
ANALISIS TITIK GENANGAN AIR AKIBAT BANJIR MENGGUNAKAN METODE RANDOM FOREST DI KECAMATAN LEDO

Christian Cahyaningtyas¹⁾, Yuliana²⁾, Noviyanti P³⁾, Totok Victor Didik Saputro⁴⁾,
Sistem Informasi¹⁾, Teknologi Informasi^{2,3)}, Pendidikan Guru Sekolah Dasar⁴⁾,

Institut Shanti Bhuana

Jl. Bukit Karmel No.01, Bengkayang, Kalimantan Barat

e-mail: christi@shantibhuana.ac.id¹⁾, yuliana@shantibhuana.ac.id²⁾, noviyanti@shantibhuana.ac.id³⁾,
totok.victor@shantibhuana.ac.id⁴⁾

Abstrak

Banjir merupakan salah satu bencana alam yang sering terjadi di berbagai wilayah Indonesia, termasuk Kabupaten Bengkayang. Tingginya curah hujan pada musim penghujan dapat menyebabkan peningkatan debit air yang dapat menimbulkan luapan sungai maupun badan air lainnya sehingga memicu terjadinya genangan. Salah satu upaya yang dapat dilakukan untuk mendukung mitigasi bencana banjir adalah melalui identifikasi dan pemetaan titik-titik genangan air. Oleh karena itu, penelitian ini bertujuan untuk memetakan titik genangan air di Kecamatan Ledo dengan memanfaatkan data penginderaan jauh, Sistem Informasi Geografis, serta algoritma *Random Forest*. Sebanyak 244 titik sampel dikumpulkan dan dilakukan ekstraksi nilai variabel sebagai dasar penyusunan dataset. Selanjutnya, dataset dibagi menjadi data latih 80% dan data uji 20% untuk proses pemodelan dan evaluasi. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* mampu mengklasifikasikan titik genangan air dengan tingkat akurasi sebesar 98%. Berdasarkan *confusion matrix*, seluruh titik yang tidak tergenang berhasil diidentifikasi dengan tepat, sedangkan hanya satu titik genangan yang mengalami kesalahan klasifikasi. Analisis *feature importance* menunjukkan bahwa variabel VV memiliki kontribusi paling besar dalam proses klasifikasi. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma *Random Forest* memiliki kinerja yang sangat baik dalam memprediksi titik genangan air dan menjadi salah satu pendukung pengambilan keputusan dalam upaya mitigasi dan penanggulangan banjir.

Kata kunci : Penginderaan Jauh; *Random Forest*; Sistem Informasi Geografis; Titik Genangan

1. Pendahuluan

Banjir merupakan bencana alam yang sering terjadi di Indonesia. Disaat musim hujan dengan curah hujan yang tinggi, otomatis debit air akan meningkat dan mengakibatkan meluapnya sungai atau badan air [1][2]. Kondisi tersebut mengakibatkan beberapa daerah yang seharusnya tetap kering menjadi tergenang oleh air akibat ketidakmampuan dalam menampung air. Banjir ini sudah menjadi masalah setiap tahunnya yang dapat memberikan dampak yang besar bagi masyarakat, baik secara sosial, ekonomi maupun lingkungan [3,4]. Dampak tersebut tidak hanya berupa kerusakan secara fisik seperti rusaknya infrastruktur, terganggunya aktivitas masyarakat dan hilangnya harta benda namun juga berdampak pada aspek kesehatan, pendidikan hingga pada kondisi psikologis para korban.

Perubahan yang dimulai dari tata guna lahan, meningkatnya urbanisasi dan menurunnya kapasitas daerah resapan air juga memperparah terjadinya banjir di banyak daerah yang ada di Indonesia. Selain itu, perubahan iklim yang disebabkan oleh intensitas curah hujan ekstrim semakin sering terjadi di beberapa waktu terakhir. Hal ini menjadi salah satu faktor meningkatnya frekuensi dan skala kejadian banjir di berbagai wilayah di Indonesia [5].

Salah satu wilayah yang mempunyai intensitas curah hujan yang cukup ekstrim adalah wilayah Kabupaten Bengkayang sehingga rawan terjadinya banjir. Terdapat beberapa kecamatan yang terendam oleh banjir pada awal tahun 2025 [6]. Kejadian banjir ini menyebabkan ribuan rumah tergenang, aktivitas masyarakat terganggu, serta memicu kerugian sosial dan ekonomi yang cukup besar. Berdasarkan laporan pemerintah dan media lokal, banjir terparah yang terjadi di Kabupaten Bengkayang berada di Kecamatan Ledo serta ini merupakan banjir terbesar yang terjadi di Kecamatan Ledo [7]. Kondisi ini menunjukkan bahwa wilayah tersebut memiliki tingkat kerawanan banjir yang tinggi dan memerlukan upaya mitigasi berbasis data yang lebih kuat dan terukur.

