

Pengembangan Sistem Deteksi Dini Gempa menggunakan Metode Convolutional Neural Network (CNN) dan eXtreme Gradient Boosting (XGBoost)

Samatha Marhaendra Putra
Universitas Gadjah Mada
Yogyakarta, Indonesia
sam.marhaendrap@mail.ugm.ac.id

Rizky Alif Ramadhan
Universitas Gadjah Mada
Jakarta, Indonesia
rizky.alif.r@mail.ugm.ac.id

Daffa Bil Nadzary
Universitas Gadjah Mada
Yogyakarta, Indonesia
daffa.bil.nadzary@mail.ugm.ac.id

Abstrak—Gempa merupakan fenomena pergeseran lempeng bumi yang dapat mengakibatkan kerusakan dari ringan hingga fatal. Gelombang gempa dapat dikategorikan menjadi tiga, yaitu P-wave, S-wave, dan *surface wave*. Sistem pendeteksi dini dibuat dengan mendeteksi adanya gelombang P-wave yang bertujuan untuk mendeteksi adanya gempa serta mengestimasi magnitudo dari gempa tersebut. Penelitian ini dilakukan dengan membuat dua sistem yang didasari pada dua model dan pendekatan yang berbeda, yakni model *Convolutional Neural Network* (CNN), serta eXtreme Gradient Boost (XGBoost). Kedua model di-*train* menggunakan dataset Stanford Earthquake Dataset (STEAD) dengan index yang dipilih yaitu sebesar 1000 titik per data. Hasil penelitian menunjukkan bahwa sistem dengan model CNN dapat menghasilkan performa klasifikasi sebesar 0.999, dengan RMSE pada model regresi sebesar 0.83. Sedangkan model XGBoost dapat menghasilkan performa klasifikasi sebesar 0.951 dengan RMSE pada model regresi sebesar 0.368. Hasil simulasi menunjukkan bahwa model dapat memprediksi terjadinya gempa 7 detik setelah P-wave terdeteksi.

Kata kunci—Gempa, deteksi dini, P-wave, CNN, XGBoost, STEAD, magnitudo.

I. LATAR BELAKANG

Gempa merupakan fenomena pergeseran lempeng bumi yang mengakibatkan terjadinya pelepasan energi secara spontan dan menghasilkan gelombang seismik [1]. Sekitar 20 ribu gempa tercatat setiap tahunnya, atau sekitar 55 setiap hari, yang bervariasi skalanya mulai dari ringan hingga berat.

Pergerakan lempeng bumi dapat diukur dengan menggunakan seismograf. Sinyal yang ditangkap pada saat terjadi gempa terbagi menjadi tiga: *Primary Wave* (P-wave), *Secondary Wave* (S-wave), dan *Surface Wave*. P-wave merupakan gelombang yang merambat sangat cepat sehingga akan ditangkap terlebih dahulu oleh sensor. P-wave pada umumnya tidak bersifat destruktif, dan dapat merambat melalui zat padat, cair, ataupun gas. Sedangkan S-wave merupakan gelombang kedua yang ditangkap oleh sensor. Berbeda dengan P-wave, S-wave merambat dibandingkan dengan P-wave serta hanya dapat merambat melalui zat padat.

Salah satu cara untuk menanggulangi jumlah korban yang besar akibat gempa adalah dengan membuat sistem pendeteksi dini gempa. Sistem pendeteksi dini gempa sudah sejak lama diteliti untuk menghasilkan sistem yang akurat dalam memprediksi magnitudo dari gempa, lokasi, hingga waktu. Gempa tergolong susah untuk dideteksi dikarenakan

sifatnya yang termasuk ke dalam jenis bencana alam, dan tidak banyak faktor yang dapat mempengaruhi terjadinya gempa. Sistem pendeteksi dini bertujuan agar informasi mengenai terjadinya gempa dapat dengan secara cepat tersampaikan khususnya di area yang menjadi fokus.

Berbagai riset yang telah dilakukan di bidang ini berfokus pada pendeteksian gelombang P-wave, dikarenakan gelombang ini merambat lebih cepat dibandingkan dengan kedua gelombang destruktif lainnya. Jika sensor mendeteksi adanya P-wave, sistem akan langsung memberikan notifikasi ke pada pusat untuk diinformasikan ke publik.

Zhang, et al. [2] membuat sistem yang dapat mendeteksi magnitudo dari gempa dengan mendeteksi fase dari P-wave yang ditangkap. Fase ini dapat dideteksi dengan melihat perubahan dari *signal-to-noise ratio* (SNR) atau rasio perubahan dari sinyal ke *noise*. Kemudian, Sugondo, et al. [3] juga membuat sistem *Automatic First Arrival Picking* (AFAP), dapat memprediksi waktu kedatangan dari P-wave. Sistem ini bekerja dengan menggunakan bantuan arsitektur *Convolutional Neural Network* (CNN) [4].

