PENGEMBANGAN MODEL KLASIFIKASI POTENSI TSUNAMI BERDASARKAN DATA GEMPA BUMI DENGAN METODE *ENSEMBLE LEARNING*

SKRIPSI

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh:
Cristian Roky
D1041211020

FAKULTAS TEKNIK
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Cristian Roky
Nim : D1041211020

menyatakan bahwa dalam skripsi yang berjudul "*Pengembangan Model Klasifikasi Potensi Tsunami Berdasarkan Data Gempa Bumi dengan Metode Ensemble Learning*" tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi manapun. Sepanjang pengetahuan Saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Saya sanggup menerima konsekuensi akademis dan hukum di kemudian hari apabila pernyataan yang dibuat ini tidak benar.

Pontianak, 23 Juli 2025

Cristian Roky
NIM D1041211020

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS TANJUNGPURA
FAKULTAS TEKNIK
Jalan Prof. Dr. H. Hadari Nawawi Pontianak 78124
Telepon (0561) 740186, WA: +6282152280907
Email : ft@untan.ac.id Website : http://teknik.untan.ac.id

HALAMAN PENGESAHAN

PENGEMBANGAN MODEL KLASIFIKASI POTENSI TSUNAMI BERDASARKAN DATA GEMPA BUMI DENGAN METODE *ENSEMBLE LEARNING*

Program Studi Sarjana Informatika
Jurusan Informatika

Oleh :

Cristian Roky
NIM D1041211020

Telah dipertahankan di depan Penguji Skripsi pada tanggal 23 Juli 2025
dan diterima sebagai salah satu persyaratan untuk memperoleh gelar sarjana.

Susunan Penguji Skripsi:

Ketua,

Rina Septiriana, S.T., M.Cs.
NIP 198709232020122001

Penguji Utama,

Tursina, S.T., M.Cs.
NIP 197801152002122003

Sekretaris,

Anggi Perwitasari, S.T., M.T.
NIP 198908192019032012

Penguji Pendamping,

M. Tsana'uddin Farid, S.Kom., M.Eng.
NIP 199609162024061001

Pontianak, 23 Juli 2025
Dekan,

Dr.-Ing. Ir. Slamet Widodo, M.T., IPM
NIP 196712231992031002

HALAMAN PERSEMBAHAN

Skripsi ini didedikasikan kepada kedua orang tua saya (Bapak NG KWET SA dan Ibu Eni) dan saudaraku (Ferdi dan Agus Trio Susanto). Apa yang mereka telah berikan melebihi daripada apa yang saya inginkan.

Dari titisan ilmu-Mu kupersembahkan goresan tanganku bagi ilmu pengetahuan Indonesia

PRAKATA

Puji syukur kami panjatkan kehadirat Tuhan Yang Maha Esa karena kasih dan karunia-Nya, sehingga penulis dapat menyelesaikan Laporan Penelitian Tugas Akhir, terhitung mulai tanggal 16 November 2025 – 23 Juli 2025. Penulis mengucapkan terima kasih kepada semua pihak yang telah mendukung selama penulis melakukan penelitian dan juga membantu penyusunan laporan ini khususnya:

1. Ibu Rina Septiriana, S.T, M.Cs dan Ibu Anggi Perwitasari S.T, M.T selaku Dosen Pembimbing tugas akhir yang telah memberikan banyak arahan dan masukkan dalam membimbing penulis, sehingga dapat menyelesaikan penelitian ini dengan baik.
2. Ibu Tursina, S.T., M.Cs dan Bapak M. Tsana'uddin Farid, S.Kom., M.Eng selaku Dosen Penguji tugas akhir yang telah memberikan banyak masukan dan saran dalam menyelesaikan penelitian ini.
3. Segenap dosen Jurusan Informatika atas segala ilmu dan bimbingannya.
4. Semua pihak yang telah membantu penulisan skripsi yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak kekurangan dan jauh dari kesempurnaan. Penulis berharap semoga skripsi ini dapat memberikan manfaat bagi kita semua.

