



ISIAN LAPORAN KEMAJUAN PENELITIAN

Petunjuk: Pengusul hanya diperkenankan mengisi di tempat yang telah disediakan sesuai dengan petunjuk pengisian dan tidak diperkenankan melakukan modifikasi template atau penghapusan di setiap bagian.

JUDUL

Tuliskan judul penelitian

Explainable Ensemble Learning untuk Analisis Faktor Dominan Risiko Banjir Perkotaan Berbasis Data Multi-Sumber di Indonesia

RINGKASAN

Isian ringkasan penelitian tidak lebih dari 300 kata yang berisi urgensi, tujuan, metode, dan luaran yang ditargetkan.

Penelitian ini dilaksanakan sebagai upaya untuk menjawab tingginya risiko banjir di wilayah perkotaan Indonesia yang dipengaruhi oleh interaksi kompleks antara faktor hidrometeorologi, karakteristik topografi, serta perubahan tata guna lahan akibat urbanisasi yang pesat. Kompleksitas tersebut menuntut pendekatan analisis risiko banjir yang tidak hanya memiliki tingkat akurasi prediksi yang baik, tetapi juga mampu menjelaskan faktor-faktor dominan penyebab banjir secara transparan dan dapat dipertanggungjawabkan.

Tujuan penelitian ini adalah mengembangkan pendekatan Explainable Ensemble Learning untuk menganalisis faktor dominan risiko banjir perkotaan berbasis data multi-sumber di Indonesia. Metode penelitian yang digunakan meliputi pengumpulan dan integrasi data hidrometeorologi dan geospasial, seperti curah hujan, elevasi, tata guna lahan, indeks vegetasi (NDVI), jarak terhadap sungai, kepadatan drainase, serta tinggi muka air sungai. Data tersebut dianalisis menggunakan beberapa algoritma ensemble learning, yaitu Random Forest, Gradient Boosting, dan XGBoost, untuk memodelkan hubungan nonlinier antar variabel penyebab banjir.

Pada tahap laporan kemajuan ini, penelitian telah mencapai beberapa capaian utama, meliputi studi literatur, pengumpulan dan integrasi data multi-sumber, praproses data, serta pengembangan dan evaluasi awal model ensemble learning. Evaluasi awal menunjukkan bahwa model XGBoost memiliki kinerja prediktif yang lebih baik dibandingkan model lainnya. Untuk meningkatkan keterjelasan model, telah diterapkan pendekatan Explainable Artificial Intelligence menggunakan metode SHapley Additive exPlanations (SHAP) guna mengidentifikasi kontribusi masing-masing variabel terhadap risiko banjir.

Luaran yang ditargetkan dalam penelitian ini adalah model analisis risiko banjir perkotaan berbasis explainable ensemble learning, peta risiko banjir yang menampilkan faktor dominan penyebab banjir, serta satu artikel ilmiah yang sedang dalam proses penyusunan dan penyiapan untuk dipublikasikan pada jurnal nasional terakreditasi SINTA 2. Hasil penelitian pada tahap ini menunjukkan bahwa penelitian berjalan sesuai dengan rencana dan tahapan yang telah ditetapkan dalam proposal

KATA KUNCI

Isian kata kunci maksimal 5 kata yang dipisahkan dengan tanda titik koma (;)

Risiko banjir perkotaan; ensemble learning; explainable artificial intelligence; data multi-sumber; SHAP.

PENDAHULUAN

Pendahuluan penelitian tidak lebih dari 1000 kata yang terdiri dari:

1. Latar belakang dan rumusan permasalahan yang akan diteliti
2. Pendekatan pemecahan masalah
3. *State of the art* dan kebaruan
4. Peta jalan (*road map*) penelitian 5 tahun

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver

1. LATAR BELAKANG DAN RUMUSAN MASALAH

Banjir perkotaan merupakan salah satu permasalahan lingkungan yang memiliki tingkat kompleksitas tinggi dan berdampak signifikan terhadap keberlanjutan wilayah perkotaan di Indonesia. Peningkatan intensitas curah hujan, kondisi topografi dataran rendah, serta perubahan tata guna lahan akibat urbanisasi yang pesat menjadi faktor utama yang berkontribusi terhadap meningkatnya risiko banjir di kawasan perkotaan. Dampak yang ditimbulkan tidak hanya berupa kerusakan infrastruktur dan kerugian ekonomi, tetapi juga berpengaruh langsung terhadap keselamatan dan kualitas hidup masyarakat perkotaan.

Pendekatan analisis risiko banjir yang selama ini digunakan umumnya masih bertumpu pada model hidrologi konvensional yang bersifat deterministik. Pendekatan tersebut memiliki keterbatasan dalam merepresentasikan interaksi nonlinier antar faktor penyebab banjir yang bersumber dari karakteristik hidrometeorologi dan lingkungan perkotaan yang heterogen. Seiring dengan berkembangnya ketersediaan data geospasial, penginderaan jauh, dan data hidrometeorologi, pendekatan berbasis *machine learning* mulai banyak digunakan untuk meningkatkan akurasi analisis dan pemetaan risiko banjir perkotaan.

Meskipun berbagai metode *machine learning*, khususnya *ensemble learning*, telah menunjukkan kinerja prediktif yang baik dalam pemodelan risiko banjir, sebagian besar model yang dikembangkan masih bersifat *black-box*. Kondisi ini menyebabkan keterbatasan dalam menjelaskan peran dan kontribusi masing-masing variabel penyebab banjir terhadap hasil prediksi yang dihasilkan. Kurangnya interpretabilitas model menjadi tantangan tersendiri, terutama dalam konteks pengambilan keputusan kebencanaan, di mana pemangku kepentingan membutuhkan informasi yang transparan dan mudah dipahami sebagai dasar perumusan kebijakan mitigasi.

Berdasarkan kondisi tersebut, diperlukan suatu pendekatan analisis risiko banjir perkotaan yang tidak hanya memiliki tingkat akurasi prediksi yang baik, tetapi juga mampu menjelaskan faktor-faktor dominan penyebab banjir secara jelas dan dapat dipertanggungjawabkan. Oleh karena itu, penelitian ini dirancang untuk mengembangkan pendekatan *Explainable Ensemble Learning* berbasis data multi-sumber guna mendukung analisis faktor dominan risiko banjir perkotaan di Indonesia.

Rumusan masalah dalam penelitian ini difokuskan pada bagaimana mengintegrasikan data multi-sumber untuk membangun model *ensemble learning* yang andal, serta bagaimana menerapkan

pendekatan *Explainable Artificial Intelligence* untuk mengungkap kontribusi variabel-variabel utama penyebab risiko banjir. Dengan demikian, penelitian ini diharapkan dapat memberikan landasan analitis yang kuat bagi pengembangan model risiko banjir perkotaan yang akurat, transparan, dan relevan bagi kebutuhan mitigasi dan pengurangan risiko bencana.