Jika bencana banjir tidak segera diatasi dengan baik dan benar maka akan semakin banyak kerusakan dan kerugian yang akan ditimbulkan serta dapat membahayakan masyarakat yang terdampak. Oleh karena itu, langkah penting dalam penanggulangan bencana banjir adalah melalui analisis dan identifikasi titik-titik genangan air. Upaya dalam meminimalkan terjadinya bencana banjir dengan cara menganalisis titik genangan air. Berdasarkan titik-titik genangan air ini dapat dipetakan wilayah yang terjadi banjir, menganalisis faktor-faktor penyebab terjadinya genangan, serta juga dapat memberikan informasi spasial yang dapat dimanfaatkan untuk mitigasi dan perencanaan penanggulangan banjir.

Dalam beberapa tahun terakhir, pemanfaatan teknologi penginderaan jauh, sistem informasi geografis (SIG) dan metode *machine learning* telah berkembang pesat dalam penelitian kebencanaan, termasuk dalam pemetaan bencana banjir [8,9,10,11]. Metode *Random Forest* menjadi salah satu algoritma paling banyak yang digunakan karena memiliki kemampuan yang baik dalam memodelkan hubungan non linear antara variabel penyebab banjir dan memberikan akurasi prediksi yang baik [12,13]. Di Indonesia, penelitian pemetaan kerawanan banjir berbasis *Random Forest* semakin berkembang. Penelitian yang dilakukan di Sumedang menunjukkan bahwa *Random Forest* mampu memetakan

kerentanan banjir dengan akurasi yang tinggi menggunakan data topografi dan klimatologi [14]. Penelitian lainnya di Jakarta menunjukkan bahwa metode *Random Forest* menghasilkan akurasi hingga 99% dalam klasifikasi level banjir berdasarkan data historis [15]. Selain itu juga penelitian di Pontianak, wilayah yang secara geografis dengan kabupaten Bengkayang menunjukkan bahwa *Random Forest* paling optimal dalam memprediksi banjir rob dibandingkan metode lain seperti KNN dan SVM [16]. Berdasarkan penelitian-penelitian yang sudah dilakukan sebelumnya, maka penggunaan metode *Random Forest* cocok diterapkan dalam penelitian analisis titik genangan air di Kecamatan Ledo.

Pemanfaatan pendekatan berbasis data spasial, penginderaan jauh dan algoritma *Random Forest*, penelitian ini diharapkan mampu memberikan hasil pemetaan titik genangan air yang lebih akurat dan dapat digunakan sebagai dasar dalam upaya mitigasi serta perencanaan penanggulangan banjir. Selain itu, penelitian ini dapat mendukung pemerintah daerah dalam pengambilan keputusan berbasis data terutama terkait perencanaan tata ruang, pembangunan infrastruktur dan strategi pengurangan risiko bencana.

2. Landasan Teori

Banjir

Banjir merupakan peristiwa tergenangnya suatu wilayah yang pada kondisi normal tidak tertutup air akibat meluapnya air dari sungai, danau, saluran drainase, maupun akibat tingginya curah hujan yang melebihi kapasitas tampung suatu daerah [17]. Banjir termasuk bencana hidrometeorologi yang dipengaruhi oleh faktor alam dan aktivitas manusia. Terdapat beberapa faktor yang dapat memengaruhi banjir seperti intensitas curah hujan yang tinggi, penggunaan lahan, kemiringan lereng dan lain sebagainya. Genangan air yang terjadi akibat banjir dapat menimbulkan berbagai dampak, seperti kerusakan infrastruktur, gangguan aktivitas masyarakat, kerugian ekonomi, hingga masalah kesehatan dan lingkungan. Oleh karena itu, identifikasi serta pemetaan wilayah genangan menjadi salah satu langkah penting dalam mendukung upaya mitigasi dan pengurangan risiko bencana banjir.

Random Forest

Random Forest merupakan algoritma *machine learning* yang termasuk dalam metode *ensemble learning*. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dari subset data yang dipilih secara acak, kemudian menggabungkan hasil prediksi dari seluruh pohon untuk menghasilkan keputusan akhir [18]. Pada kasus klasifikasi, hasil akhir ditentukan berdasarkan suara terbanyak (*majority voting*) dari seluruh pohon yang terbentuk.

Keunggulan *Random Forest* terletak pada kemampuannya dalam menangani data berukuran besar, mengurangi risiko *overfitting*, serta mampu memodelkan hubungan yang kompleks dan nonlinier antarvariabel. Selain itu, algoritma ini juga dapat mengukur tingkat kepentingan setiap variabel melalui *feature importance*, sehingga dapat diketahui variabel mana yang memberikan kontribusi terbesar terhadap hasil klasifikasi. Karena karakteristik tersebut, *Random Forest* banyak digunakan dalam berbagai bidang, termasuk analisis lingkungan, penginderaan jauh, pemetaan kerawanan bencana, dan klasifikasi tutupan lahan.