Pontianak, 23 Juli 2025
Penulis,

Cristian Roky

ABSTRAK

Tsunami merupakan gelombang laut besar yang disebabkan oleh gangguan tiba-tiba di dasar laut, seperti gempa bumi tektonik atau letusan gunung berapi. Penelitian ini bertujuan untuk membandingkan performa teknik Bagging (*Random Forest*) dan *Boosting* (XGBoost) dalam mengklasifikasikan potensi tsunami berdasarkan data gempa bumi. Selain itu, penelitian ini juga menganalisis pengaruh imputasi dan SMOTE-ENN terhadap performa model. Data yang akan digunakan terdiri dari 4000 data, yang bersumber dari United States Geological Survey (USGS). Model klasifikasi dikembangkan secara menyeluruh dari awal (*from scratch*), mulai dari proses pengumpulan data hingga pemodelan, tanpa menggunakan model bawaan yang telah tersedia. Enam model dibangun berdasarkan kombinasi imputasi data, SMOTE-ENN, serta algoritma XGBoost atau *Random Forest*. *Hyperparameter tuning* dilakukan menggunakan teknik *Bayesian Optimization* untuk memperoleh kombinasi *hyperparameter* terbaik setiap model. Hasil klasifikasi menunjukkan bahwa baik XGBoost maupun *Random Forest* mampu mengklasifikasikan potensi tsunami dengan baik, bahkan ketika menghadapi masalah data tidak lengkap dan distribusi kelas yang tidak seimbang. Model XGBoost menunjukkan performa terbaik dengan akurasi tertinggi sebesar 0,964 pada skenario imputasi tanpa SMOTE-ENN, mengindikasikan keunggulan model ini dalam menangani data gempa. SMOTE-ENN terbukti efektif meningkatkan *recall* pada kelas tsunami, namun menurunkan *precision*. Proses pengembangan model disusun dalam bentuk modul Python, sehingga setiap tahap mulai dari *data preprocessing* hingga pemodelan dapat dilakukan secara efisien. Modul ini dapat memudahkan proses replikasi dan pengembangan lanjutan secara fleksibel. Selain itu, modul ini juga memungkinkan penerapan teknik imputasi untuk mengatasi *missing value* dan SMOTE-ENN untuk mengatasi ketidakseimbangan data, sehingga menghasilkan model klasifikasi yang lebih andal.

Kata Kunci : Klasifikasi Potensi Tsunami, XGBoost, *Random Forest*, SMOTE-ENN.

ABSTRACT

Tsunamis are large ocean waves triggered by sudden disturbances on the seafloor, such as tectonic earthquakes or volcanic eruptions. This study aims to compare the performance of Bagging (Random Forest) and Boosting (XGBoost) techniques in classifying tsunami potential based on earthquake data. In addition, the study also analyzes the impact of imputation and SMOTE-ENN on model performance. The dataset used consists of 4,000 entries sourced from the United States Geological Survey (USGS). The classification models are developed entirely from scratch without relying on pre-built models. Six models were built based on combinations of data imputation, SMOTE-ENN, and either the XGBoost or Random Forest algorithm. Hyperparameter tuning was performed using Bayesian Optimization to obtain the best hyperparameter combinations for each model. The classification results show that both XGBoost and Random Forest are capable of accurately classifying tsunami potential, even when faced with incomplete data and imbalanced class distributions. The XGBoost model achieved the best performance with the highest accuracy of 0.964 in the imputation-only scenario without SMOTE-ENN, indicating its superiority in handling earthquake data. SMOTE-ENN proved effective in improving recall for the tsunami class, although it reduced precision. The model development process is organized into Python modules, allowing each stage—from data preprocessing to build model—to be conducted efficiently. These modules facilitate flexible replication and further development. Moreover, the modular design enables the application of imputation techniques to handle missing values and SMOTE-ENN to address data imbalance, resulting in more robust classification models.

Keywords : Tsunami classification, XGBoost, Random Forest, SMOTE-ENN

DAFTAR ISI

HALAMAN PERNYATAAN ii
HALAMAN PENGESAHAN iii
HALAMAN PERSEMBAHAN iv
PRAKATA v
ABSTRAK vi
ABSTRACT vii
DAFTAR ISI viii
DAFTAR TABEL xii
DAFTAR GAMBAR xiii
DAFTAR KODE PROGRAM xv
DAFTAR LAMPIRAN xvi
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Pembatasan Masalah 3
BAB II TINJAUAN PUSTAKA 5
2.1 Kajian Terkait 5
2.2 Gempa Bumi 8
2.3 Tsunami 9
2.4 *Machine Learning* (ML) 9
2.4.1 Supervised Learning 9
2.4.2 Unsupervised Learning 10
2.4.3 Reinforcement Learning 10