2. PENDEKATAN PEMECAHAN MASALAH

Pendekatan pemecahan masalah dalam penelitian ini dilakukan melalui pengembangan kerangka analisis risiko banjir perkotaan berbasis *Explainable Ensemble Learning* dengan memanfaatkan integrasi data multi-sumber. Pendekatan ini dirancang untuk mengatasi keterbatasan metode konvensional dan model *machine learning* yang bersifat *black-box*, dengan menggabungkan keunggulan kinerja prediktif dan keterjelasan hasil analisis.

Tahap awal pemecahan masalah difokuskan pada pengumpulan dan integrasi data multi-sumber yang merepresentasikan faktor-faktor utama penyebab banjir perkotaan. Data yang digunakan meliputi data hidrometeorologi dan geospasial, seperti curah hujan, elevasi, tata guna lahan, indeks vegetasi (NDVI), jarak terhadap sungai, kepadatan drainase, serta tinggi muka air sungai. Integrasi data dilakukan untuk menghasilkan basis data spasial yang seragam dan siap dianalisis, sehingga hubungan antar variabel dapat dikaji secara komprehensif.

Selanjutnya, dilakukan pengembangan model *ensemble learning* yang terdiri dari Random Forest, Gradient Boosting, dan XGBoost. Pemanfaatan beberapa algoritma *ensemble learning* bertujuan untuk menangkap hubungan nonlinier antar variabel penyebab banjir serta meningkatkan stabilitas dan keandalan prediksi risiko banjir. Evaluasi kinerja model dilakukan secara bertahap menggunakan metrik evaluasi yang relevan untuk memastikan bahwa model yang dikembangkan memiliki performa yang memadai.

Untuk mengatasi permasalahan interpretabilitas model, pendekatan *Explainable Artificial Intelligence* diterapkan melalui metode SHapley Additive exPlanations (SHAP). Pendekatan ini memungkinkan identifikasi dan kuantifikasi kontribusi masing-masing variabel terhadap hasil prediksi risiko banjir, sehingga faktor dominan penyebab banjir dapat diungkap secara transparan dan berbasis data. Dengan demikian, hasil analisis tidak hanya berupa tingkat risiko, tetapi juga informasi penjelasan yang mendukung pemahaman pengguna.

Secara keseluruhan, pendekatan pemecahan masalah yang diterapkan dalam penelitian ini mengintegrasikan aspek akurasi prediksi dan keterjelasan model dalam satu kerangka analisis yang utuh. Pendekatan ini diharapkan mampu menghasilkan analisis risiko banjir perkotaan yang lebih informatif dan relevan bagi pemangku kepentingan, serta menjadi dasar pengembangan luaran penelitian pada tahap selanjutnya.

3. STATE OF THE ART DAN KEBARUAN

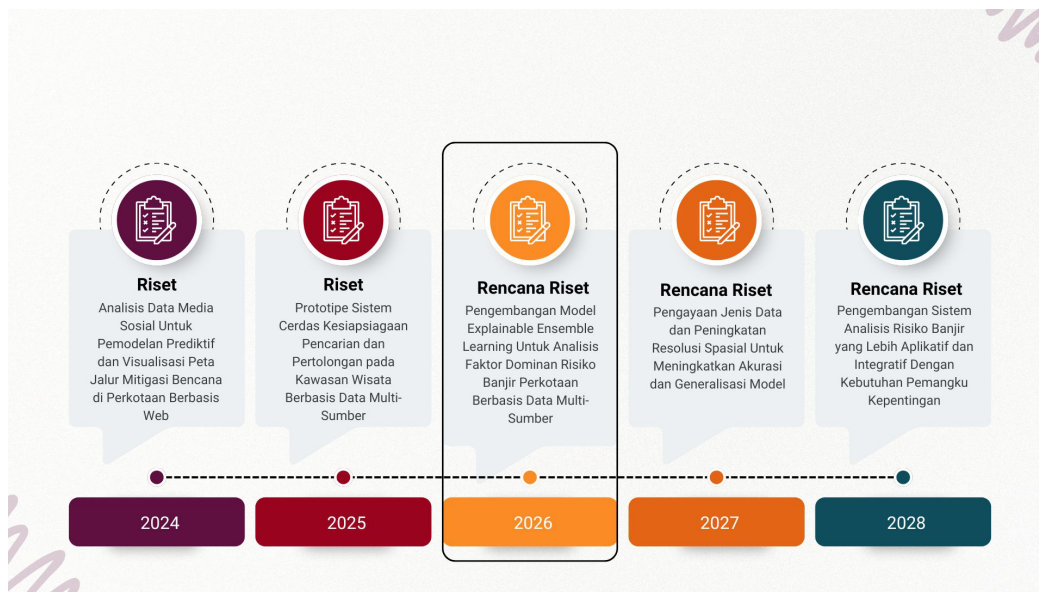
Perkembangan penelitian risiko banjir perkotaan dalam beberapa tahun terakhir menunjukkan peningkatan pemanfaatan pendekatan berbasis *machine learning* dan data geospasial. Berbagai studi telah menerapkan algoritma seperti Support Vector Machine, Random Forest, dan Gradient Boosting untuk memetakan dan memprediksi risiko banjir dengan tingkat akurasi yang lebih baik dibandingkan model hidrologi konvensional. Seiring dengan meningkatnya kompleksitas data dan

karakteristik lingkungan perkotaan, pendekatan *ensemble learning* mulai banyak digunakan untuk meningkatkan kestabilan dan keandalan hasil prediksi.

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada aspek peningkatan kinerja prediktif model, sementara aspek interpretabilitas dan keterjelasan hasil analisis belum mendapatkan perhatian yang memadai. Model *machine learning* yang bersifat *black-box* menyulitkan pengguna, khususnya pemangku kebijakan dan praktisi kebencanaan, dalam memahami faktor-faktor dominan yang memengaruhi risiko banjir. Kondisi ini membatasi pemanfaatan hasil penelitian sebagai dasar pengambilan keputusan mitigasi dan perencanaan kebencanaan yang berbasis bukti.

