

BAB I

PENDAHULUAN

1.1 Latar Belakang

Wilayah Jawa Barat merupakan salah satu kawasan dengan tingkat kerawanan gempa bumi yang sangat tinggi akibat kondisi geografis dan tatanan tektoniknya. Posisi wilayah ini berada tepat pada zona pertemuan lempeng tektonik aktif, di mana Lempeng Indo-Australia yang bergerak ke utara bertumbukan dan menunjam ke bawah Lempeng Eurasia yang relatif diam, membentuk zona subduksi di bagian selatan pulau Jawa. Selain itu, daratan Jawa Barat juga dilewati oleh jalur cincin api dan berbagai sesar aktif yang membentang luas, seperti Sesar Cimandiri, Sesar Lembang, Sesar Baribis, dan Sesar Pelabuhan Ratu. Tingginya aktivitas seismik dari elemen-elemen tektonik tersebut menuntut adanya upaya mitigasi yang terencana dan terukur. Dalam konteks inilah, probabilitas gempa hadir sebagai ukuran kemungkinan terjadinya gempa bumi di suatu wilayah dalam kurun waktu tertentu, yang dinyatakan dalam bentuk nilai statistik sebagai dasar penilaian risiko bencana seismik. Konsep probabilitas gempa ini memunculkan kebutuhan untuk memahami besaran peluang kegempaan, di mana pemodelan yang akurat sangat penting untuk mitigasi risiko dan perencanaan tata ruang. Menurut (Distribusi et al. 2022) yang menganalisis distribusi spasial dan temporal parameter seismotektonik di Jawa Barat, pola probabilitas terjadinya gempa bumi dapat dipetakan menggunakan parameter seismologi tertentu seperti *a-value* dan *b-value*. Studi lain oleh (Hamid 2024) menekankan pentingnya pemetaan risiko bencana alam dalam upaya manajemen lingkungan berkelanjutan, di mana probabilitas gempa menjadi salah satu indikator vital untuk pengelompokan tingkat kerawanan wilayah. Berbagai penelitian juga menunjukkan bahwa probabilitas gempa sangat dipengaruhi oleh berbagai faktor seperti magnitudo, kedalaman, lokasi episenter,

serta aktivitas tektonik secara historis, sehingga penilaiannya memerlukan data yang komprehensif dan valid sebagai dasar prediksi. Dengan demikian, probabilitas gempa tidak sekadar angka statistik, tetapi juga merepresentasikan tingkat keterpaparan risiko masyarakat terhadap bahaya seismik yang berdampak pada proses perumusan kebijakan mitigasi bencana di kawasan rawan.

Kemajuan ilmu pengetahuan dan teknologi telah mendorong pemanfaatan model statistik dan pembelajaran mesin (*machine learning*) dalam memprediksi probabilitas gempa, yang menjadikan estimasi probabilitas tersebut semakin presisi dan adaptif terhadap dinamika data seismik. (Shukla et al. 2025) melalui pengembangan model *ensemble* Random Forest menunjukkan peningkatan kinerja prediksi probabilitas kejadian seismik, di mana model ini memanfaatkan fitur statistik dari data historis untuk memberikan keluaran probabilistik yang relevan bagi kebutuhan mitigasi risiko. Selanjutnya, (Hu et al. 2025) memperkenalkan pendekatan *neural network* multimodal untuk *forecasting* probabilitas gempa dalam jangka menengah dengan hasil yang cukup signifikan dalam meningkatkan keakuratan prediksi spasial dan temporal. (Syaifuddin and Prabowo 2024) juga menemukan bahwa integrasi metode *ensemble learning* mampu mengoptimalkan hasil prediksi, terutama ketika probabilitas gempa dinilai berdasarkan variabel magnitudo dan lokasi kejadian. (James, Gareth and Witten, Daniela and Hastie, Trevor and Tibshirani 2021) menegaskan relevansi pendekatan statistik dan *machine learning* untuk menginterpretasikan probabilitas dalam konteks ketidakpastian kegempaan, memperkuat dasar pemodelan prediktif yang berorientasi pada evaluasi risiko bencana. Keterpaduan metode kuantitatif ini mampu menyajikan output probabilitas gempa yang adaptif, sehingga pada akhirnya, pemanfaatan probabilitas gempa secara komprehensif menjadi kunci utama dalam mendukung keberhasilan kebijakan mitigasi kegempaan yang berbasis data, serta memperkecil dampak sosial ekonomi di masyarakat.

