagaan dan mitigasi, sedangkan wilayah dengan kelas Rendah tetap perlu mendapatkan perhatian dalam bentuk pemantauan dan pencegahan agar tidak mengalami peningkatan dampak pada kejadian banjir berikutnya.

Selain itu, hasil penelitian ini juga menunjukkan pentingnya kualitas dan kelengkapan data dalam proses klasifikasi. Atribut-atribut yang digunakan dalam penelitian ini telah mampu merepresentasikan tingkat dampak banjir secara umum, namun masih terdapat peluang untuk meningkatkan akurasi model dengan menambahkan variabel lain yang relevan. Variabel seperti curah hujan, kondisi topografi, kepadatan penduduk, dan jarak terhadap sungai berpotensi memberikan informasi tambahan yang dapat memperjelas perbedaan antar kelas dampak banjir.

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

Dari sisi metodologi, penggunaan perangkat lunak RapidMiner Studio memberikan kemudahan dalam proses pemodelan, pengujian, dan evaluasi algoritma KNN. Visualisasi alur proses dan hasil evaluasi yang ditampilkan oleh RapidMiner membantu peneliti dalam memahami setiap tahapan pengolahan data dan interpretasi hasil. Hal ini menunjukkan bahwa penggunaan perangkat lunak data mining yang tepat dapat mendukung proses penelitian secara sistematis dan terstruktur, terutama bagi penelitian yang berfokus pada analisis data bencana.

3.3 Evaluasi Model

Setelah proses klasifikasi selesai dilakukan, hasil yang diperoleh tidak hanya menunjukkan kemampuan algoritma K-Nearest Neighbor (KNN) dalam mengelompokkan tingkat dampak banjir, tetapi juga perlu dianalisis lebih lanjut melalui tahapan evaluasi model. Evaluasi ini menjadi bagian penting dalam penelitian karena memberikan gambaran mengenai sejauh mana model klasifikasi yang dibangun mampu bekerja secara akurat dan konsisten pada data uji yang digunakan. Dengan melakukan evaluasi, kualitas hasil klasifikasi dapat diukur secara objektif, sekaligus menjadi dasar dalam menentukan konfigurasi parameter yang paling optimal.

Evaluasi model klasifikasi dilakukan untuk menilai kinerja algoritma K-Nearest Neighbor (KNN) berdasarkan hasil pengujian yang telah diperoleh. Evaluasi difokuskan pada perbandingan nilai akurasi dan classification error antara penggunaan nilai k = 3 dan k = 5. Berdasarkan hasil evaluasi, model KNN dengan nilai k = 3 menunjukkan kinerja yang lebih baik dibandingkan dengan k = 5. Hal ini ditunjukkan oleh nilai akurasi yang lebih tinggi, yaitu 89,47%, serta classification error yang lebih rendah sebesar 10,53%. Nilai error yang relatif kecil menunjukkan bahwa model dengan k = 3 mampu meminimalkan kesalahan dalam proses klasifikasi tingkat dampak banjir. Sebaliknya, pada penggunaan nilai k = 5, diperoleh akurasi yang lebih rendah, yaitu 84%, dengan classification error sebesar 15,79%. Peningkatan nilai error ini menunjukkan bahwa penambahan jumlah tetangga terdekat tidak selalu meningkatkan performa model, dan pada dataset penelitian ini justru menyebabkan penurunan ketepatan klasifikasi. Perbedaan hasil evaluasi ini menunjukkan bahwa pemilihan nilai parameter k sangat berpengaruh terhadap performa algoritma KNN. Dalam penelitian ini, nilai k = 3 dinilai lebih optimal karena mampu memberikan hasil klasifikasi yang lebih akurat dan konsisten dibandingkan dengan k = 5. Oleh karena itu, model KNN dengan k = 3 dipilih sebagai model terbaik untuk mengklasifikasikan tingkat dampak banjir di Provinsi Sumatera Utara.

Secara keseluruhan, hasil dan pembahasan dalam penelitian ini menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) mampu mengklasifikasikan tingkat dampak banjir di Provinsi Sumatera Utara dengan baik, terutama pada penggunaan nilai k = 3. Meskipun masih terdapat beberapa keterbatasan, seperti jumlah data yang relatif terbatas dan potensi tumpang tindih antar kelas, hasil penelitian ini memberikan gambaran awal yang jelas mengenai penerapan KNN dalam klasifikasi dampak bencana banjir. Dengan pengembangan lebih lanjut, baik dari sisi data maupun metode, pendekatan ini berpotensi memberikan kontribusi yang signifikan dalam mendukung pengambilan keputusan berbasis data pada bidang kebencanaan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Nearest Neighbor (KNN) mampu digunakan untuk mengklasifikasikan tingkat dampak banjir di Provinsi Sumatera Utara ke dalam tiga kelas, yaitu Tinggi, Sedang, dan Rendah, berdasarkan data dampak bencana banjir yang meliputi kerusakan dan korban banjir. Pengujian menunjukkan bahwa pemilihan nilai parameter k berpengaruh terhadap kinerja model klasifikasi. Nilai k = 3 menghasilkan performa terbaik dengan tingkat akurasi sebesar 89,47% dan classification error sebesar 10,53%, sedangkan penggunaan k = 5 menghasilkan akurasi yang lebih rendah, yaitu 84% dengan classification error sebesar 15,79%. Hal ini menunjukkan bahwa model KNN dengan k = 3 lebih optimal dan konsisten dalam mengklasifikasikan tingkat dampak banjir pada dataset penelitian ini. Hasil klasifikasi yang diperoleh dapat dimanfaatkan sebagai informasi pendukung bagi instansi terkait dalam upaya mitigasi dan penanggulangan dampak bencana banjir. Untuk penelitian selanjutnya, disarankan menggunakan jumlah data yang lebih besar, menambahkan variabel pendukung lainnya, serta membandingkan algoritma KNN dengan metode klasifikasi lain guna memperoleh hasil yang lebih optimal

REFERENCES

[1] F. Seftiani and R. Handawati, “Pemetaan Tingkat Bahaya Banjir Menggunakan Metode Penginderaan Jauh Di Kecamatan Jatinegara Kota Administrasi Jakarta Timur,” vol. 2, no. 2, pp. 48–55, 2023.