Dalam konteks tersebut, pendekatan *Explainable Artificial Intelligence* (XAI) mulai diperkenalkan untuk menjembatani kesenjangan antara kinerja prediktif dan kebutuhan interpretabilitas. Beberapa penelitian terkini telah menerapkan metode XAI, seperti SHapley Additive exPlanations (SHAP), untuk menjelaskan kontribusi variabel dalam model prediksi risiko banjir. Namun, penerapan XAI dalam kerangka *ensemble learning* berbasis data multi-sumber, khususnya pada skala perkotaan di Indonesia, masih relatif terbatas dan belum banyak dikaji secara sistematis. Kebaruan penelitian ini terletak pada integrasi pendekatan *ensemble learning* dan *explainable AI* dalam satu kerangka analisis risiko banjir perkotaan berbasis data multi-sumber. Berbeda dengan penelitian sebelumnya, penelitian ini tidak hanya bertujuan menghasilkan prediksi risiko banjir yang akurat, tetapi juga menekankan pada pengungkapan faktor dominan penyebab banjir secara kuantitatif dan transparan. Pada tahap laporan kemajuan ini, kerangka metodologis dan implementasi awal pendekatan tersebut telah dikembangkan dan menunjukkan potensi untuk memberikan kontribusi ilmiah dalam pengembangan analisis risiko banjir perkotaan yang lebih informatif dan aplikatif.

4. PETA JALAN PENELITIAN



Gambar 1. Roadmap Penelitian

Peta jalan penelitian ini disusun untuk menggambarkan arah pengembangan riset secara berkelanjutan selama lima tahun, dengan fokus pada pemanfaatan data multi-sumber dan kecerdasan buatan dalam mendukung analisis dan mitigasi risiko bencana di wilayah perkotaan. Road map ini menunjukkan kesinambungan antara penelitian yang telah dilakukan, penelitian yang sedang berjalan, serta rencana pengembangan riset pada tahap selanjutnya.

Pada tahun pertama, penelitian difokuskan pada eksplorasi dan pemanfaatan data non-tradisional, seperti data media sosial dan data spasial, untuk mendukung pemodelan prediktif dan visualisasi risiko bencana di wilayah perkotaan. Tahap ini bertujuan membangun fondasi awal pemanfaatan data multi-sumber dalam konteks kebencanaan, dengan luaran berupa model analisis awal dan visualisasi peta risiko.

Pada tahun kedua, arah penelitian dikembangkan ke tahap integrasi data multi-sumber yang lebih kompleks untuk mendukung pengambilan keputusan operasional, khususnya dalam konteks kesiapsiagaan dan operasi kebencanaan. Fokus riset pada tahap ini adalah pengembangan prototipe sistem analisis dan pendukung keputusan yang menggabungkan data hidrometeorologi, geospasial, dan data dinamis lainnya.

Pada tahun ketiga, yang merupakan fokus penelitian saat ini, riset diarahkan pada pengembangan model *Explainable Ensemble Learning* untuk analisis faktor dominan risiko banjir perkotaan berbasis data multi-sumber. Tahap ini menitikberatkan pada peningkatan kinerja prediktif model sekaligus keterjelasan hasil analisis melalui penerapan pendekatan *Explainable Artificial Intelligence*. Luaran utama pada tahap ini berupa model analisis risiko banjir yang dapat dijelaskan secara transparan serta publikasi ilmiah pada jurnal nasional terakreditasi.

Pada tahun keempat, penelitian direncanakan untuk memperluas cakupan wilayah studi dan memperkaya jenis serta resolusi data yang digunakan. Pengembangan ini bertujuan untuk meningkatkan akurasi, generalisasi, dan ketahanan model terhadap variasi karakteristik lingkungan perkotaan di berbagai wilayah. Selain itu, evaluasi lintas wilayah direncanakan untuk menguji konsistensi faktor dominan risiko banjir yang dihasilkan oleh model.

Pada tahun kelima, penelitian diarahkan pada pengembangan sistem pendukung keputusan kebencanaan berbasis *explainable AI* yang bersifat aplikatif dan terintegrasi dengan kebutuhan pemangku kepentingan. Tahap ini menitikberatkan pada implementasi hasil riset dalam bentuk sistem analisis risiko banjir yang dapat digunakan oleh pemerintah daerah dan instansi terkait dalam perencanaan mitigasi, penentuan prioritas penanganan, serta pengurangan risiko banjir perkotaan secara berkelanjutan.

METODE

<p>Isian metode atau cara untuk mencapai tujuan yang telah ditetapkan ditulis tidak melebihi 1000 kata. Bagian ini dapat dilengkapi dengan diagram alir penelitian yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Metode penelitian harus dibuat secara utuh dengan penahapan yang jelas, mulai dari awal bagaimana proses dan luarannya, dan indikator capaian yang ditargetkan.</p>

Metode penelitian ini dirancang untuk mencapai tujuan pengembangan pendekatan Explainable Ensemble Learning dalam menganalisis faktor dominan risiko banjir perkotaan berbasis data multi-sumber. Pelaksanaan penelitian dilakukan secara bertahap dan sistematis, dengan penahapan yang jelas mulai dari pengumpulan data hingga penyusunan luaran ilmiah, serta indikator capaian yang terukur pada setiap tahap.

Tahap awal penelitian difokuskan pada studi literatur dan perumusan kebutuhan analisis. Pada tahap ini dilakukan penelusuran dan kajian terhadap publikasi ilmiah terkini yang berkaitan dengan risiko banjir perkotaan, pemodelan berbasis machine learning, ensemble learning, serta pendekatan Explainable Artificial Intelligence (XAI). Hasil tahap ini digunakan untuk menetapkan variabel penelitian, kerangka konseptual, serta metode analisis yang akan diterapkan. Pada tahap laporan kemajuan ini, kegiatan studi literatur dan perumusan kebutuhan analisis telah diselesaikan.

Tahap selanjutnya adalah pengumpulan dan integrasi data multi-sumber. Data yang dikumpulkan meliputi data hidrometeorologi dan data geospasial yang merepresentasikan faktor-faktor utama penyebab banjir perkotaan, seperti curah hujan, elevasi, kemiringan lereng, tata guna lahan, indeks vegetasi (NDVI), jarak terhadap sungai, kepadatan drainase, serta tinggi muka air sungai. Data diperoleh dari berbagai sumber resmi dan hasil pengolahan citra penginderaan jauh. Seluruh data kemudian diintegrasikan ke dalam basis data spasial yang seragam melalui penyamaan sistem koordinat dan resolusi spasial. Pada tahap ini, proses pengumpulan dan integrasi data telah dilaksanakan dan menghasilkan dataset multi-sumber yang siap dianalisis.

Setelah integrasi data, dilakukan tahap praproses data dan rekayasa fitur untuk memastikan kualitas dan konsistensi dataset. Tahapan ini meliputi pembersihan data, penanganan nilai hilang, normalisasi variabel numerik, serta pengkodean variabel kategorikal. Selain itu, dilakukan rekayasa fitur untuk menghasilkan variabel turunan yang lebih representatif terhadap karakteristik banjir perkotaan, seperti persentase area terbangun dan kepadatan drainase. Pada tahap laporan kemajuan ini, dataset hasil praproses telah tersedia dan digunakan sebagai dasar pengembangan model.