Sistem Informasi Geografis

Sistem Informasi Geografis (SIG) merupakan sistem berbasis komputer yang digunakan untuk mengumpulkan, menyimpan, mengelola, menganalisis, dan menyajikan data yang memiliki referensi geografis [19]. SIG memungkinkan pengguna untuk mengintegrasikan berbagai jenis data spasial dan atribut sehingga dapat digunakan untuk menghasilkan informasi yang mendukung proses pengambilan keputusan. Komponen utama SIG meliputi perangkat keras (*hardware*), perangkat lunak (*software*), data, metode, dan sumber daya manusia. Dalam bidang kebencanaan, SIG berperan penting dalam pemetaan wilayah terdampak, analisis kerawanan, pemodelan spasial, serta visualisasi hasil analisis dalam bentuk peta tematik. Kemampuan SIG dalam mengolah dan menganalisis data spasial menjadikannya alat yang efektif dalam mendukung mitigasi dan penanggulangan bencana banjir.

Penginderaan Jauh

Penginderaan jauh merupakan ilmu dan teknologi untuk memperoleh informasi mengenai objek, wilayah, atau fenomena di permukaan bumi tanpa melakukan kontak langsung dengan objek yang diamati [17]. Informasi diperoleh melalui sensor yang dipasang pada satelit, pesawat udara, atau wahana lainnya yang merekam energi elektromagnetik yang dipantulkan atau dipancarkan oleh objek di permukaan bumi.

Data penginderaan jauh memiliki cakupan wilayah yang luas, waktu akuisisi yang cepat, dan kemampuan pemantauan secara berkala sehingga banyak digunakan dalam berbagai bidang, termasuk pemetaan sumber daya alam, lingkungan, pertanian, serta kebencanaan. Dalam penelitian banjir, citra satelit digunakan untuk mengidentifikasi wilayah genangan, menganalisis perubahan kondisi permukaan, dan menghasilkan informasi spasial yang mendukung proses mitigasi bencana. Salah satu data penginderaan jauh yang banyak digunakan adalah citra Sentinel-1 dan Sentinel-2. Sentinel-1 menggunakan sensor radar yang mampu merekam kondisi permukaan bumi meskipun tertutup awan, sedangkan Sentinel-2 menggunakan sensor optik yang dapat dimanfaatkan untuk menghitung berbagai indeks spektral, termasuk indeks air.

Normalized Difference Water Index (NDWI)

Normalized Difference Water Index (NDWI) merupakan indeks spektral yang digunakan untuk mendeteksi keberadaan air di permukaan bumi [20]. NDWI memanfaatkan perbedaan reflektansi antara kanal hijau (Green) dan kanal inframerah dekat (*Near Infrared/NIR*). Nilai NDWI yang tinggi umumnya menunjukkan keberadaan badan air atau wilayah yang memiliki kandungan air tinggi, sedangkan nilai rendah menunjukkan vegetasi atau lahan kering. NDWI banyak digunakan dalam pemetaan genangan banjir karena mampu membedakan area yang tergenang air dengan area non-air secara efektif. Indeks ini juga sering dikombinasikan dengan data radar untuk meningkatkan akurasi identifikasi genangan.

Backscatter Sentinel-1 (VV)

Backscatter merupakan nilai pantulan gelombang radar yang direkam oleh sensor *Synthetic Aperture Radar* (SAR). Pada citra Sentinel-1, salah satu polarisasi yang umum digunakan adalah VV (*Vertical Transmit–Vertical Receive*), yaitu gelombang radar yang dipancarkan dan diterima kembali dalam orientasi vertikal. Nilai *backscatter* sangat dipengaruhi oleh kondisi permukaan bumi. Area yang tergenang air cenderung memiliki nilai *backscatter* yang lebih rendah dibandingkan permukaan kasar atau vegetasi karena permukaan air memantulkan energi radar menjauhi sensor. Oleh karena itu, nilai VV sering digunakan sebagai indikator penting dalam deteksi dan pemetaan genangan banjir [20].

Elevasi dan Kemiringan Lereng (Slope)

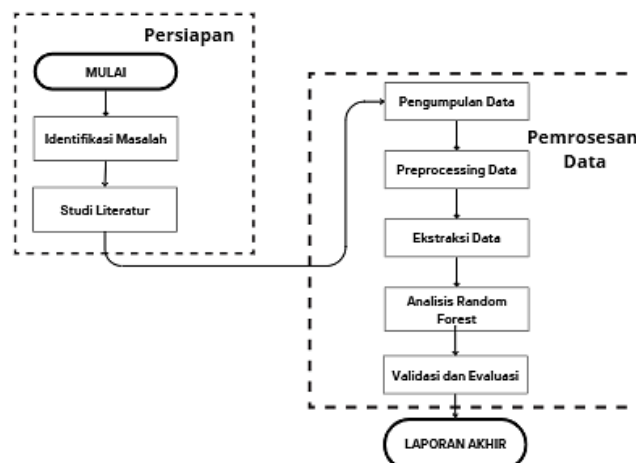
Elevasi merupakan tinggi suatu lokasi terhadap permukaan laut rata-rata. Elevasi berpengaruh terhadap arah aliran air dan potensi terjadinya genangan. Wilayah dengan elevasi rendah umumnya memiliki risiko genangan yang lebih tinggi dibandingkan wilayah dengan elevasi tinggi. Kemiringan lereng (*slope*) menunjukkan tingkat kecuraman permukaan tanah. Daerah dengan lereng landai cenderung memiliki kecepatan aliran air yang lebih rendah sehingga berpotensi mengalami genangan lebih lama. Sebaliknya, daerah dengan lereng curam memungkinkan air mengalir lebih cepat sehingga risiko genangan relatif lebih kecil. Elevasi dan kemiringan lereng merupakan faktor topografi yang berpengaruh terhadap arah aliran air dan potensi terbentuknya genangan, sehingga sering digunakan dalam pemodelan kerawanan banjir [21].