Berdasarkan penelitian serupa yang telah dilakukan, penulis mendesain sebuah sistem pendeteksi dini gempa yang berfokus pada pendeteksian gelombang P-wave. Sistem ini dibuat menggunakan dua model yang akan dibandingkan, yakni CNN serta XGBoost. Sistem yang dirancang bertujuan untuk mendeteksi adanya gempa yang akan datang dari gelombang P-wave yang terdeteksi, membedakan antara gelombang gempa dengan *noise*, serta melakukan estimasi dari magnitudo dari gempa yang akan datang.

II. TUJUAN

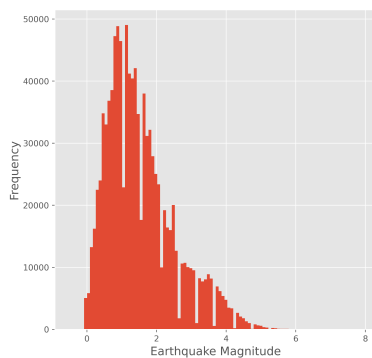
Adapun kegiatan ini dilakukan dengan tujuan:

- 1) Membuat model klasifikasi dan regresi berbasis *machine learning* dan *deep learning* untuk melakukan deteksi dini terhadap gempa yang dapat digunakan secara *real-time*.
- 2) Melakukan komparasi performa antara model klasifikasi dan regresi berbasis *machine learning* dan *deep learning* untuk melakukan deteksi dini terhadap gempa.
- 3) Melakukan simulasi sistem deteksi gempa dini menggunakan model *machine learning* dan *deep learning*.

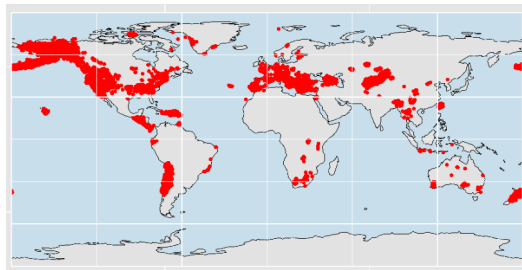
III. METODE

A. Dataset

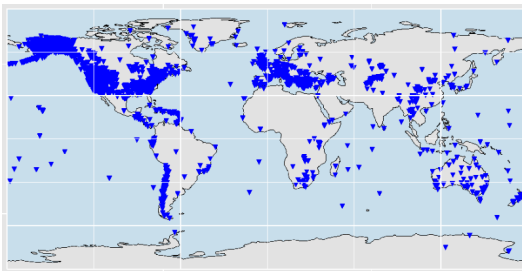
Model sistem pendeteksi dini gempa ini dibuat dengan menggunakan dataset STEAD (Stanford Earthquake Dataset) [5]. STEAD merupakan dataset berukuran besar dan berkala global yang berisikan 1 juta data sinyal gempa yang tercatat di berbagai stasiun serta 300 ribu data *noise* yang berbentuk gelombang. Tiap baris data berisikan 6000 titik hasil *sampling* sinyal yang ditangkap selama proses terjadinya gempa dalam waktu 60 detik, dengan *sampling rate* sebesar 100 Hz. Sinyal ini juga terdiri dari tiga kanal pada tiap baris pengamatannya, yang menangkap gerakan sinyal pada poros utara-selatan, timur-barat, serta secara vertikal. Gambar 3.1 menunjukkan persebaran dari magnitudo gempa serta jarak gempa dari stasiun pada dataset STEAD. Lanjut, Gambar 3.2 juga menunjukkan peta persebaran lokasi terjadinya gempa beserta dengan peta stasiun yang tersedia.



Gambar 3.1 Persebaran magnitudo data gempa yang tercatat.



(a)



(b)

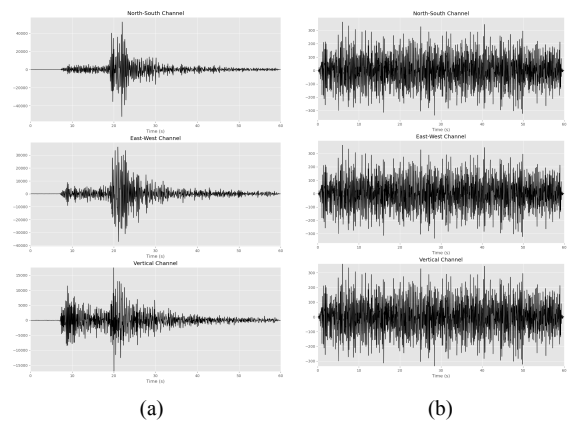
Gambar 3.2 Peta persebaran lokasi terjadinya gempa (a) dan lokasi stasiun penangkap sinyal.