Dengan semakin vitalnya pemanfaatan probabilitas gempa untuk mendukung keberhasilan kebijakan mitigasi kegempaan yang berbasis data, penelitian ini secara spesifik bertujuan untuk melakukan analisis komparatif antara dua model pembelajaran mesin, yakni Random Forest dan Support Vector Machine, dalam memprediksi probabilitas terjadinya gempa bumi di wilayah Jawa Barat dalam rentang waktu tertentu. Signifikansi penelitian ini terletak pada kemampuannya untuk memberikan kontribusi empiris terhadap optimalisasi pemodelan probabilitas gempa, melalui pendayagunaan metode machine learning yang telah terbukti adaptif dan akurat dalam menangkap pola spasial-temporal dari data seismik historis. Kehadiran studi ini sangat penting mengingat karakteristik seismotektonik Jawa Barat yang kompleks, sehingga diperlukan pendekatan prediktif yang mampu mengakomodasi aspek multivariat dan non-linieritas data kegempaan yang secara konvensional kerap menghambat ketepatan estimasi probabilitas. Selain itu, dengan membandingkan dua pendekatan machine learning yang populer dan relatif banyak diadopsi dalam literatur terkini, penelitian ini juga memfasilitasi pemilihan model yang paling unggul untuk diaplikasikan dalam sistem peringatan dini maupun perencanaan mitigasi berbasis risiko. Urgensi penelitian ini semakin tinggi karena tantangan minimnya solusi prediksi probabilitas gempa yang akurat dan aplikatif masih menjadi isu di lapangan, khususnya di kawasan rawan seperti Jawa Barat, yang mengakibatkan rendahnya efektivitas strategi mitigasi dan pengurangan dampak bencana. Oleh sebab itu, hasil penelitian ini diharapkan tidak hanya mempertegas efektivitas masing-masing model, tetapi juga memberikan rekomendasi berbasis bukti ilmiah bagi pengambil kebijakan dan pemangku kepentingan dalam merancang perencanaan tata ruang serta tindakan pencegahan yang lebih responsif dan terukur. Secara keseluruhan, analisis mendalam terhadap perbandingan performa kedua model prediksi probabilitas gempa ini diharapkan dapat menjadi pijakan utama dalam memperkuat ketahanan kawasan Jawa Barat terhadap ancaman seismik yang terus berkembang.

Untuk merealisasikan tujuan tersebut, penelitian ini akan mengintegrasikan himpunan data (*dataset*) historis kegempaan yang diakses melalui WebGIS Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dengan data statistik kewilayahan dari Badan Pusat Statistik (BPS) Provinsi Jawa Barat. Penggabungan kedua sumber data ini tidak hanya merepresentasikan populasi dan sampel aktivitas seismik, tetapi juga menjadi parameter krusial untuk memperkirakan tingkat keterpaparan serta besaran dampak kerugian sosial-ekonomi jika gempa bumi terjadi di wilayah tersebut. Mengingat karakteristik data gempa yang pada umumnya bersifat tidak seimbang (*imbalanced data*) dan memiliki pola persebaran yang sangat kompleks, diperlukan instrumen analitik yang tangguh. Dalam konteks ini, model *Random Forest* dihipotesiskan akan menunjukkan performa dan tingkat akurasi yang lebih unggul dibandingkan *Support Vector Machine* (SVM). Keunggulan *Random Forest* sebagai algoritma berbasis *ensemble tree* memungkinkannya untuk mengelola ketidakseimbangan kelas data spasial secara lebih stabil dan meminimalkan risiko bias prediksi (*overfitting*) dengan lebih optimal dibandingkan pendekatan *hyperplane* pada SVM. Beranjak dari realitas fenomena tektonik, urgensi pemodelan yang berorientasi pada dampak mitigasi, hingga hipotesis yang telah diuraikan, maka peneliti bermaksud untuk mengkaji permasalahan ini secara komprehensif dalam sebuah penelitian skripsi yang berjudul “Analisis Komparatif Model Pembelajaran Mesin *Random Forest* dan *Support Vector Machine* dalam Memprediksi Probabilitas Gempa di Jawa Barat”.

1.2 Rumusan Masalah

1. Bagaimana perbandingan performa model *Random Forest* dan *Support Vector Machine* dalam memprediksi probabilitas gempa bumi di Jawa Barat, khususnya dalam mengatasi masalah data yang tidak seimbang (*imbalanced data*) dan berkarakteristik non-linier?