Curah Hujan

Curah hujan merupakan jumlah air hujan yang jatuh pada suatu wilayah dalam periode tertentu dan biasanya dinyatakan dalam satuan milimeter (mm) [22]. Curah hujan merupakan salah satu faktor utama yang mempengaruhi terjadinya banjir dan genangan air. Semakin tinggi intensitas dan durasi hujan, semakin besar volume air yang harus ditampung oleh sungai, drainase, maupun daerah resapan. Dalam analisis genangan banjir, data curah hujan digunakan untuk menggambarkan kondisi hidrologis suatu wilayah dan menjadi salah satu variabel penting dalam pemodelan spasial maupun *machine learning*.

3. Metode Penelitian

Penelitian ini dilakukan di Kecamatan Ledo, Kabupaten Bengkayang, Kalimantan Barat. Kecamatan Ledo merupakan wilayah yang mengalami banjir terparah pada tahun 2025 yang ketinggian airnya mencapai lebih dari 1,5 meter [23]. Secara geografis, Kecamatan Ledo memiliki topografi beragam, ada yang berada pada dataran rendah hingga wilayah perbukitan, sehingga terdapat variasi tingkat genangan air. Adapun alur dalam penelitian ini disajikan pada gambar 1.

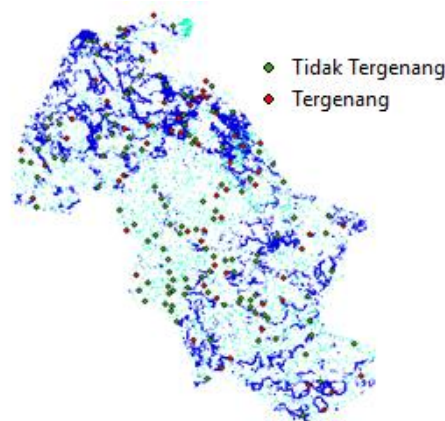


Gambar 1. Alur Penelitian

Pada gambar 1 merupakan diagram alur pelaksanaan penelitian yang akan dilakukan. Pada tahap persiapan dimulai dengan identifikasi masalah untuk menentukan isu utama yang akan diselesaikan dalam penelitian yaitu tentang analisis titik genangan banjir di Kecamatan Ledo. Selanjutnya studi literatur yaitu dilakukan dengan mencari referensi yang

relevan dengan topik penelitian. Hasil dari tahap persiapan ini menjadi dasar dalam menentukan variabel, data yang akan digunakan serta metode analisis yang cocok untuk diterapkan pada penelitian. Tahapan selanjutnya adalah tahapan pemrosesan data. Tahap ini diawali dengan pengumpulan data yaitu data citra satelit sentinel 1, sentinel 2, curah hujan, elevasi, penggunaan lahan dan kemiringan lereng. Data citra sentinel 1 untuk mencari nilai polarisasi VV (Vertikal - Vertikal), sentinel 2 untuk mencari nilai NDWI (Normalized Difference Water Index) serta data elevasi diunduh dari *platform copernicus* melalui laman *google earth engine* [24], data curah hujan diunduh dari data portal CHRS [25], sedangkan data penggunaan lahan dan kemiringan lereng diunduh dari platform *im carto* sebagai sumber data spasial pendukung [20,21]. Data yang sudah dikumpulkan sebelum dapat digunakan harus dilakukan *preprocessing* data terlebih dahulu. Pada tahap *preprocessing* ini menggunakan bantuan perangkat lunak ArcGis. Proses *preprocessing* ini meliputi pemotongan citra sesuai wilayah yang diteliti, penyamaan titik koordinat dan *cell size, resampling*, penghapusan *noise* dan lain sebagainya. Pada tahap *preprocessing*, masing-masing variabel yang digunakan memerlukan *preprocessing* yang berbeda-beda sesuai dengan karakteristiknya. Setelah data sudah siap maka dilanjutkan ke tahap selanjutnya yaitu tahapan ekstraksi data. Pada tahap ini dimulai dengan membuat titik sampel pada area penelitian terlebih dahulu. Selanjutnya dilakukan ekstraksi data yaitu proses pengambilan nilai atau fitur dari masing-masing variabel untuk dijadikan dataset *input* algoritma. Pada tahap ini, titik sampel kemudian diklasifikasikan menjadi 2 kategori yaitu tergenang dan tidak tergenang berdasarkan dengan interpretasi visual citra satelit. Pada tahapan ini juga dilakukan pembersihan data yang kosong atau tidak valid agar data tersebut dapat langsung digunakan untuk analisis lebih lanjut. Selanjutnya *eksport* data tersebut untuk dijadikan dataset dan digunakan dalam analisis dengan metode *Random Forest*. Pada tahap ini, model dibangun untuk mempelajari pola dari data pelatihan dengan memanfaatkan sejumlah pohon keputusan yang dikombinasikan untuk menghasilkan prediksi yang lebih stabil. Proses pemodelan dilakukan dengan menggunakan bantuan *google colab* dimana dataset dibagi menjadi 2 yaitu data latih (80%) dan data uji (20%). Setelah pemodelan terbentuk maka dilakukan evaluasi dan validasi data untuk mengukur performa model. Hasil dari keseluruhan proses dari pemrosesan ini selanjutnya dirangkum dan disajikan dalam bentuk artikel.