Pada data tersebut juga terdapat meta data yang berisi jenis gelombang, magnitudo, kedalaman, kategori pengamatan, apakah *noise* atau gempa, dan serta fitur lain yang tercatat pada pengamatan gempa.

B. Sinyal Gelombang Gempa

1) Seismometer dan Seismograf

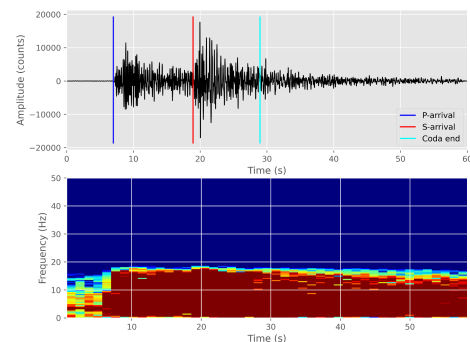
Seismograf adalah sebuah alat yang dapat merekam pergerakan gelombang dari bawah tanah pada saat terjadinya gempa [1]. Di dalam seismograf terdapat sebuah seismometer yang berfungsi sebagai pendulum yang dapat merespon terhadap getaran tersebut. Umumnya, stasiun pengamatan memiliki instrumen yang dapat mencatat tiga sumbu pergerakan: utara-selatan (sumbu y), timur-barat (sumbu x), dan secara vertikal (sumbu z). Di antara ketiga sumbu tersebut, gelombang vertikal mengandung informasi yang lebih banyak dikarenakan jumlah *noise* yang lebih sedikit. Gambar 3.3 menunjukkan contoh hasil gelombang yang ditangkap stasiun dalam tiga kanal.



Gambar 3.3 Contoh hasil pengamatan gelombang gempa pada tiga kanal. (a) Gelombang gempa, (b) *Noise*.

2) *Waveform* dan Spectrogram

Sinyal hasil tangkapan stasiun dapat direpresentasikan ke dalam beberapa bentuk, seperti *waveform*/bentuk gelombang dan spectrogram. Spectrogram merupakan representasi visual dari spektrum frekuensi pada sebuah sinyal yang bersifat *time dependent*. Spektrogram dapat diperoleh melalui transformasi Fourier atau transformasi wavelet [6]. Spectrogram umumnya digunakan pada bidang seismologi dikarenakan kegunaannya yang dapat menyimpan gabungan dari informasi pada domain waktu, frekuensi, serta amplitudo secara langsung. Gambar 3.4 menunjukkan plot *waveform* dan spectrogram dari satu sinyal pengamatan.



Gambar 3.4 Plot *waveform* dan spectrogram dari suatu sinyal pengamatan.

Dari hasil Gambar 3.4 juga menunjukkan bagaimana hasil tangkapan dari sensor terhadap sinyal dalam bentuk *waveform* dapat dipecah menjadi tiga bagian: gelombang P-wave, S-wave, serta *surface wave*. Informasi

mengenai P-wave menjadi sangat penting dikarenakan propertinya yang dapat diakses terlebih dahulu. Pada rentang jarak 50 hingga 500 km, P-wave secara umum dapat merambat dengan kecepatan 8 km per detik, sedangkan S-wave sebesar 3.45 km per detik.

3) Magnitudo dan Intensitas dari Sebuah Gempa

Secara umum, besaran dari sebuah gempa dapat direpresentasikan ke dalam dua besaran: Magnitudo dan intensitas. Intensitas dari suatu gempa merepresentasikan efek dari pergerakan tanah dari titik lokasi, sedangkan magnitudo dari suatu gempa merepresentasikan kekuatan serta 'ukuran' rerata dari sebuah gempa. Magnitudo dari gempa umumnya direpresentasikan ke dalam skala Richter, yang bersifat logaritmik. Sedangkan intensitas dari gempa umumnya direpresentasikan ke dalam bentuk Modified Mercalli Intensity scale (MMI). Berbeda dengan skala Richter, MMI mengkategorikan gempa ke dalam beberapa tingkatan yang dibedakan berdasarkan efek yang dirasakan. Tabel 1 menunjukkan efek yang dirasakan dari gempa berdasarkan tingkatan MMI beserta magnitudo dari gempa [7].

TABEL 1. TABEL SKALA MMI BESERTA DENGAN SKALA RICHTER.