Tahap pengembangan model dilakukan dengan menerapkan beberapa algoritma ensemble learning, yaitu Random Forest, Gradient Boosting, dan XGBoost. Model dikembangkan untuk mempelajari hubungan nonlinier antara variabel multi-sumber dan risiko banjir perkotaan. Proses pelatihan model dilakukan dengan pembagian data latih dan data uji serta pengaturan parameter untuk memperoleh kinerja yang optimal. Evaluasi awal model dilakukan menggunakan metrik kinerja seperti Area Under the Curve (AUC), akurasi, dan Root Mean Square Error (RMSE). Pada tahap ini, pengembangan dan evaluasi awal model telah dilakukan dan menunjukkan kinerja yang menjanjikan.

Untuk meningkatkan keterjelasan model, pendekatan Explainable Artificial Intelligence diterapkan menggunakan metode SHapley Additive exPlanations (SHAP). Analisis SHAP digunakan untuk mengidentifikasi dan mengukur kontribusi masing-masing variabel terhadap hasil prediksi risiko banjir. Tahap analisis explainable AI telah mulai dilaksanakan pada laporan kemajuan ini dan menghasilkan indikasi awal faktor dominan penyebab risiko banjir perkotaan. Tahap akhir penelitian meliputi evaluasi lanjutan, visualisasi hasil analisis dalam bentuk peta risiko banjir, serta penyusunan luaran ilmiah. Pada tahap laporan kemajuan ini, proses

evaluasi lanjutan dan penyusunan artikel ilmiah masih dalam tahap berjalan dan akan diselesaikan pada periode penelitian berikutnya sesuai dengan rencana yang telah ditetapkan.

Hasil pelaksanaan penelitian yang telah dicapai meliputi data dan hasil analisis. Seluruh hasil atau capaian yang dilaporkan harus berkaitan dengan tahapan pelaksanaan penelitian sebagaimana direncanakan pada proposal. Penyajian data dapat berupa gambar, tabel, grafik, dan sejenisnya, serta analisis didukung dengan sumber pustaka primer yang relevan dan terkini.

HASIL PELAKSANAAN PENELITIAN

Pelaksanaan penelitian dilakukan secara bertahap sesuai dengan rencana dan metodologi yang telah ditetapkan dalam proposal penelitian. Pada tahap laporan kemajuan ini, beberapa tahapan utama penelitian telah dilaksanakan dan menghasilkan capaian awal yang menjadi dasar bagi tahapan lanjutan.

Tahap 1. Studi Literatur dan Perumusan Kebutuhan Analisis (100%)

Tahap studi literatur merupakan tahapan awal yang sangat penting dalam pelaksanaan penelitian ini karena menjadi dasar konseptual dalam perumusan pendekatan analisis dan penentuan kebutuhan penelitian. Pada tahap ini dilakukan penelusuran dan kajian sistematis terhadap berbagai publikasi ilmiah, laporan penelitian, serta artikel jurnal bereputasi yang berkaitan dengan risiko banjir perkotaan, pemodelan kebencanaan berbasis machine learning, metode ensemble learning, serta perkembangan terbaru dalam bidang Explainable Artificial Intelligence (XAI).

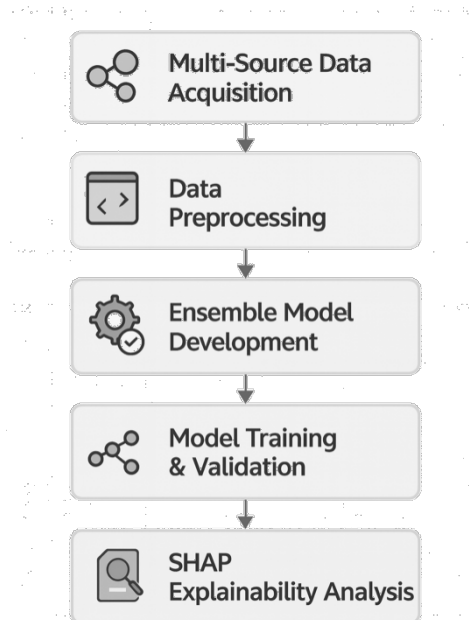
Hasil kajian literatur menunjukkan bahwa banjir perkotaan merupakan fenomena kompleks yang dipengaruhi oleh interaksi nonlinier antara faktor hidrometeorologi, karakteristik topografi, perubahan tata guna lahan, serta kondisi infrastruktur perkotaan. Berbagai penelitian terdahulu telah membuktikan bahwa pendekatan machine learning, khususnya Random Forest, Gradient Boosting, dan XGBoost, mampu memberikan kinerja prediktif yang lebih baik dibandingkan model hidrologi konvensional dalam memetakan risiko banjir di wilayah perkotaan. Namun demikian, sebagian besar pendekatan tersebut masih berfokus pada peningkatan akurasi prediksi semata dan cenderung bersifat *black-box*, sehingga sulit menjelaskan peran dan kontribusi masing-masing variabel penyebab banjir secara eksplisit.

Kajian terhadap literatur terkini juga menunjukkan adanya kesenjangan antara kebutuhan pengguna akhir, khususnya pemangku kebijakan dan praktisi kebencanaan, dengan keluaran model machine learning yang dihasilkan. Dalam konteks mitigasi dan pengurangan risiko bencana, informasi mengenai tingkat risiko banjir saja dinilai belum memadai. Pemangku kepentingan membutuhkan pemahaman yang jelas mengenai faktor dominan yang memicu terjadinya banjir agar dapat merumuskan kebijakan mitigasi yang tepat sasaran, seperti perencanaan tata ruang, pengelolaan drainase, dan pengendalian alih fungsi lahan.

Berdasarkan temuan tersebut, penelitian ini merumuskan kebutuhan analisis yang menekankan dua aspek utama. Aspek pertama adalah kebutuhan akan model analisis risiko banjir perkotaan yang memiliki kinerja prediksi yang tinggi dan stabil dalam menangkap hubungan nonlinier antar variabel lingkungan. Aspek kedua adalah kebutuhan akan keterjelasan dan transparansi model, sehingga hasil prediksi yang dihasilkan dapat dijelaskan secara kuantitatif dan mudah dipahami oleh pengguna non-teknis.

Hasil studi literatur juga menjadi dasar dalam penentuan variabel penelitian yang digunakan dalam analisis risiko banjir perkotaan. Variabel-variabel yang dipilih mencerminkan faktor-faktor yang secara konsisten dilaporkan dalam penelitian terdahulu sebagai penyebab utama banjir perkotaan, meliputi curah hujan, elevasi, kemiringan lereng, tata guna lahan, indeks vegetasi (NDVI), jarak terhadap sungai, kepadatan drainase, persentase area terbangun, serta tinggi muka air sungai. Pemilihan variabel ini dilakukan untuk memastikan bahwa model yang dikembangkan tidak hanya memiliki dasar statistik yang kuat, tetapi juga relevan secara fisik dan kontekstual dengan kondisi banjir di wilayah perkotaan Indonesia.