Adapun hasil dari *preprocessing* data berupa peta sebaran titik sampel genangan air yang disajikan pada gambar 2. Dimana titik sampel telah diklasifikasikan menjadi 2 kategori yaitu tergenang yang ditandai dengan titik warna merah dan tidak tergenang yang ditandai dengan warna hijau. Sebaran titik yang merata pada wilayah studi menandakan bahwa *preprocessing* telah menghasilkan dataset yang representatif, mencakup variasi kondisi lingkungan seperti dataran rendah, dataran tinggi, pemukiman dan lain sebagainya. Selain itu data yang kosong, bermasalah atau tidak valid telah dihapus pada tahapan ini sehingga dataset siap digunakan tanpa resiko menghasilkan bias pada proses analisis lebih lanjut. Sehingga hasil dari *preprocessing* ini dapat dikatakan menghasilkan dataset bersih, terstandarisasi dan siap ekstraksi dan diolah dengan menggunakan algoritma *Random Forest*.



Gambar 2. Sampel Titik Genangan

4. Hasil Penelitian

Berdasarkan hasil ekstraksi data pada titik-titik sampel yang telah ditentukan, diperoleh sejumlah variabel yang digunakan sebagai masukan dalam proses pemodelan genangan air. Setiap titik sampel diekstraksi nilai-nilainya dari berbagai sumber data, seperti elevasi, penggunaan lahan, indeks vegetasi air (NDWI), nilai *backscatter* radar Sentinel-1 (VV), kemiringan lereng, dan curah hujan. Variabel-variabel tersebut dipilih karena memiliki pengaruh terhadap terbentuknya genangan air dan digunakan sebagai fitur untuk melatih model *Random Forest*. Selain itu, setiap titik diberi label status tergenang atau tidak tergenang berdasarkan interpretasi visual citra satelit. Adapun ringkasan hasil ekstraksi data untuk beberapa titik sampel ditampilkan pada Tabel 1 berikut.

Tabel 1. Dataset titik genangan air

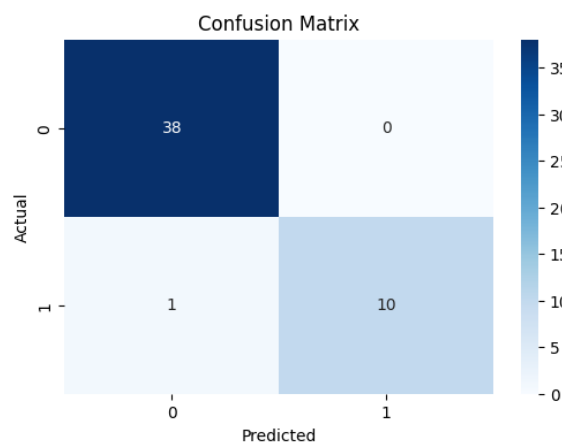
Elevasi	PL	NDWI	VV	Slope	CH	Status
---------	----	------	----	-------	----	--------

15	2	0,09223	0,329761	235406	649	1
17	2	0,099458	-0,279745	305038	589	0
17	4	0,005454	5,36934	383290	586	1
18	2	0,010664	-0,685667	561669	663	0
19	2	0,107216	2,17443	158302	587	1
20	2	0,166508	-0,020047	332872	600	0
20	5	0,080615	0,539032	207335	586	1
20	2	0,088342	0,033229	223347	589	1
21	6	0,207203	-0,180956	798191	613	0
21	2	0,025804	-0,649964	256480	650	0
dst

Tabel 1 menunjukkan hasil ekstraksi data dari titik sampel yang digunakan sebagai *input* pada proses pemodelan. Setiap baris merepresentasikan satu titik sampel dengan nilai variabel yang diperoleh dari berbagai sumber data spasial. Variabel Elevasi menunjukkan ketinggian lokasi berdasarkan data DEM, sedangkan PL merupakan nilai klasifikasi penggunaan lahan. Nilai NDWI berasal dari citra Sentinel-2 dan merepresentasikan indeks kandungan air pada permukaan. Variabel VV merupakan nilai *backscatter* dari citra Sentinel-1 yang sensitif terhadap kondisi permukaan basah. Sementara itu, *Slope* menggambarkan tingkat kemiringan lereng dan CH adalah nilai curah hujan pada lokasi tersebut. Kolom Status digunakan sebagai label klasifikasi, dengan nilai 1 menunjukkan titik yang tergenang dan 0 menunjukkan titik tidak tergenang berdasarkan interpretasi visual citra. Dataset ini menjadi dasar dalam proses pelatihan dan pengujian model *Random Forest* untuk memprediksi titik-titik genangan air di wilayah studi.