2. Bagaimana perbandingan kemampuan kedua model tersebut dalam memetakan pola probabilitas spasial-temporal gempa bumi, serta integrasinya dengan data kerentanan wilayah di Jawa Barat?
3. Berdasarkan hasil komparasi kinerja dan pemetaan spasial, model manakah yang paling optimal untuk direkomendasikan sebagai instrumen pendukung keputusan dalam sistem peringatan dini dan strategi mitigasi risiko bencana di Jawa Barat?

1.3 Tujuan Penelitian

1. Menganalisis dan membandingkan performa model Random Forest dan Support Vector Machine dalam memprediksi probabilitas gempa bumi di Jawa Barat, dengan fokus pada keandalan kedua model dalam menangani data yang tidak seimbang (imbalanced data) serta berkarakteristik non-linier.
2. Mengevaluasi perbandingan kemampuan kedua model tersebut dalam memetakan pola probabilitas kegempaan secara spasial-temporal, sekaligus melihat keselarasan hasil pemetaan tersebut jika diintegrasikan dengan data kerentanan wilayah di Jawa Barat.
3. Menentukan dan merekomendasikan model machine learning yang paling optimal dan akurat untuk diaplikasikan sebagai instrumen pendukung pengambilan keputusan strategis dalam sistem peringatan dini serta perencanaan mitigasi risiko bencana di Jawa Barat.

1.4 Manfaat Penelitian

A. Manfaat Teoritis:

Studi ini secara signifikan berkontribusi pada pengayaan literatur dalam domain data science dan seismologi terapan, dengan memberikan pencermatan kritis atas distingsi performa dua model pembelajaran mesin Random Forest, dan Support Vector Machine dalam konteks prediksi probabilitas gempa di wilayah berkarakteristik data tidak seimbang serta non-linier. Hasil analisis komparatif yang dihasilkan mampu memperluas wawasan konseptual terkait

pemilihan dan penerapan algoritma machine learning untuk pemetaan spasial-temporal risiko kegempaan, sekaligus mempertegas urgensi integrasi antara pendekatan statistik-analitik dengan data kerentanan wilayah. Temuan penelitian ini diharapkan memfasilitasi elaborasi teoritis mengenai efektivitas model prediktif pada situasi data dan fenomena kebencanaan yang kompleks dan dinamis, serta menawarkan dasar argumentatif untuk pengembangan konsep hybrid modeling pada isu mitigasi bencana, khususnya di bidang geoinformatika dan disaster risk reduction.

B. Manfaat Praktis:

Secara praktis, hasil penelitian ini menyediakan referensi empirik dan rekomendatif dalam optimalisasi model machine learning sebagai instrumen pendukung pengambilan keputusan strategis pada sistem peringatan dini dan perencanaan mitigasi risiko bencana di Provinsi Jawa Barat. Dengan mengevaluasi dan mengidentifikasi model prediktif paling akurat, penelitian ini mampu meningkatkan efektivitas pengelolaan data spasial-temporal kejadian gempa sebagai basis mitigasi berbasis sains, memperkuat kapabilitas lembaga terkait dalam menyiapkan intervensi antisipatoris, serta mendorong penguatan sistem monitoring bencana yang adaptif terhadap variasi kerentanan wilayah. Implikasi aplikatif dari studi ini diharapkan dapat diaplikasikan secara langsung pada pengembangan kebijakan perlindungan masyarakat, penguatan tata kelola mitigasi risiko kegempaan, dan inovasi pada ekosistem informasi kebencanaan yang terintegrasi di tingkat pemerintah daerah maupun nasional.

1.5 Batasan Masalah

Penelitian komparatif algoritma pembelajaran mesin untuk prediksi kegempaan memerlukan batasan ruang lingkup yang jelas guna memastikan validitas temuan, replikasi metodologis, serta menjaga fokus penelitian. Mengingat kompleksitas spasial-temporal data kegempaan di Jawa Barat serta tantangan inherent pada data

yang tidak seimbang (imbalanced data), penelitian ini secara eksplisit membatasi ruang lingkupnya pada parameter operasional dan metodologis berikut:

1. Fokus Algoritma Pembelajaran Mesin:

Analisis secara eksklusif dikonstruksi pada perbandingan kinerja prediktif antara model Random Forest dan Support Vector Machine (SVM). Ruang lingkup ini tidak mencakup eksplorasi algoritma lain, hibridisasi model, maupun penggunaan meta-ensemble lebih lanjut.