ISSN: 3110-0864 |(Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 108-118
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.103

[2] B. Mufarida, "BNPB Catat 74 Bencana Landa Tanah Air di Awal 2025, Banjir Mendominasi." Accessed: Jan. 19, 2026. [Online]. Available: https://nasional.sindonews.com/read/1516047/15/bnpb-catat-74-bencana-landa-tanah-air-di-awal-2025-banjir-mendominasi-1736820077?

[3] Z. Sembiring *et al.*, "Didaktik : Jurnal Ilmiah PGSD FKIP UNIVERSITAS MANDIRI ISSN Cetak : 2477-5673 ISSN Online : 2614-722X Volume 11 Nomor 02 , Juni 2025 Didaktik : Jurnal Ilmiah PGSD FKIP UNIVERSITAS MANDIRI ISSN Cetak : 2477-5673 ISSN Online : 2614-722X Volume 11 Nomor 02 , Juni 2025," vol. 11, pp. 366–371, 2025.

[4] A. Penyebab, D. A. N. Dampak, B. Di, and K. Anggrung, "terjadi di Indonesia, termasuk di Kelurahan Anggrung," vol. 5, no. 6, 2024.

[5] Elfa Harahap, "Kabupaten-Kota yang Terdampak Banjir Longsor di Sumut Bertambah," Mistar.ID. Accessed: Jan. 19, 2026. [Online]. Available: https://mistar.id/news/medan/kabupatenkota-yang-terdampak-banjir-longsor-di-sumut-bertambah

[6] N. Anggraini, B. Pangaribuan, A. P. Siregar, and G. Sintampalam, "ANALISIS PEMETAAN DAERAH RAWAN BANJIR DI KOTA MEDAN TAHUN 2020," vol. 4, no. 2, pp. 27–33, 2021.

[7] N. Bayes *et al.*, "Perbandingan Teknik Klasifikasi Data Mining untuk Penentuan Jenis Jamur Beracun," vol. 5, no. 3, pp. 23–31, 2024.

[8] S. Setyaningtyas, B. I. Nugroho, and Z. Arif, "TINJAUAN PUSTAKA SISTEMATIS PADA DATA MINING : STUDI KASUS ALGORITMA K-MEANS CLUSTERING," vol. 10, no. 2, pp. 52–61, 2022.

[9] N. A. Ramadhani and H. A. Rosyid, "Algoritma – Algoritma Data Mining untuk Klasifikasi Data," vol. 2, no. 12, pp. 550–556, 2022, doi: 10.17977/um068v2i122022pxxx-xxx.

[10] J. M. Informatika, S. I. Misi, C. P. Inayati, S. Lestanti, and S. Budiman, "RANCANG BANGUN APLIKASI PENGENALAN WAJAH," vol. 8, pp. 1–12, 2025.

[11] P. Data *et al.*, "Jurnal PIPA: Pendidikan Ilmu Pengetahuan Alam," vol. 05, no. 01, pp. 42–45, 2024.

[12] A. Naïve, "Comparison of Data Mining Methods for Predition of Floods with Naïve Bayes and KNN Algorithm Perbandingan Metode Data Mining untuk Prediksi Banjir Dengan," pp. 40–48, 2022.

[13] P. E. Putra, M. A. Amrullah, Y. H. Fauzi, and R. F. Saputri, "Penerapan Algoritma C4 . 5 untuk Klasifikasi Tingkat Korban Banjir di Indonesia," vol. 3, no. c, pp. 1–7, 2025.

[14] S. M. Natzir, "Perbandingan Kinerja Model Pembelajaran Mesin dalam Prediksi Banjir menggunakan KNN , Naive Bayes , dan Random Forest," vol. 14, no. c, pp. 59–64, 2023.

[15] J. Akbar, M. Ali, and S. Yudono, "Water Level Classification for Detect Flood Disaster Status using KNN and SVM," vol. 13, pp. 298–302, 2024.

[16] A. Khusaeri, S. Ilham, D. Nurhasanah, D. Delpidat, and B. N. Sari, "Algoritma c4.5 untuk pemodelan daerah rawan banjir studi kasus kabupaten karawang jawa barat," vol. 9, pp. 132–136, 2017.

[17] M. W. Martadiansyah, A. Ghufron, R. A. Hidayah, and D. Salzabila, "Perbandingan Algoritma C4 . 5 dengan Naive Bayes Untuk Klasifikasi Curah Hujan," vol. 3, no. c, pp. 8–17, 2025.

[18] P. P. Haryoto, H. Okprana, and I. S. Saragih, "Algoritma C4 . 5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru," vol. 2, no. 5, pp. 358–364, 2021.

[19] E. Pitaloka *et al.*, "Penerapan Machine Learning untuk Prediksi Bencana Banjir," vol. 01, 2024, doi: 10.21456/vol14iss1pp62-76.

[20] A. Rizky, S. Achmadi, A. F. Setiawan, and T. Informatika, "PENERAPAN DATA MINING DENGAN ALGORITMA KNN UNTUK MENENTUKAN TINGKAT KERUSAKAN DRAINASE KOTA JOMBANG," vol. 8, no. 5, pp. 8471–8478, 2024.