2.5 *Ensemble Learning*..... 10
2.5.1 *Bootstrap Aggregating* (Bagging).....11
2.5.2 *Boosting*.....11
2.5.3 Stacking..... 12
2.6 Klasifikasi 12
2.7 *Extreme Gradient Boosting* (XGBoost) 12
2.8 *Random Forest* 13
2.9 Reverse Geocoding..... 15
2.10 Imbalance Data..... 15
2.11 SMOTE-ENN..... 15
2.12 Python 16
2.13 Hyperparameter Tuning..... 16
2.14 Stratified K-Fold Cross-Validation 17
2.15 Pearson Correlation Coefficient 17
2.16 Confusion Matrix 18
2.17 Modul 19
2.18 Pengujian Unit..... 19
2.19 Pengujian Integrasi..... 20
2.20 Streamlit..... 20
BAB III METODOLOGI PENELITIAN..... 21
3.1 Alat Bantu Penelitian..... 21
3.1.1 Perangkat Lunak..... 21
3.1.2 Perangkat Keras 21
3.2 Data Penelitian 21
3.3 Metodologi Penelitian 24
3.3.1 Pengumpulan Data 25

3.3.2 *Exploratory Data Analysis* (EDA) 26
3.3.3 Data Preprocessing 26
3.3.4 Pemodelan 28
3.3.5 Evaluasi Model 41
3.3.6 Analisis Hasil 42
3.3.7 Implementasi 42
3.3.8 Penarikan Kesimpulan 43
BAB IV IMPLEMENTASI DAN HASIL 44
4.1 Pengumpulan Data 44
4.2 *Exploratory Data Analysis* (EDA) 44
4.3 Data Preprocessing 48
4.3.1 Data Cleaning 48
4.3.2 Feature Construction 49
4.3.3 Data Transformation 51
4.3.4 Feature Selection 52
4.4 Pemodelan 56
4.5 Hasil Evaluasi 65
4.5.1 Hasil Validasi Model 65
4.5.2 Hasil Pengujian Model 74
4.6 Analisis Hasil 81
4.6.1 Analisis Hasil Klasifikasi 81
4.6.2 Analisis Hasil Penelitian 86
4.7 Hasil Implementasi 89
4.8 Pengujian Implementasi 92
4.8.1 Pengujian Unit 92
4.8.2 Pengujian Integrasi 95

BAB V KESIMPULAN 98
5.1 Kesimpulan 98
5.2 Saran 99
DAFTAR PUSTAKA 100

DAFTAR TABEL

Tabel 2. 1 Rangkuman Kajian Terkait 7
Tabel 2. 2 *Confusion Matrix* 19
Tabel 3. 1 Deskripsi Data Utama 22
Tabel 3. 2 Deskripsi Data Tambahan 23
Tabel 3. 3 Daftar *Hyperparameter* yang Disetel (*Tuning*) 31
Tabel 4. 1 Skenario Pemodelan 56
Tabel 4. 2 Hasil Validasi Silang Model Pertama 65
Tabel 4. 3 Hasil Validasi Silang Model Kedua 67
Tabel 4. 4 Hasil Validasi Silang Model Ketiga 68
Tabel 4. 5 Hasil Validasi Silang Model Keempat 69
Tabel 4. 6 Hasil Validasi Silang Model Kelima 70
Tabel 4. 7 Hasil Validasi Silang Model Keenam 72
Tabel 4. 8 Perbandingan Hasil Klasifikasi 82
Tabel 4. 9 Hasil Pengujian Unit 92
Tabel 4. 10 Hasil Pengujian Integrasi 95

DAFTAR GAMBAR

Gambar 2. 1 *Stratified K-Fold Cross Validation* 17
Gambar 3. 1 Metodologi Penelitian 24
Gambar 3. 2 Diagram *Pipe and Filter* 25
Gambar 3. 3 Fitur Sources dan Source_Counter 27
Gambar 3. 4 Fitur *Location* 27
Gambar 3. 5 Fitur MagType, Net, dan *Location* 28
Gambar 3. 6 Diagram Alir Proses Boosting pada XGBoost 30
Gambar 3. 7 Diagram Alir Proses Bagging pada *Random Forest* 31
Gambar 3. 8 Diagram Alir *Bayesian Optimization* 33
Gambar 3. 9 Diagram Alir SMOTE-ENN 34
Gambar 3. 10 Ilustrasi Pengaruh SMOTE-ENN terhadap Distribusi Data 35
Gambar 3. 11 Alur Kerja Skenario Pertama 36
Gambar 3. 12 Alur Kerja Skenario Kedua 37
Gambar 3. 13 Alur Kerja Skenario Ketiga 38
Gambar 3. 14 Alur Kerja Skenario Keempat 39
Gambar 3. 15 Alur Kerja Skenario Kelima 40
Gambar 3. 16 Alur Kerja Skenario Keenam 41
Gambar 3. 17 Diagram Komponen 43
Gambar 4. 1 Hasil Pengumpulan Data 44
Gambar 4. 2 *Missing Value* pada Setiap Kolom 45
Gambar 4. 3 Jumlah *Outlier* pada Setiap Fitur 46
Gambar 4. 4 Korelasi *Pearson* antar Fitur 47
Gambar 4. 5 Distribusi Gempa 48
Gambar 4. 6 Hasil Pembagian Data 49
Gambar 4. 7 Jumlah Sumber Data 50
Gambar 4. 8 Visualisasi Data *Location* 51
Gambar 4. 9 Hasil *Encoding* 52
Gambar 4. 10 Hasil *Feature Importance* XGBoost 53
Gambar 4. 11 Hasil Seleksi Fitur dengan XGBoost 54
Gambar 4. 12 Hasil *Feature Importance Random Forest* 55
Gambar 4. 13 Hasil Seleksi Fitur dengan *Random Forest* 56