Dataset yang digunakan terdiri dari total 244 titik sampel, yang masing-masing berisi nilai variabel hasil ekstraksi spasial yang telah disajikan pada tabel 1. Dataset ini kemudian dibagi menjadi dua bagian untuk keperluan pemodelan, yaitu data *training* sebanyak 80% dan data testing sebanyak 20%. Pembagian ini bertujuan untuk memastikan bahwa model *Random Forest* dapat dilatih menggunakan sebagian besar data, sementara kinerjanya dievaluasi secara obyektif menggunakan data yang tidak pernah dilihat model sebelumnya. Proporsi ini juga cukup umum digunakan untuk menjaga keseimbangan antara kemampuan pembelajaran model dan validasi performanya.

Untuk mengevaluasi performa model *Random Forest* dalam mengklasifikasikan titik tergenang dan tidak tergenang, dilakukan pengujian menggunakan 20% data uji yang sebelumnya tidak dilibatkan dalam proses pelatihan. Evaluasi dilakukan dengan menggunakan *confusion matrix*, yang berfungsi untuk menggambarkan perbandingan antara hasil prediksi model dengan hasil interpretasi visual citra satelit. *Confusion matrix* memberikan informasi yang lebih detail mengenai jumlah prediksi yang tepat maupun yang keliru untuk masing-masing kelas, sehingga memudahkan dalam menilai kemampuan model dalam mendeteksi titik genangan secara akurat. Adapun hasil *confusion matrix* untuk model *Random Forest* ditunjukkan pada Gambar berikut :



Gambar 3. *Confusion Matrix*

Berdasarkan dari gambar 3, *confusion matrix* pada data uji menunjukkan bahwa model *Random Forest* memiliki performa klasifikasi yang sangat baik dalam membedakan titik tergenang dan tidak tergenang. Dari total 49 sampel pengujian, model berhasil mengidentifikasi seluruh 38 titik yang sebenarnya tidak tergenang dengan tepat (*true negative*), tanpa menghasilkan satupun kesalahan prediksi genangan pada lokasi yang kering (*false positive* = 0). Pada kelas tergenang, model mampu memprediksi dengan benar 10 titik genangan (*true positive*), sementara hanya satu titik yang terlewat atau salah diklasifikasikan sebagai tidak tergenang (*false negative*). Hasil ini menunjukkan bahwa model memiliki sensitivitas yang tinggi dalam mendeteksi genangan serta sangat andal dalam menghindari prediksi positif yang keliru. Secara keseluruhan, pola pada *confusion matrix* menggambarkan kinerja model yang sangat kuat dan stabil dalam proses klasifikasi genangan air.

	precision	recall	f1-score	support
0	0.97	1.00	0.99	38
1	1.00	0.91	0.95	11
accuracy			0.98	49
macro avg	0.99	0.95	0.97	49
weighted avg	0.98	0.98	0.98	49

Gambar 4. Classification Report

Berdasarkan dari gambar 4, hasil evaluasi model menunjukkan kinerja klasifikasi yang sangat baik, dengan akurasi keseluruhan mencapai 98%. Pada kelas 0, model memiliki *precision* 0,97 dan *recall* 1,00, yang berarti hampir seluruh sampel kelas ini dapat teridentifikasi dengan benar. Sementara itu, kelas 1 menghasilkan *precision* sempurna sebesar 1,00 dan *recall* 0,91, menunjukkan bahwa meskipun semua prediksi kelas 1 benar, masih terdapat beberapa sampel kelas 1 yang tidak terdeteksi oleh model. Nilai rata-rata *macro* dan *weighted* yang tinggi juga mengindikasikan performa yang stabil di seluruh kelas.

Tabel 2. Feature Important

Variabel	Important
VV	50,92%
NDWI	16,28%
Slope	15,04%
Elevation	10,97%
Rainfall	5,45%
Land Use	1,34%

Sedangkan berdasarkan tabel 2 *feature important* dari variabel menunjukkan bahwa fitur VV memiliki pengaruh terbesar terhadap proses klasifikasi dengan kontribusi sekitar 50,92%, diikuti NDWI (16,28%), kemiringan lereng atau *slope* (15,04%), elevasi (10,97%), curah hujan atau CH (5,45%), dan penggunaan lahan atau PL (1,34%). Hal ini menunjukkan bahwa informasi radar (VV) dan indeks air (NDWI) menjadi faktor paling dominan dalam membedakan kelas pada dataset tersebut.