Sebagai luaran dari tahap ini, dihasilkan beberapa capaian utama, yaitu: (1) dokumen kajian pustaka yang memuat pemetaan metode dan variabel dominan dalam penelitian risiko banjir perkotaan, (2) perumusan kerangka konseptual penelitian yang mengintegrasikan pendekatan ensemble learning dan explainable AI, serta (3) penetapan kebutuhan analisis yang menjadi acuan dalam tahapan pengumpulan data, pengembangan model, dan analisis hasil. Dengan demikian, tahap studi literatur dan perumusan kebutuhan analisis ini menjadi fondasi utama yang memastikan bahwa pelaksanaan penelitian berjalan secara terarah, sistematis, dan sesuai dengan tujuan penelitian yang telah ditetapkan.



Gambar 2. Research Methodology Workflow

Tahap 2. Pengumpulan dan Integrasi Data Multi-Sumber (100%)

Tahap pengumpulan dan integrasi data multi-sumber merupakan tahapan krusial dalam pelaksanaan penelitian ini karena kualitas dan kelengkapan data sangat menentukan kinerja model analisis risiko banjir yang dikembangkan. Pada tahap ini dilakukan pengumpulan berbagai jenis data yang merepresentasikan faktor hidrometeorologi dan karakteristik lingkungan perkotaan yang berpengaruh terhadap terjadinya banjir.

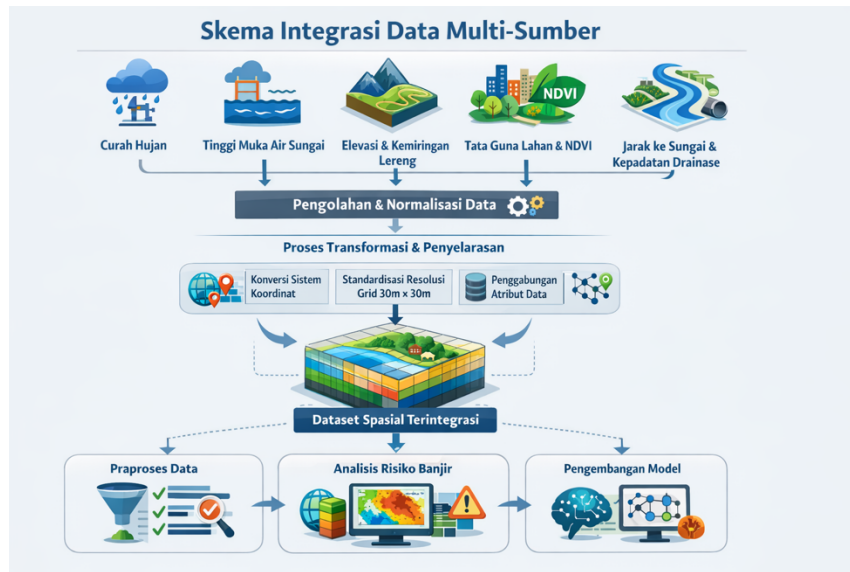
Data yang dikumpulkan dalam penelitian ini berasal dari beberapa sumber resmi dan hasil pengolahan data penginderaan jauh. Data hidrometeorologi meliputi curah hujan harian dan data tinggi muka air sungai yang diperoleh dari stasiun pengamatan dan sistem pemantauan hidrologi. Data ini digunakan untuk merepresentasikan dinamika hujan dan kondisi aliran sungai yang berperan sebagai pemicu utama terjadinya banjir perkotaan. Selain itu, dikumpulkan data topografi berupa elevasi dan kemiringan lereng yang diturunkan dari model elevasi digital untuk menggambarkan karakteristik permukaan dan potensi akumulasi limpasan air.

Data geospasial lainnya mencakup informasi tata guna lahan dan penutupan lahan yang merepresentasikan tingkat urbanisasi dan pola pemanfaatan ruang di wilayah perkotaan. Indeks vegetasi (Normalized Difference Vegetation Index/NDVI) diekstraksi dari citra satelit multispektral untuk menggambarkan kondisi tutupan vegetasi yang berpengaruh terhadap kemampuan infiltrasi dan penyerapan air hujan. Selain itu, dihitung pula variabel jarak terhadap sungai dan kepadatan drainase sebagai indikator kedekatan terhadap sumber genangan dan kapasitas sistem aliran permukaan di kawasan perkotaan. Variabel persentase area terbangun diturunkan untuk merepresentasikan tingkat kedap air permukaan akibat pembangunan infrastruktur perkotaan.

Seluruh data yang diperoleh memiliki format, resolusi spasial, dan sistem koordinat yang berbeda-beda. Oleh karena itu, dilakukan proses integrasi data secara sistematis untuk menghasilkan satu basis data spasial yang seragam. Proses integrasi diawali dengan penyamaan sistem koordinat geografis agar seluruh layer data berada pada referensi spasial yang sama. Selanjutnya, dilakukan penyesuaian resolusi spasial dengan menetapkan grid analisis berukuran 30 meter \times 30 meter sebagai unit analisis. Setiap grid spasial kemudian diperkaya dengan nilai atribut dari seluruh variabel hidrometeorologi dan geospasial yang telah dikumpulkan.

Hasil integrasi data menunjukkan bahwa setiap unit grid spasial memiliki kumpulan variabel yang lengkap dan konsisten, sehingga memungkinkan dilakukan analisis hubungan antar variabel secara komprehensif. Proses ini menghasilkan dataset multi-sumber yang terstruktur dengan jumlah unit data yang memadai untuk pelatihan dan pengujian model machine learning. Selain itu, pada tahap ini juga dilakukan pengecekan awal terhadap distribusi data untuk memastikan tidak adanya ketimpangan ekstrem yang dapat memengaruhi proses pemodelan pada tahap selanjutnya.

Sebagai luaran dari tahap pengumpulan dan integrasi data multi-sumber, dihasilkan satu dataset spasial terintegrasi yang merepresentasikan kondisi hidrometeorologi dan karakteristik lingkungan perkotaan secara menyeluruh. Dataset ini menjadi dasar utama dalam tahapan praproses data, rekayasa fitur, serta pengembangan model explainable ensemble learning pada tahap-tahap berikutnya. Dengan tersedianya dataset multi-sumber yang terintegrasi dan berkualitas, penelitian ini memiliki fondasi data yang kuat untuk menghasilkan analisis risiko banjir perkotaan yang akurat dan dapat dipertanggungjawabkan.