Gambar 4. 14 Perbedaan Distribusi Kelas Sebelum dan Setelah SMOTE-ENN 57
Gambar 4. 15 Sampel Data yang Dihasilkan SMOTE-ENN..... 58
Gambar 4. 16 Hasil *Hyperparameter Tuning* Pemodelan Skenario Pertama..... 59
Gambar 4. 17 Hasil *Hyperparameter Tuning* Pemodelan Skenario Kedua..... 60
Gambar 4. 18 Hasil *Hyperparameter Tuning* Pemodelan Skenario Ketiga..... 61
Gambar 4. 19 Hasil *Hyperparameter Tuning* Pemodelan Skenario Keempat..... 62
Gambar 4. 20 Hasil *Hyperparameter Tuning* Pemodelan Skenario Kelima 63
Gambar 4. 21 Hasil *Hyperparameter Tuning* Pemodelan Skenario Keenam..... 64
Gambar 4. 22 Visualisasi Perbandingan Skor F1 Model Pertama..... 66
Gambar 4. 23 Visualisasi Perbandingan Skor F1 Model Kedua 67
Gambar 4. 24 Visualisasi Perbandingan Skor F1 Model Ketiga 69
Gambar 4. 25 Visualisasi Perbandingan Skor F1 Model Keempat 70
Gambar 4. 26 Visualisasi Perbandingan Skor F1 Model Kelima..... 71
Gambar 4. 27 Visualisasi Perbandingan Skor F1 Model Keenam 73
Gambar 4. 28 *Heatmap* Validasi Model 74
Gambar 4. 29 Hasil Pengujian Model Pertama 75
Gambar 4. 30 *Confusion Matrix* Model Pertama 76
Gambar 4. 31 Hasil Pengujian Model Kedua..... 76
Gambar 4. 32 *Confusion Matrix* Model Kedua..... 77
Gambar 4. 33 Hasil Pengujian Model Ketiga..... 77
Gambar 4. 34 *Confusion Matrix* Model Ketiga..... 78
Gambar 4. 35 Hasil Pengujian Model Keempat..... 78
Gambar 4. 36 *Confusion Matrix* Model Keempat..... 79
Gambar 4. 37 Hasil Pengujian Model Kelima..... 79
Gambar 4. 38 *Confusion Matrix* Model Kelima..... 80
Gambar 4. 39 Hasil Pengujian Model Keenam 80
Gambar 4. 40 *Confusion Matrix* Model Keenam 81
Gambar 4. 41 Perbandingan Akurasi..... 82
Gambar 4. 42 Perbandingan *Precision* 84
Gambar 4. 43 Perbandingan *Recall* 85
Gambar 4. 44 Perbandingan Skor F1..... 86
Gambar 4. 45 Tampilan Antarmuka Aplikasi 92

DAFTAR KODE PROGRAM

Kode Program 4. 1 Pembagian Data 47
Kode Program 4. 2 Pembuatan Fitur Source_Counter 48
Kode Program 4. 3 Pembuatan Fitur Location 49
Kode Program 4. 4 Transformasi Data Kategorikal 50
Kode Program 4. 5 *Feature Importance* XGBoost 51
Kode Program 4. 6 Seleksi Fitur dengan XGBoost 51
Kode Program 4. 7 *Fitur Importance Random Forest* 52
Kode Program 4. 8 Seleksi Fitur dengan *Random Forest* 53
Kode Program 4. 9 Pemodelan Skenario Pertama 57
Kode Program 4. 10 Pemodelan Skenario Kedua 58
Kode Program 4. 11 Pemodelan Skenario Ketiga 59
Kode Program 4. 12 Pemodelan Skenario Keempat 60
Kode Program 4. 13 Pemodelan Skenario Kelima 61
Kode Program 4. 14 Pemodelan Skenario Keenam 62