5. Kesimpulan

Berdasarkan hasil ekstraksi data dari 244 titik sampel, diperoleh enam variabel utama elevasi, penggunaan lahan, NDWI, *backscatter* VV, *slope*, dan curah hujan yang berperan sebagai fitur dalam pemodelan genangan air menggunakan algoritma *Random Forest*. Variabel-variabel tersebut berhasil menggambarkan kondisi fisik dan lingkungan yang mempengaruhi terbentuknya genangan air, dengan label status genangan ditetapkan melalui interpretasi visual citra satelit. Proses pemodelan dilakukan dengan membagi dataset menjadi 80% data latih dan 20% data uji, sehingga memungkinkan evaluasi performa model secara objektif. Hasil pengujian menunjukkan bahwa model *Random Forest* memiliki kinerja yang sangat baik, dengan akurasi mencapai 98%. *Confusion matrix* memperlihatkan bahwa model mampu mengklasifikasikan seluruh titik tidak tergenang secara *tepat* dan hanya keliru pada satu titik genangan. Temuan ini mengindikasikan bahwa model memiliki tingkat presisi dan sensitivitas yang tinggi dalam mendeteksi genangan air. Evaluasi melalui *classification report* juga menunjukkan nilai *precision*, *recall*, dan *f1-score* yang konsisten tinggi di kedua kelas. Analisis *feature importance* mengungkapkan bahwa variabel VV menjadi penentu terbesar dalam proses klasifikasi, diikuti NDWI, *slope*, dan elevasi. Dominasi variabel terkait kondisi permukaan dan kandungan air menunjukkan bahwa kombinasi informasi radar dan indeks air sangat efektif dalam membedakan antara titik yang tergenang dan tidak tergenang. Secara keseluruhan, model *Random Forest* terbukti andal dan dapat digunakan sebagai alat prediksi yang akurat untuk pemetaan genangan air di wilayah studi.

Daftar Pustaka

- [1] M. Yusup, P. I. S. Tarigan, K. Noviansah, R. Ridwana, and S. A. Aliyan, "Identifikasi Genangan Banjir Menggunakan Sentinel 1 dan Korelasinya Dengan Kerawanan Banjir Di Kabupaten Barito Selatan," *Geo Image (Spat.)*, vol. 12, no. 1, pp. 62–70, 2023.
- [2] A. P. Ambarini, F. Bioresita, and N. Hayati, "Analisis Sebaran Spasial Genangan Banjir Terkait Tutupan Lahan di Kota Banjarmasin , Kalimantan Selatan Menggunakan Citra Sentinel 1 dan 2," *Geoid (Journal Geod. geomatics)eomatics*), vol. 19, no. 2, pp. 326–337, 2024.
- [3] F. A. Budiarto and F. Bioresita, "Pemanfaatan Citra Sentinel-1 SAR dan Metode Change Detection Approach Untuk Analisis Sebaran Spasial Wilayah Banjir dan Area Terdampak (Studi Kasus: Banjir Kabupaten Aceh Utara 2022)," *JGISE J. Geospatial Inf. Sci. Eng.*, vol. 6, no. 2, pp. 153–162, 2023, doi: 10.22146/jgise.87585.
- [4] R. A. Kenranto, H. Hidayat, and F. Bioresita, "Analisis Genangan Banjir Terhadap Penutup Lahan di Wilayah Tangerang Menggunakan Data Citra Sentinel-1 dan Sentinel-2," *JGISE J. Geospatial Inf. Sci. Eng.*, vol. 7, no. 1, pp. 14–22, 2024, doi: 10.22146/jgise.87579.