Magnitudo (SR)	Tingkatan	Level MMI	Efek
1.0-1.9	Mikro	I	Tidak terasa, namun tercatat oleh seismograf.
2.0-2.9	Minor	I	Sedikit terasa, tidak ada dampak terhadap bangunan.
3.0-3.9		II s.d. III	Getaran dapat dirasakan dan diamati.
4.0-4.9	Kecil	IV s.d. V	Pergerakan objek-objek di dalam bangunan yang dapat dirasakan.
5.0-5.9	Sedang	VI s.d. VII	Dapat mengakibatkan kerusakan pada bangunan yang rapuh.
6.0-6.9	Kuat	VII s.d. IX	Kerusakan pada kebanyakan bangunan. Dapat dirasakan hingga ratusan km dari pusat gempa.
7.0-7.9	Major	VIII atau lebih tinggi	Kerusakan yang dapat menghancurkan bangunan, dengan jarak yang lebih jauh.

Dari tingkatan tersebut, dapat dikatakan gempa dengan magnitudo >4 sudah cukup untuk membuat panik dan menjadi fokus utama. Pada level IV dan di atas, efek dari gempa sudah dapat dirasakan tidak hanya oleh orang-orang, namun juga mengakibatkan pergerakan dari objek-objek di dalam bangunan.

C. Waveform Feature Extraction

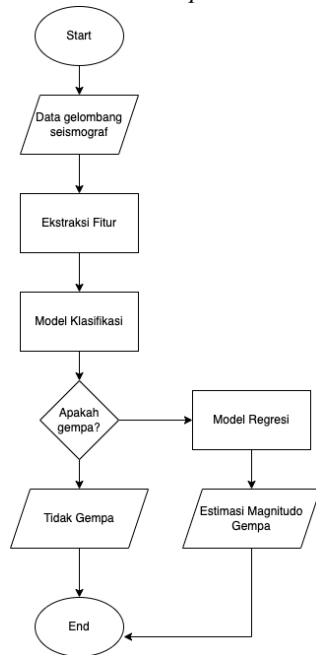
Ekstraksi fitur adalah proses pengurangan dimensi di mana sekumpulan awal data mentah direduksi menjadi grup yang lebih mudah dikelola untuk diproses. Pada gelombang atau sinyal, terdapat dua tipe ekstraksi fitur : Yang pertama adalah domain waktu yang menunjukkan bagaimana sinyal berubah terhadap waktu. Berikut adalah komponen sinial atau gelombang pada domain waktu :

Mean	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$
Max	$\max(x_i)$
Peak	$P_m = \max(x_i)$
Peak to Peak	$P_k = \max(x_i) - \min(x_i)$
RMS	$\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
Variance	$\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}$
Standard Deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$
Power	$\frac{1}{N} \sum_{i=1}^N x_i^2$
Crest Factor	$\frac{P_m}{RMS}$
Form Factor	$\frac{RMS}{\bar{x}}$
Pulse Indicator	$\frac{P_m}{\bar{x}}$
Margin	$\frac{P_m}{\sqrt{\frac{1}{N} \sum_{i=1}^N x_i }}$
Kurtosis	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$
Skewness	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$

Kemudian ada pula domain frekuensi yang menunjukkan seberapa banyak sinyal terletak di dalam setiap pita frekuensi yang diberikan pada rentang frekuensi. Sinyal yang berada pada domain waktu harus dikonversi terlebih dahulu menggunakan FFT atau *Fast Fourier Transform*. Berikut adalah kompon pada domain frekuensi [8]:

Mean of band Power Spectrum	$S_\mu = \frac{1}{n} \sum_{i=1}^n S(f)_i$
Max of band power spectrum	$S_{MAX} = \max(S(f)_i)$
Sum of total band power	$S_{SBP} = \sum_{i=1}^n S(f)_i$
Peak of band power	
Variance of band power	$S_V = \frac{\sum_{i=1}^n (S(f)_i - S_\mu)^2}{n-1}$
Standard Deviation of band power	
Skewness of band power	$\hat{S}_S = \frac{1}{n} \frac{\sum_{i=1}^n (S(f)_i - S_\mu)^3}{S_V^{3/2}}$
Kurtosis of band power	$S_K = \frac{1}{n} \frac{\sum_{i=1}^n (S(f)_i - S_\mu)^4}{S_V^{4/2}}$
Relative Spectral Peak per band	$S_{RSPPB} = \frac{\max(S(f)_i)}{\frac{1}{n} \sum_{i=1}^n S(f)_i}$