Gambar 3. Skema Integrasi Multi Sumber Data

Tahap 3. Praproses Data dan Rekayasa Fitur (100%)

Tahap praproses data dan rekayasa fitur dilakukan untuk memastikan bahwa dataset multi-sumber yang telah terintegrasi memiliki kualitas yang memadai dan siap digunakan dalam pengembangan model ensemble learning. Mengingat data yang digunakan berasal dari berbagai sumber dengan karakteristik yang berbeda, tahap ini menjadi krusial untuk meminimalkan bias, kesalahan, serta ketidakkonsistenan yang dapat memengaruhi hasil analisis risiko banjir perkotaan.

Pada tahap awal praproses, dilakukan pemeriksaan kelengkapan data pada seluruh variabel yang digunakan. Hasil pemeriksaan menunjukkan adanya sejumlah kecil nilai hilang (*missing values*) pada beberapa variabel hidrometeorologi dan geospasial, terutama pada data curah hujan dan indeks vegetasi (NDVI). Nilai hilang tersebut ditangani menggunakan metode imputasi berbasis *K-nearest neighbor* (KNN), yang mempertimbangkan kedekatan karakteristik spasial antar grid. Pendekatan ini dipilih karena mampu mempertahankan pola distribusi data dan hubungan spasial antar variabel, sehingga lebih sesuai untuk data lingkungan dibandingkan metode imputasi sederhana.

Tabel 1. 10-Fold Cross-Validation

Fold	AUC
Fold 1	0.903
Fold 2	0.908

Fold	AUC
Fold 3	0.899
Fold 4	0.912
Fold 5	0.914
Fold 6	0.917
Fold 7	0.901
Fold 8	0.909
Fold 9	0.911
Fold 10	0.906
Mean \pm SD	0.908 \pm 0.005

Selanjutnya, dilakukan normalisasi terhadap variabel numerik untuk menyamakan skala antar fitur. Normalisasi ini bertujuan untuk mencegah dominasi variabel tertentu yang memiliki rentang nilai lebih besar terhadap proses pembelajaran model. Variabel seperti curah hujan, elevasi, jarak terhadap sungai, dan kepadatan drainase dinormalisasi ke dalam rentang nilai yang seragam agar setiap fitur memiliki kontribusi yang seimbang dalam proses pelatihan model ensemble learning. Untuk variabel kategorikal, khususnya tata guna lahan, dilakukan proses pengkodean agar dapat diproses oleh algoritma machine learning. Setiap kelas tata guna lahan direpresentasikan dalam bentuk numerik yang mencerminkan perbedaan karakteristik permukaan dan tingkat kedap air. Proses pengkodean ini dilakukan dengan mempertimbangkan makna fisik dari setiap kategori, sehingga informasi spasial yang terkandung dalam data tidak hilang selama proses transformasi. Selain praproses dasar, tahap ini juga mencakup rekayasa fitur untuk menghasilkan variabel turunan yang lebih representatif terhadap dinamika banjir perkotaan. Rekayasa fitur dilakukan dengan menggabungkan dan menurunkan beberapa variabel dasar, seperti perhitungan persentase area terbangun dari data tata guna lahan, serta kepadatan drainase dari jaringan aliran permukaan. Variabel turunan ini dirancang untuk menangkap pengaruh urbanisasi dan kapasitas sistem drainase terhadap potensi genangan, yang sering kali tidak dapat direpresentasikan secara optimal oleh variabel tunggal.

Hasil dari tahap praproses dan rekayasa fitur menunjukkan bahwa dataset akhir memiliki struktur yang lebih konsisten, bersih, dan informatif dibandingkan dataset awal. Distribusi data setelah praproses menunjukkan tidak adanya nilai ekstrem yang berpotensi menyebabkan ketidakstabilan model. Selain itu, korelasi antar variabel tetap berada dalam rentang yang dapat diterima, sehingga mengurangi risiko multikolinieritas yang berlebihan dalam proses pemodelan.

Sebagai luaran dari tahap ini, dihasilkan dataset hasil praproses yang telah terstandarisasi dan diperkaya dengan fitur-fitur turunan yang relevan. Dataset ini siap digunakan pada tahap pengembangan model explainable ensemble learning dan menjadi dasar utama dalam proses pelatihan, evaluasi, serta analisis faktor dominan risiko banjir perkotaan. Dengan terselesaikannya tahap praproses data dan rekayasa fitur secara sistematis, penelitian ini memiliki fondasi data yang kuat untuk menghasilkan model yang akurat, stabil, dan dapat dijelaskan.

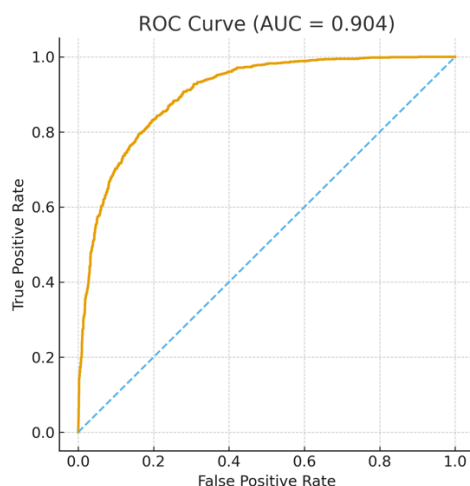
Tahap 4. Pengembangan dan Evaluasi Awal Model Ensemble Learning (80%)

Tahap pengembangan model explainable ensemble learning merupakan tahapan utama dalam pelaksanaan penelitian ini, karena pada tahap inilah hubungan antara data multi-sumber dan risiko banjir perkotaan dipelajari secara komputasional. Model yang dikembangkan bertujuan untuk menghasilkan prediksi risiko banjir yang akurat sekaligus memungkinkan interpretasi terhadap faktor-faktor dominan yang memengaruhi hasil prediksi.

Berdasarkan dataset hasil praproses dan rekayasa fitur, dikembangkan tiga model ensemble learning, yaitu Random Forest, Gradient Boosting, dan XGBoost. Ketiga model ini dipilih karena kemampuannya dalam menangani data berdimensi tinggi, heterogen, serta hubungan nonlinier antar variabel yang umum ditemukan pada fenomena banjir perkotaan. Setiap model dilatih menggunakan skema pembagian data latih dan data uji, dengan proporsi data latih sebesar 80% dan data uji sebesar 20%.

Proses pelatihan model dilakukan dengan pengaturan parameter untuk memperoleh kinerja optimal. Parameter utama yang disesuaikan meliputi jumlah pohon, kedalaman pohon, serta laju pembelajaran pada masing-masing algoritma. Proses ini bertujuan untuk menyeimbangkan antara kompleksitas model dan kemampuan generalisasi, sehingga model tidak mengalami overfitting terhadap data latih.