- [5] Suhadi, F. Maburoh, A. Wiyanto, and Ikra, “Analisis fenomena perubahan iklim terhadap curah hujan ekstrim,” *Opt. J. Pendidik. Fis.*, vol. 7, no. 1, pp. 94–100, 2023.
- [6] Narwati, “11 Kecamatan di Bengkayang Kalbar Dilanda Bencana Banjir,” *Antaraneews.com*, Bengkayang, 2025. [Online]. Available: <https://kalbar.antaraneews.com/berita/621470/11-kecamatan-di-bengkayang-kalbar-dilanda-bencana-banjir>
- [7] D. Eno, “12 Ribu Warga Bengkayang Terdampak Banjir, 3.468 Rumah Terendam,” *suarakalbar.com*, 2025. [Online]. Available: <https://www.suarakalbar.co.id/2025/02/12-ribu-warga-bengkayang-terdampak-banjir-3-468-rumah-terendam/>
- [8] S. N. Aziza, L. Somantri, and I. Setiawan, “Analisis Pemetaan Tingkat Rawan Banjir Di Kecamatan Bontang Barat Kota Bontang Berbasis Sistem Informasi Geografis,” *J. Pendidik. Geogr. Undiksha*, vol. 9, no. 2, pp. 109–120, 2021.
- [9] C. Cahyaningtyas, M. Sari, and D. S. Pranata, “Pengembangan Model SIG untuk Pemetaan Daerah Risiko Rawan Bencana Banjir di Kabupaten Bengkayang,” *Jutisi J. Ilm. Tek. Inf. dan Sist. Inf.*, vol. 14, no. 1, pp. 60–69, 2024.
- [10] P. O. Andewi, K. A. Seputra, K. Y. E. Aryanto, L. Joni, and E. Dewi, “Integrasi Teknologi Penginderaan Jauh dan Machine Learning Pada Web GIS Untuk Pemetaan Potensi Banjir,” *J. Pendidik. Teknol. dan Kejur.*, vol. 22, no. 1, pp. 12–23, 2025.
- [11] H. Farhadi and M. Najafzadeh, “Flood Risk Mapping By Remote Sensing Data and Random Forest Technique,” *Water MDPI*, vol. 13, no. 3115, pp. 1–25, 2021, doi: 10.3390/w13213115.
- [12] A. S. Shaik, N. Shaik, and D. C. K. Priya, “Predictive Modeling in Remote Sensing Using Machine Learning Algorithms,” *Int. J. Curr. Sci. Res. Rev.*, vol. 07, no. 06, pp. 4116–4123, 2024, doi: 10.47191/ijcsrr/v7-i6-62.
- [13] F. Rahmasari, M. Rifany, T. Priharyanto, and R. Kurniawan, “Perbandingan Algoritma Machine Learning dalam Klasifikasi Status Banjir di Sumatera Utara,” *Semin. Nas. Sains Data 2024 (SENADA 2024)*, pp. 307–318, 2024.
- [14] R. D. Ismanto, H. L. Fitriana, J. Manalu, A. A. Purboyo, and I. Prasasti, “Development of Flood-Hazard-Mapping Model Using Random Forest and Frequency Ratio in Sumedang Regency , West Java , Indonesia,” *Geomatics Environ. Eng.*, vol. 17, no. 6, pp. 129–157, 2023.
- [15] L. S. Qamarani and M. Riasetiawan, “Klasifikasi Level Banjir Menggunakan Random Forest dan Support Vector Machine,” *Indones. J. Electron. Instrum. Syst.*, vol. 14, no. 2, pp. 199–208, 2024.
- [16] E. Hermawan, S. D. Panjaitan, and E. F. Ripanti, “Sistem Prediksi Banjir Rob Kota Pontianak Berbasis Machine Learning Menggunakan Framework Streamlit,” *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 10, no. 3, pp. 351–361, 2024.
- [17] H. S. Munawar, A. W. A. Hammad, and S. T. Waller, “Remote Sensing Methods for Flood Prediction: A Review,” *Sensors*, vol. 22, no. 960, pp. 1–21, 2022, doi: 10.3390/s22030960.
- [18] Y. Liao, Z. Wang, C. Lai, and C. Y. Xu, “A Framework on Fast Mapping of Urban Flood Based on a Multi-Objective Random Forest Model,” *Int. J. Disaster Risk Sci.*, vol. 14, no. 2, pp. 253–268, 2023, doi: 10.1007/s13753-023-00481-2.
- [19] G. Schumann, L. Giustarini, A. Tarpanelli, B. Jarihani, and S. Martinis, “Flood Modeling and Prediction Using Earth Observation Data,” *Surv. Geophys.*, vol. 44, no. 5, pp. 1553–1578, 2023, doi: 10.1007/s10712-022-09751-y.
- [20] U. Nazir *et al.*, “Improved Flood Mapping for Efficient Policy Design by Fusion of Sentinel-1, Sentinel-2 and Landsat-9 Imagery to Identify Population and Infrastructure Exposed to Floods,” *Int. Geosci. Remote Sens. Symp.*, vol. 2023-July, pp. 1591–1594, 2023, doi: 10.1109/IGARSS52108.2023.10282530.
- [21] T. Islam, E. B. Zeleke, M. Afroz, and A. M. Melesse, “A Systematic Review of Urban Flood Susceptibility Mapping: Remote Sensing, Machine Learning, and Other Modeling Approaches,” *Remote Sens.*, vol. 17, no. 3, 2025, doi: 10.3390/rs17030524.
- [22] F. Karim, M. A. Armin, D. Ahmedt-Aristizabal, L. Tychsen-Smith, and L. Petersson, “A Review of Hydrodynamic and Machine Learning Approaches for Flood Inundation Modeling,” *Water (Switzerland)*, vol. 15, no. 3, pp. 1–21, 2023, doi: 10.3390/w15030566.
- [23] Admin, “Banjir Terparah di Bengkayang, 5.216 Warga Terdampak dan Mulai Terima Bantuan,” *Suara Nusantara*. [Online]. Available: <https://www.suaranusantara.co.id/2025/01/banjir-terparah-di-bengkayang-5216.html>
- [24] “Google Earth Engine.” [Online]. Available: <https://code.earthengine.google.com/>
- [25] “CHRS Data Portal.” [Online]. Available: <https://chrsdata.eng.uci.edu/>
- [26] “Im.Carto Geo data Base : Digital Elevation Model,” *imcarto.webflo.io*. [Online]. Available: <https://imcarto.webflow.io/gdb>
- [27] “Im.Carto Geo Data Base : Land Use 2017,” *imcarto.webflo.io*. [Online]. Available: <https://imcarto.webflow.io/gdb>