D. Desain Sistem Pendeteksi Gempa Bumi



Gambar 3.5 Sistem Pendeteksi Gempa Bumi

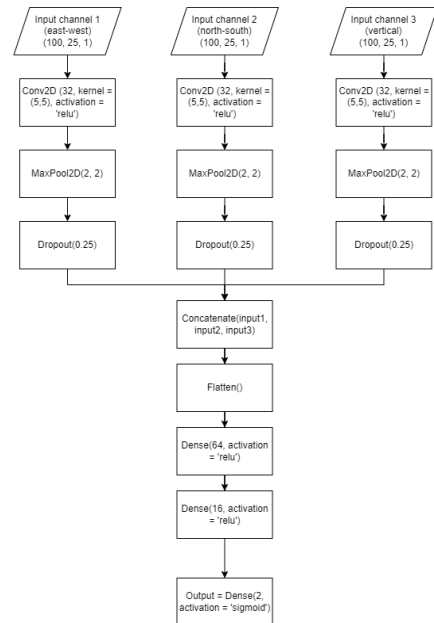
Data gelombang dari seismograf akan diekstraksi fiturnya atau dijadikan spectrogram. Kemudian data yang sudah diekstraksi tersebut dimasukkan ke dalam model klasifikasi. Jika terdeteksi gempa, maka akan dimasukkan ke dalam model regresi untuk dilakukan estimasi besar magnitudonya. Jika tidak maka akan diberikan keluaran yang menyatakan bahwa gelombang tersebut bukanlah gelombang gempa.

E. Convolutional Neural Network

Convolutional neural network (CNN) merupakan salah satu kelas dari *deep learning* yang umumnya diterapkan pada bidang citra. Jaringan saraf dalam, seperti CNN, sudah menunjukkan hasil yang signifikan di berbagai jenis penerapan, termasuk dalam bidang sinyal dan citra [9]. CNN bekerja dengan menggunakan filter yang umumnya disebut sebagai kernel. Dengan menggunakan filter tersebut, CNN dapat membentuk fitur-fitur tak terlihat yang dibentuk dari hasil konvolusi piksel.

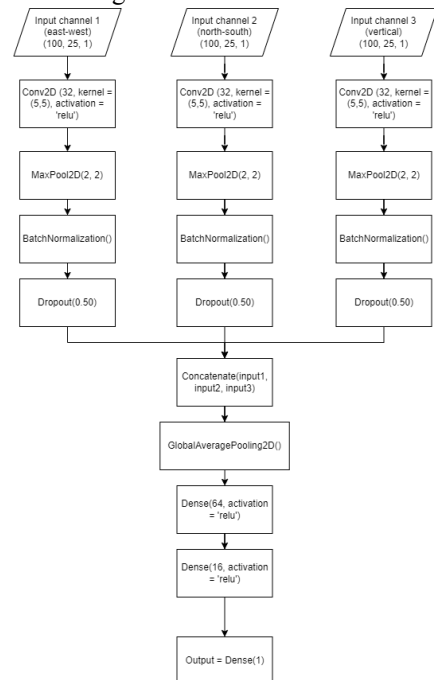
Pada penelitian ini, dibangun sebuah model CNN yang bertujuan untuk mendeteksi gempa beserta dengan estimasi magnitudo dari suatu gempa. Pada model CNN, input merupakan sebuah tensor berdimensi 4 yang merepresentasikan (ukuran input) \times (tinggi input) \times (lebar input) \times (ukuran kanal input). Dikarenakan dataset masih berupa data sinyal *time series*, data dikonversi terlebih dahulu ke dalam bentuk gambar. Data diubah ke dalam bentuk spectrogram yang sudah dicuplik. Spectrogram dipilih dibandingkan dengan *waveform* dikarenakan memiliki informasi yang lebih banyak dibandingkan dengan *waveform*. Terlebih lagi, terdapat perbandingan yang cukup signifikan dari energi yang dibawa oleh sinyal *noise* dengan sinyal gempa, yang dapat mempermudah proses klasifikasi.

Pada proses klasifikasi, model CNN dibangun dengan arsitektur multi input yang menerima tiga kanal spektrogram dengan *output* prediksi sejumlah dua (gempa dan non gempa), dihubungkan dengan *activation function* sigmoid. Gambar 3.6 menunjukkan kerangka arsitektur dari model CNN yang dibentuk.



Gambar 3.6 Arsitektur model CNN klasifikasi gempa.

Model selanjutnya adalah model regresi, untuk mengestimasi magnitudo dari suatu gempa. Seperti halnya dengan model klasifikasi, model ini juga menerima inputan spectrogram dari tiga kanal observasi. Gambar 3.7 menunjukkan kerangka arsitektur dari model CNN regresi.