2. Sumber Data dan Variabel:

Dataset yang digunakan terbatas pada data historis kejadian gempa bumi yang diakses melalui WebGIS BMKG dan data statistik karakteristik wilayah (seperti kepadatan penduduk dan infrastruktur) dari BPS Provinsi Jawa Barat. Aspek prediktif dibatasi hanya pada variabel yang termaktub dalam kedua sumber data tersebut, tanpa integrasi data eksternal lain seperti citra satelit atau sensor in-situ.

3. Area Studi dan Ruang Spasial:

Unit analisis spasial didefinisikan secara ketat sesuai batas administratif Provinsi Jawa Barat. Penelitian ini tidak membahas probabilitas gempa untuk wilayah di luar cakupan administratif tersebut, meskipun secara geografis berdekatan.

4. Batasan Temporal Data:

Seluruh analisis prediksi dan evaluasi kinerja model dibatasi pada periode data historis gempa yang tersedia secara publik pada WebGIS BMKG hingga rentang waktu terakhir saat pengumpulan data dilakukan, tanpa melakukan simulasi atau proyeksi ke masa depan di luar cakupan data aktual tersebut.

5. Skenario Evaluasi Kinerja dan Preprocessing:

Proses persiapan data, termasuk penanganan kelas data yang tidak seimbang, difokuskan pada praktik standar dalam literatur machine

learning. Evaluasi performa model dibatasi pada metrik utama seperti Accuracy, Precision, Recall, F1-Score, dan Area Under the Curve (AUC).

6. Metodologi dan Teknik Implementasi:

Desain eksperimen mengadopsi pendekatan kuantitatif dengan pipeline implementasi standar (pembagian data training/testing dan validasi) menggunakan pustaka (library) umum seperti Scikit-Learn. Pengujian model dibatasi pada eksperimen berbasis data sekunder secara offline, tanpa uji validasi eksternal berbasis data baru atau implementasi live-testing pada sistem monitoring aktual.

Dengan batasan-batasan tegas di atas, penelitian ini diharapkan mampu menghasilkan analisis komparatif yang terukur dan tajam mengenai performa Random Forest dan SVM, tanpa meluas pada variabel atau kondisi di luar kendali teknis penelitian

1.6 Mata Kuliah yang Mendasari Penelitian

Penelitian ini dibangun di atas fondasi utama mata kuliah Machine Learning. Teori dan praktik yang dipelajari dalam mata kuliah ini menjadi landasan krusial bagi pengembangan sistem prediksi probabilitas gempa bumi yang dilakukan. Penerapan mata kuliah ini dalam penelitian mencakup aspek-aspek berikut:

1. Penerapan *Supervised Learning*: Konsep pembelajaran terawasi (*supervised learning*) digunakan sebagai kerangka dasar untuk melatih model memprediksi target biner (kejadian gempa) berdasarkan data historis berlabel. Pemahaman ini diterapkan untuk memetakan variabel independen (seperti magnitudo masa lalu dan lokasi grid) menjadi probabilitas prediksi yang akurat.
2. Algoritma Klasifikasi Non-Linear: Teori mendalam mengenai algoritma kompleks diterapkan secara langsung dalam pemilihan model Random Forest dan Support Vector Machine (SVM). Mata kuliah ini mendasari pemahaman tentang bagaimana metode *ensemble* pada Random Forest mampu menangani

overfitting serta bagaimana *Kernel Trick* pada SVM efektif dalam memisahkan data seismik yang memiliki pola distribusi non-linear.

3. Teknik *Feature Engineering* dan *Preprocessing*: Mata kuliah ini memberikan landasan teknis dalam pengolahan data mentah menjadi format yang siap dimodelkan. Hal ini mencakup teknik normalisasi data, penanganan *missing values*, serta strategi penanganan ketidakseimbangan kelas (*imbalanced data*) yang sangat umum ditemukan pada dataset bencana alam.
4. Evaluasi dan Validasi Model: Metodologi untuk mengukur kinerja model secara objektif diadopsi dari mata kuliah ini. Penggunaan metrik evaluasi seperti *Area Under the ROC Curve* (AUC-ROC), *Brier Score*, dan *Log Loss*, serta teknik validasi silang (*cross-validation*) berbasis waktu (*time-series split*), diterapkan guna memastikan bahwa model yang dibangun memiliki kemampuan generalisasi yang baik terhadap data baru.