DAFTAR LAMPIRAN

Lampiran A Hasil Hyperparameter Tuning A-1
Lampiran B Pengujian Unit dan Pengujian Integrasi B-1
Lampiran C Repository GitHub C-1

BAB I
PENDAHULUAN

1.1 Latar Belakang

Tsunami adalah gelombang laut besar yang dihasilkan oleh gangguan tiba-tiba di dasar laut, seperti pergerakan lempeng tektonik akibat gempa bumi, letusan gunung berapi, maupun tanah longsor di bawah laut. Indonesia terletak di pertemuan tiga lempeng tektonik utama, yaitu Eurasia, Indo-Australia, dan Pasifik menjadikannya salah satu wilayah paling rawan gempa di dunia. Menurut (Reid & Mooney, 2023), frekuensi terjadinya tsunami besar adalah sekitar 2,5 kali per tahun, dan sekitar 80% tsunami disebabkan oleh gempa bumi. Peristiwa tsunami besar, seperti di Samudra Hindia pada tahun 2004 dan Jepang pada tahun 2011, menunjukkan dampak destruktif yang luar biasa, mencakup kerusakan infrastruktur dan korban jiwa dalam skala besar. Saat ini, deteksi tsunami dilakukan menggunakan sistem *Deep-Ocean Assessment and Reporting of Tsunami* (DART), yaitu gabungan sensor di dasar laut dan pelampung yang mengirim data lewat satelit. Sistem ini membutuhkan biaya infrastruktur yang tinggi untuk pengadaan dan perawatan, serta belum tersedia di seluruh wilayah Indonesia. Selain itu, sistem ini membutuhkan waktu dalam analisis data sensor, oleh sebab itu diperlukan pendekatan berbasis data yang mampu melakukan klasifikasi secara cepat dan akurat. Salah satu pendekatan yang dapat digunakan adalah dengan model *machine learning* yang dilatih menggunakan data gempa bumi untuk mengklasifikasikan potensi tsunami.

Machine learning merupakan pembelajaran oleh mesin menggunakan data. Dalam konteks klasifikasi potensi tsunami, *machine learning* menawarkan keunggulan dari segi biaya dan kecepatan dalam mendeteksi potensi tsunami. Dengan memanfaatkan data gempa seperti magnitudo, kedalaman, dan lokasi, algoritma *machine learning* dapat dilatih untuk mengenali pola-pola yang berhubungan dengan kejadian tsunami. Klasifikasi ini juga mendukung pengembangan model klasifikasi bencana yang berbasis data ilmiah dan dapat membantu pemerintah dalam merancang kebijakan mitigasi risiko yang lebih baik. Oleh karena itu, klasifikasi potensi tsunami merupakan hal yang penting dalam pengelolaan mitigasi bencana. Dalam konteks ini, *Ensemble Learning*

digunakan untuk mengklasifikasi potensi tsunami berdasarkan data gempa bumi. *Ensemble Learning* merupakan sebuah metode dalam *machine learning* yang menggabungkan model untuk meningkatkan akurasi dan keandalan hasil. *Ensemble Learning* menggabungkan hasil dari banyak model sehingga menghasilkan prediksi yang lebih stabil dan akurat (Mienye & Sun, 2022).

Bootstrap Aggregating (Bagging) dan *Boosting* merupakan teknik di dalam *Ensemble Learning* yang terbukti efektif dalam mengurangi kesalahan, karena dibangun berdasarkan banyak pohon. Bagging bekerja dengan membangun model independen secara paralel, sedangkan *Boosting* secara iteratif meningkatkan kinerja model dengan memperbaiki kesalahan dari model sebelumnya. Berdasarkan penelitian (S. Wang et al., 2021), teknik Bagging pada *Random Forest* menunjukkan kinerja yang lebih baik dalam mengklasifikasikan daerah rawan longsor, sedangkan pada penelitian (Fatima et al., 2023), teknik *Boosting* pada XGBoost lebih unggul dalam mengklasifikasikan kelas keselamatan penumpang kapal.