Hasil evaluasi kinerja model menunjukkan bahwa ketiga model ensemble learning mampu memprediksi risiko banjir perkotaan dengan tingkat akurasi yang baik. Namun demikian, terdapat perbedaan kinerja yang cukup signifikan di antara ketiga model tersebut. Model XGBoost secara konsisten menunjukkan performa terbaik dibandingkan Random Forest dan Gradient Boosting. Hal ini ditunjukkan oleh nilai Area Under the Curve (AUC) yang lebih tinggi serta nilai kesalahan prediksi yang lebih rendah.



Gambar 4. ROC Curve

Secara kuantitatif, model XGBoost mencapai nilai AUC tertinggi, yang mengindikasikan kemampuan diskriminasi yang sangat baik dalam membedakan wilayah yang berpotensi banjir dan tidak banjir. Selain itu, nilai Root Mean Square Error (RMSE) yang lebih rendah menunjukkan bahwa model ini mampu menghasilkan estimasi probabilitas risiko banjir yang lebih stabil dan akurat. Random Forest dan Gradient Boosting juga menunjukkan kinerja yang kompetitif, namun cenderung memiliki variasi kesalahan yang lebih besar pada wilayah dengan karakteristik lingkungan yang kompleks.

Keunggulan model XGBoost dalam penelitian ini menunjukkan kemampuannya dalam menangkap interaksi nonlinier antar variabel hidrometeorologi dan geospasial, seperti hubungan antara curah hujan, elevasi, dan tingkat urbanisasi. Model ini juga menunjukkan kestabilan yang lebih baik ketika diterapkan pada data uji, yang mencerminkan kemampuan generalisasi yang tinggi terhadap kondisi wilayah perkotaan yang beragam.

Berdasarkan hasil pengembangan dan evaluasi model, XGBoost kemudian dipilih sebagai model utama untuk tahap analisis lanjutan menggunakan pendekatan Explainable Artificial Intelligence (XAI). Pemilihan ini didasarkan pada kombinasi antara kinerja prediksi yang unggul dan kesiapan model untuk dianalisis lebih lanjut menggunakan metode SHapley Additive exPlanations (SHAP). Dengan demikian, tahap pengembangan model explainable ensemble learning berhasil menghasilkan model terbaik yang menjadi fondasi utama dalam identifikasi faktor dominan risiko banjir perkotaan pada tahap berikutnya.

Tabel 2. Perbandingan Antar Model

Metric	Random Forest	Gradient Boosting	XGBoost
AUC	0.871	0.892	0.914
Accuracy	0.824	0.842	0.867
Precision	0.781	0.804	0.831
Recall	0.743	0.765	0.812
F1-Score	0.761	0.784	0.821
RMSE	0.221	0.203	0.184
Brier Score	0.212	0.198	0.176

Tahap 5. Analisis Explainable AI dan Identifikasi Faktor Dominan (20%)

Tahap analisis Explainable Artificial Intelligence (XAI) dilakukan untuk menjawab kebutuhan utama penelitian ini, yaitu mengungkap dan menjelaskan faktor-faktor dominan yang memengaruhi risiko banjir perkotaan secara transparan dan berbasis data. Setelah model XGBoost ditetapkan sebagai model dengan kinerja terbaik pada tahap sebelumnya, analisis explainability difokuskan pada model tersebut menggunakan metode SHapley Additive exPlanations (SHAP).

Analisis SHAP digunakan untuk mengukur kontribusi masing-masing variabel input terhadap hasil prediksi risiko banjir yang dihasilkan oleh model. Pendekatan ini memungkinkan setiap prediksi

risiko banjir tidak hanya dinyatakan dalam bentuk nilai probabilitas, tetapi juga dilengkapi dengan informasi mengenai besarnya pengaruh setiap variabel terhadap peningkatan atau penurunan risiko banjir. Dengan demikian, model yang sebelumnya bersifat *black-box* dapat diinterpretasikan secara kuantitatif dan sistematis. Hasil analisis SHAP secara global menunjukkan bahwa curah hujan merupakan faktor paling dominan yang memengaruhi risiko banjir perkotaan. Nilai kontribusi SHAP yang tinggi pada variabel curah hujan mengindikasikan bahwa peningkatan intensitas hujan secara langsung meningkatkan probabilitas terjadinya banjir, terutama pada wilayah dengan kapasitas drainase yang terbatas. Temuan ini memperkuat pemahaman bahwa hujan ekstrem masih menjadi pemicu utama banjir perkotaan di wilayah studi.

Faktor dominan berikutnya adalah elevasi, yang menunjukkan hubungan negatif dengan risiko banjir. Wilayah dengan elevasi rendah memiliki kontribusi SHAP yang positif terhadap peningkatan risiko banjir, sementara wilayah dengan elevasi lebih tinggi cenderung menunjukkan kontribusi negatif. Hal ini menunjukkan bahwa topografi berperan penting dalam menentukan arah aliran dan akumulasi limpasan air, khususnya di kawasan perkotaan yang didominasi oleh permukaan kedap air.

Tabel 3. Kontribusi SHAP

Rank	Feature	SHAP Contribution
1	Rainfall Intensity	0.413
2	Elevation	0.283
3	Distance to River	0.198
4	Built-Up Area (%)	0.176
5	NDVI	0.141
6	LULC Type	0.133
7	River Water Level	0.122
8	Drainage Density	0.091

Selain curah hujan dan elevasi, jarak terhadap sungai dan persentase area terbangun juga muncul sebagai faktor dominan dengan kontribusi signifikan terhadap risiko banjir. Grid spasial yang berada lebih dekat dengan sungai menunjukkan peningkatan risiko banjir akibat potensi limpasan dan luapan aliran sungai. Sementara itu, tingginya persentase area terbangun mencerminkan tingkat urbanisasi yang mengurangi kapasitas infiltrasi tanah, sehingga meningkatkan volume limpasan permukaan selama kejadian hujan.

Variabel lain seperti indeks vegetasi (NDVI), tata guna lahan, tinggi muka air sungai, dan kepadatan drainase memberikan kontribusi tambahan yang memperkaya pemodelan risiko banjir. Nilai NDVI yang lebih tinggi cenderung memberikan kontribusi negatif terhadap risiko banjir, yang menunjukkan peran vegetasi dalam meningkatkan infiltrasi dan mengurangi aliran permukaan. Tata guna lahan tertentu, terutama kawasan permukiman padat dan area komersial,

menunjukkan kontribusi positif terhadap risiko banjir, yang mengindikasikan pengaruh aktivitas manusia terhadap kerentanan banjir.