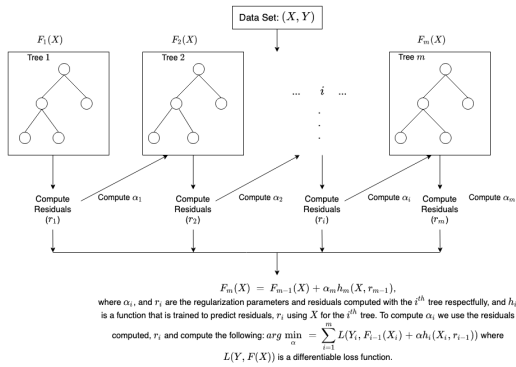


Gambar 3.7 Arsitektur model CNN regresi magnitudo gempa.

F. Extreme Gradient Boosting

XGBoost adalah perpustakaan peningkatan gradien terdistribusi yang dioptimalkan yang dirancang untuk pelatihan model pembelajaran mesin yang efisien dan dapat diskalakan. Merupakan metode pembelajaran ensemble yang menggabungkan prediksi dari beberapa model lemah untuk menghasilkan prediksi yang lebih kuat. XGBoost adalah singkatan dari "Extreme Gradient Boosting" dan telah menjadi salah satu algoritme pembelajaran mesin yang paling populer dan banyak digunakan karena kemampuannya untuk menangani kumpulan data besar dan

kemampuannya untuk mencapai kinerja canggih di banyak tugas pembelajaran mesin seperti sebagai klasifikasi dan regresi [10].



Gambar 3.8 Cara Kerja XGBoost

G. Bayesian Optimization

Bayesian Optimization menyediakan teknik berprinsip berdasarkan Teorema Bayes untuk mengarahkan pencarian masalah optimasi global yang efisien dan efektif. Algoritma optimasi ini bekerja dengan membangun model probabilistik dari fungsi tujuan, yang disebut fungsi pengganti, yang kemudian dicari secara efisien dengan fungsi akuisisi sebelum sampel kandidat dipilih untuk evaluasi pada fungsi tujuan sebenarnya.

Bayesian Optimization sering digunakan dalam pembelajaran mesin terapan untuk menyetel hyperparameter dari model berperforma baik tertentu pada kumpulan data validasi [11].

IV. HASIL DAN ANALISIS

A. Preprocessing

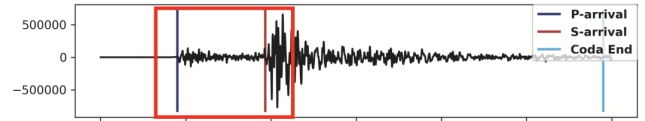
Data yang bersumber dari STEAD memiliki 1 juta baris data dan untuk setiap baris memiliki 3 kanal pengamatan gelombang. Pada penelitian kali ini, penulis melakukan *undersampling* dengan proporsi sebagaimana berikut :

- Gempa dengan magnitudo ≥ 4 sebanyak $\sim 24,000$ baris data
- Gempa dengan magnitudo < 4 sebanyak $\sim 24,000$ baris data
- Noise atau yang bukan gempa sebanyak $\sim 24,000$ baris data

Penentuan sampel tersebut didasari pada skala MMI dan skala richter pada Tabel 1 yang disederhanakan menjadi beberapa kategori sebagaimana yang telah disebutkan di atas. Pada dataset tersebut, gempa yang memiliki magnitudo ≥ 4 hanya memiliki sekitar 24,000 baris data. Maka dari itu saat melakukan *undersampling*, untuk kategori gempa lain juga harus disesuaikan jumlah baris datanya.

Selain melakukan *undersampling*, penulis juga hanya mengambil sebagian gelombang gempa, yaitu pada bagian P-wave saja. Berdasarkan EDA, rata-rata interval antara P-wave dan S-wave adalah 6.7 detik. Namun, pada penelitian ini penulis membulatkannya menjadi 10 detik. **Dengan demikian penulis hanya akan mengambil 1000 nilai gelombang pertama**

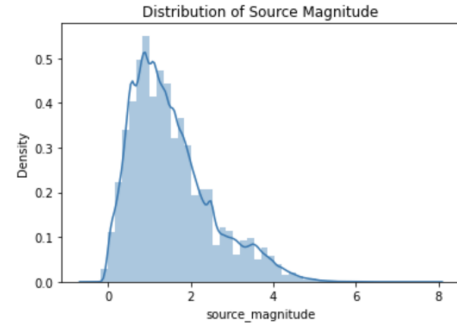
dari P-wave arrival time pada dataset STEAD. Untuk gelombang *noise*, karena tidak memiliki gelombang yang dilabeli sebagai P-wave maka penulis mengambil 1000 nilai gelombang secara acak dan berurutan dari gelombang *noise*.