Dalam penelitian ini, XGBoost dan *Random Forest* dipilih sebagai representasi dari masing-masing teknik *Ensemble Learning*, yaitu *Boosting* dan Bagging, karena keduanya telah terbukti secara luas sebagai algoritma yang andal dan efisien (Mienye & Sun, 2022). *Random Forest* merupakan algoritma Bagging yang banyak digunakan karena kemampuannya dalam mengurangi varians sehingga mencegah *overfitting* dan meningkatkan stabilitas prediksi. Di sisi lain, XGBoost adalah algoritma *Boosting* yang dikenal unggul dalam mengurangi bias sehingga mencegah *underfitting*, serta meningkatkan akurasi. Oleh karena itu, diperlukan penelitian lebih lanjut untuk membandingkan kedua algoritma ini, guna memahami kekuatan dan kelemahan masing-masing algoritma serta menentukan algoritma yang paling efektif dalam mengklasifikasikan gempa yang berpotensi tsunami. Penelitian ini tidak hanya membandingkan performa model, tetapi juga mengembangkan model klasifikasi dari awal (*from scratch*). Penelitian ini menjadi penting mengingat studi mengenai klasifikasi potensi tsunami masih terbatas, baik dari segi perbandingan algoritma maupun ketersediaan implementasi praktisnya. Selain itu, hasil perbandingan performa model *machine learning* yang telah dilakukan umumnya sulit diakses dan digunakan ulang oleh

pengguna lain (Semmelrock et al., 2025). Oleh karena itu, penelitian ini membangun modul Python yang tidak hanya merepresentasikan hasil analisis performa, tetapi juga memungkinkan implementasi ulang perbandingan performa algoritma secara terstruktur. Dengan demikian, penelitian ini diharapkan berkontribusi secara akademik melalui pengayaan literatur *Ensemble Learning* dalam kebencanaan dan secara praktis melalui implementasi modul klasifikasi yang mudah diterapkan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana performa teknik Bagging (*Random Forest*) dibandingkan dengan *Boosting* (XGBoost) dalam mengklasifikasikan potensi tsunami berdasarkan data gempa bumi?
2. Bagaimana membangun modul Python yang dapat membandingkan performa model klasifikasi untuk mendukung proses pengembangan model?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah sebagai berikut:

1. Membandingkan performa teknik Bagging (*Random Forest*) dan *Boosting* (XGBoost) untuk menentukan algoritma yang paling sesuai dalam mengklasifikasikan potensi tsunami berdasarkan data gempa bumi.
2. Membangun modul Python yang dapat membandingkan performa model klasifikasi untuk mendukung proses pengembangan model.

1.4 Pembatasan Masalah

Peneliti menetapkan batasan masalah untuk mencegah terjadinya penyimpangan dalam penyusunan penelitian ini. Berikut adalah batasan masalah pada penelitian ini.

1. Algoritma yang digunakan dalam *Ensemble Learning* terbatas pada dua algoritma, yaitu XGBoost yang mewakili pendekatan *Boosting*

dan *Random Forest* yang mewakili pendekatan Bagging, tanpa melibatkan algoritma *Ensemble Learning* lainnya.

2. *Data preprocessing* dilakukan sesuai kebutuhan pada setiap skenario dan dilakukan saat proses pembangunan model. Penanganan *outlier* dan imputasi dilakukan pada skenario yang menggunakan SMOTE-ENN. *Data preprocessing* ini dilakukan di dalam tahap pembangunan model untuk mencegah kebocoran data. Pada skenario yang tidak membutuhkan SMOTE-ENN dan imputasi, data akan digunakan dalam bentuk aslinya tanpa perubahan.
3. Pengembangan model merujuk pada proses menyeluruh yang dimulai dari tahap pengumpulan data hingga pemodelan. Pengembangan model tidak merujuk pada mengembangkan model yang telah ada sebelumnya.
4. *Hyperparameter tuning* menggunakan teknik *Bayesian Optimization* diterapkan pada seluruh skenario model, dengan tujuan untuk memperoleh performa terbaik dari masing-masing model secara adil.
5. Implementasi dilakukan secara terbatas pada pembuatan modul Python yang dilengkapi dengan antarmuka sederhana. Antarmuka ini dirancang untuk menerima masukan berupa data yang terkait dengan gempa bumi, memproses data tersebut melalui modul, dan menghasilkan keluaran sesuai dengan modul yang digunakan.
6. Pengujian terhadap implementasi dilakukan dengan pengujian unit dan integrasi, serta tidak melibatkan pengujian waktu eksekusi (*runtime*) dan memori.