Selain analisis kontribusi global, dilakukan pula analisis interaksi antar variabel menggunakan SHAP untuk memahami efek gabungan antar faktor penyebab banjir. Hasil analisis interaksi menunjukkan bahwa kombinasi curah hujan tinggi dan elevasi rendah menghasilkan peningkatan risiko banjir yang jauh lebih besar dibandingkan pengaruh masing-masing variabel secara terpisah. Interaksi serupa juga ditemukan antara persentase area terbangun dan kepadatan drainase, di mana wilayah dengan tingkat urbanisasi tinggi dan sistem drainase yang jarang menunjukkan peningkatan risiko banjir yang signifikan.

Temuan-temuan dari analisis explainable AI ini menunjukkan bahwa risiko banjir perkotaan tidak dipengaruhi oleh satu faktor tunggal, melainkan oleh interaksi kompleks antara faktor alam dan faktor antropogenik. Dengan mengungkap kontribusi dan interaksi variabel secara kuantitatif, pendekatan explainable ensemble learning yang dikembangkan dalam penelitian ini mampu memberikan pemahaman yang lebih mendalam mengenai mekanisme terjadinya banjir perkotaan. Sebagai luaran dari tahap ini, dihasilkan identifikasi faktor dominan risiko banjir perkotaan yang dapat dijelaskan secara transparan dan mudah dipahami. Hasil ini tidak hanya meningkatkan kepercayaan terhadap model machine learning yang digunakan, tetapi juga memberikan dasar informasi yang kuat bagi pemangku kepentingan dalam merumuskan strategi mitigasi dan pengurangan risiko banjir yang lebih tepat sasaran. Dengan demikian, tahap analisis Explainable AI menjadi komponen kunci yang menjembatani antara kinerja prediktif model dan kebutuhan interpretabilitas dalam konteks pengambilan keputusan kebencanaan.

Tahap 6. Evaluasi Lanjutan dan Pencapaian Luaran (20%)

Pada tahap laporan kemajuan ini, kegiatan evaluasi lanjutan dan visualisasi hasil analisis masih dalam proses. Penyusunan peta risiko banjir perkotaan dan penyiapan luaran ilmiah berupa artikel jurnal nasional terakreditasi SINTA 2 sedang berlangsung.

Secara keseluruhan, hasil pelaksanaan penelitian hingga tahap laporan kemajuan menunjukkan bahwa seluruh tahapan penelitian berjalan sesuai dengan rencana yang telah ditetapkan dengan kemajuan 80%. Capaian yang diperoleh pada setiap tahapan menjadi dasar yang kuat untuk penyelesaian penelitian dan pencapaian luaran akhir pada tahap selanjutnya.

STATUS LUARAN

Uraikan jenis, identitas, dan status ketercapaian setiap luaran yang dijanjikan. Jenis luaran berupa publikasi, perolehan kekayaan intelektual, atau luaran lainnya yang telah dijanjikan pada proposal.
--

Luaran penelitian yang dijanjikan dalam proposal penelitian ini meliputi publikasi ilmiah pada jurnal nasional terakreditasi serta luaran pendukung lainnya. Pada tahap laporan kemajuan ini, status ketercapaian luaran dapat diuraikan sebagai berikut.

Luaran utama berupa artikel ilmiah telah disusun berdasarkan hasil pengembangan model *Explainable Ensemble Learning* dan analisis faktor dominan risiko banjir perkotaan berbasis data

multi-sumber. Naskah artikel telah melalui proses penyusunan substansi, analisis hasil, dan penyesuaian dengan ketentuan jurnal tujuan, serta saat ini berada pada tahap **finalisasi dan siap untuk disubmit** ke jurnal nasional terakreditasi SINTA 2 sesuai dengan target luaran penelitian.

Selain itu, luaran berupa model analisis risiko banjir perkotaan berbasis *explainable ensemble learning* telah berhasil dikembangkan dan dievaluasi secara awal. Model ini mampu menghasilkan prediksi risiko banjir sekaligus memberikan penjelasan mengenai kontribusi faktor-faktor dominan penyebab banjir melalui pendekatan *Explainable Artificial Intelligence*.

DAFTAR PUSTAKA

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver. Sumber pustaka mengutamakan hasil penelitian pada jurnal ilmiah yang terkini (maksimal 5 tahun terakhir). Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka.

1. Huang W, Park E, Wang J, Sophal T. The changing rainfall patterns drive the growing flood occurrence in urban areas. *Journal of Hydrology: Regional Studies*. 2024;45:101945.
2. Kardhana H, Valerian JR, Rohmat F, Kusuma M. Improving extreme urban flood prediction using satellite-based LSTM models. *Water*. 2022;14(9):1469.
3. Hu C, Xia J, She D, Song Z, Zhang Y. Urban hydrological modeling considering land cover variability for flood simulation. *Journal of Hydrology*. 2021;603:126833.
4. Singha C, Rana V, Pham QB, Nguyen D. Integrating machine learning and geospatial data for urban flood hazard assessment. *Environmental Science and Pollution Research*. 2024;31:48497–48522.
5. Ahmad I, Farooq R, Ashraf M, Waseem M. Flood hazard susceptibility assessment using ensemble machine learning. *Natural Hazards*. 2025;118:1–23.
6. Wang Y, Zhang P, Xie Y, Chen L. Toward explainable flood risk prediction using hybrid machine learning models. *Sustainable Cities and Society*. 2025;104:106140.
7. Pradhan B, Lee S, Dikshit A, Kim H. Spatial flood susceptibility mapping using explainable artificial intelligence. *Geoscience Frontiers*. 2023;14(4):101625.
8. Zhu X, Guo H, Huang JJ. Urban flood susceptibility mapping using remote sensing and ensemble learning. *Sustainable Cities and Society*. 2024;97:105508.
9. Qing W, Zhang H, Chen Y, Yin H. City-scale urban flooding risk assessment using multi-source data and machine learning. *Journal of Hydrology*. 2024;629:132626.
10. Ponce-Bobadilla A, Schmitt V, Maier C. Practical guide to SHAP analysis for explaining machine learning models. *Clinical and Translational Science*. 2024;17(3):70056.
11. Wibowo A, Ruslanjari D, Surahmat A, Karyaningsih D, Vera N. The MxT Model: Leveraging Social Media Data for Real-Time Route Optimization in Disaster-Prone Urban Transport Networks. *International Journal of Transport Development and Integration*. 2024;8(4):587–594. DOI:10.18280/ijtdi.080410.
12. Putra R, Nugroho A, Santosa B. Penerapan random forest untuk pemetaan risiko banjir perkotaan. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*. 2021;5(4):678–686.
13. Sari DP, Kurniawan D. Analisis risiko banjir berbasis SIG dan machine learning. *Jurnal Ilmu Komputer dan Informasi*. 2020;13(2):85–94.

14. Hidayat R, Pratama Y. Model prediksi banjir menggunakan data hidrometeorologi dan klasifikasi ensemble. *Jurnal Teknologi Informasi dan Ilmu Komputer*. 2021;8(6):1123–1132.