Gambar 4.1 Cuplikan P-wave pada gelombang gempa

B. Exploratory Data Analysis

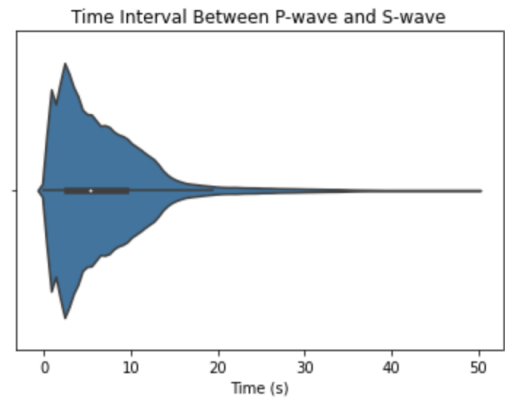
1) Sebaran Magnitudo pada dataset

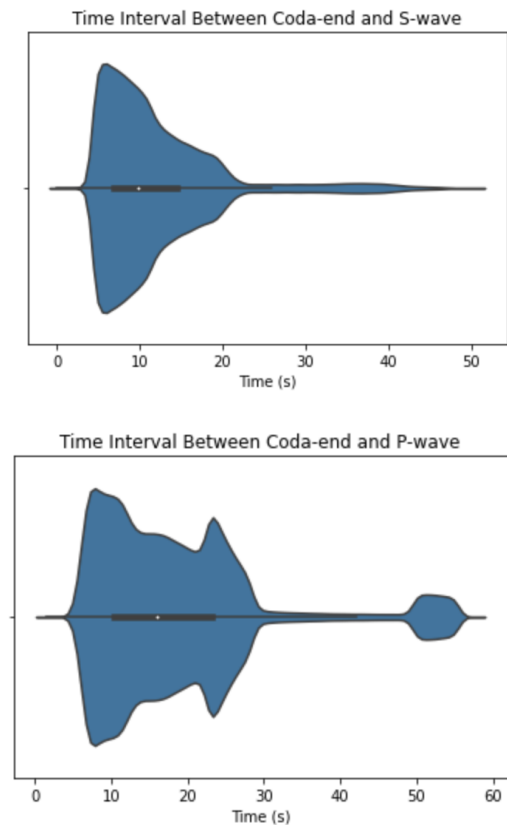


Gambar 4.2 Sebaran Magnitudo

Pada dataset tersebut, gempa yang dicatat lebih banyak gempa yang memiliki magnitudo < 4 . Rata-rata magnitudo pada dataset tersebut adalah 1.52. Hal ini menunjukkan bahwa kebanyakan atau rata-rata **gempa yang terjadi di Dunia memiliki magnitudo yang cenderung kecil dan tidak membahayakan.**

2) Sebaran Waktu Interval antar Jenis Gelombang





Gambar 4.3 Sebaran waktu interval antar jenis gelombang

Gambar di atas menunjukkan waktu interval antar berbagai jenis gelombang (P-wave, S-wave, dan akhir gelombang S-wave). **Didapat rata-rata interval P-wave dengan S-wave adalah 6.7 detik, rata-rata interval S-wave sampai akhir gelombang S-wave adalah 11.8 detik, dan rata-rata interval P-wave sampai akhir gelombang S-wave adalah 18.6 detik.** Berdasarkan data di atas, 18.6 detik adalah waktu rata-rata antara P-wave pertama sampai S-wave habis dan waktu tersebutlah waktu yang krusial untuk melakukan deteksi terhadap gempa, karena setelah waktu tersebut akan muncul Surface wave yaitu gelombang gempa sungguhan. Dengan demikian, **Sistem yang dikembangkan harus mampu mendeteksi gempa sebelum 18.6 detik setelah P-wave pertama terdeteksi.**

C. Feature Extraction and Engineering

Pada model XGBoost, **gelombang pada 3 kanal di ekstrak** masing-masing untuk mendapatkan nilai-nilai sebagaimana berikut : MIN, MAX, MEAN, RMS, VAR, STD, POWER, PEAK, P2P, CREST, FACTOR, SKEW, KURT OSIS, MAX_f, SUM_f, MEAN_f, VAR_f, PEAK_f, SKEW_f, KURTOSIS_f. Kemudian dilakukan **merge pada kolom untuk setiap kanal menjadi satu set data** yang akan dilakukan pemodelan.

Berbeda dengan model XGBoost, **Gelombang gempa akan dijadikan gambar spectrogram terlebih dahulu sebelum dimasukkan ke dalam model.** Model CNN dapat mengekstrak fitur dari gambar tersebut dan mencari pola yang tidak terlihat secara langsung. Fitur-fitur ini diekstraksi dengan bantuan beberapa layer pada CNN

seperti Conv2D, pooling layer, dan fully-connected layer (FC Layer).

D. Modelling

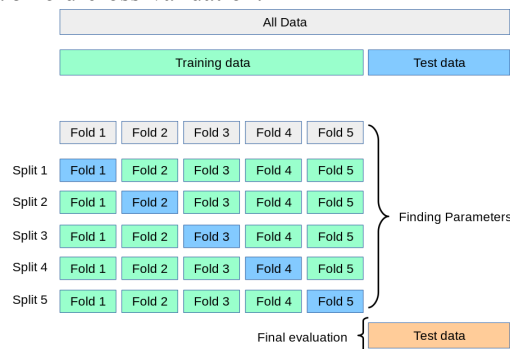
Pemodelan akan dilakukan dua kali yaitu klasifikasi dengan *metrics* F1 score dan regresi dengan *metrics* RMSE. Sebelum dimodelkan, dilakukan pembagian dataset menjadi data latih dan data uji dengan proporsi $\frac{1}{3}$.

Pada pemodelan XGBoost, untuk menentukan parameter terbaik, dilakukanlah *hyperparameter tuning* menggunakan bayesian optimization. Bayesian Optimization akan mengoptimasi parameter pada XGBoost untuk mendapatkan performa yang lebih baik.

Arsitektur model CNN dapat dilihat pada Gambar 3.6 dan Gambar 3.7. Pada masing-masing model, terdapat *dropout layer* yang merupakan salah satu teknik regularisasi yang bertujuan untuk menghilangkan beberapa koneksi antara satu *layer* ke *layer* yang lain dengan tujuan agar model tidak *overfit* terhadap *training dataset*. Ketiga kanal inputan di-*reshape* terlebih dahulu ke dalam bentuk tensor berdimensi empat, yang akan diproses secara individual untuk kemudian digabung pada *layer concatenate*. Untuk input sendiri merupakan gambar spectrogram yang telah diambil 1000 nilai dari saat munculnya P-wave. Oleh karena itu, ukuran dari masing-masing spectrogram adalah sebesar 100×25 piksel.

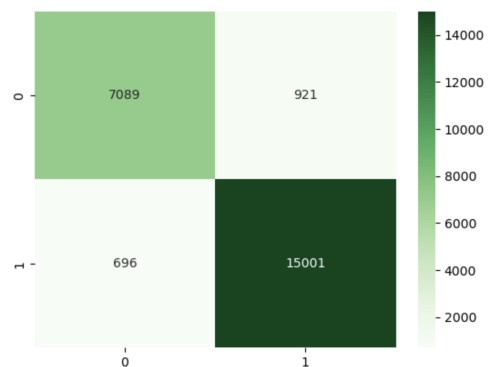
E. Validation

Model XGBoost divalidasi menggunakan *metrics* F1 Score untuk klasifikasi dan RMSE untuk regresi dengan skema 5-fold-cross-validation.



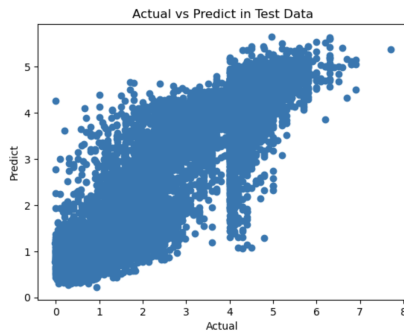
Gambar 4.4 Skema 5-fold Cross-Validation

Didapat skor F1 adalah 0.951 dan skor RMSE adalah 0.368. Setelah melakukan validasi, kemudian dilakukan pengujian pada data uji. Untuk klasifikasi adalah sebagai berikut :



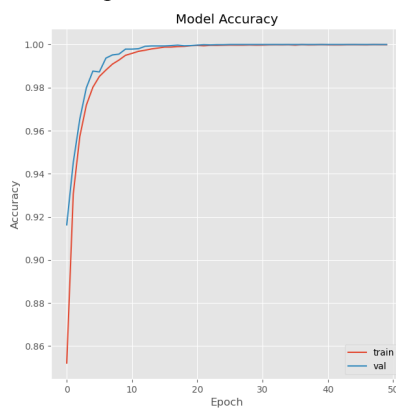
Gambar 4.5 Plot *confusion matrix* hasil klasifikasi model XGBoost terhadap *testing set*.

Kemudian untuk regresi, hasilnya adalah sebagai berikut :

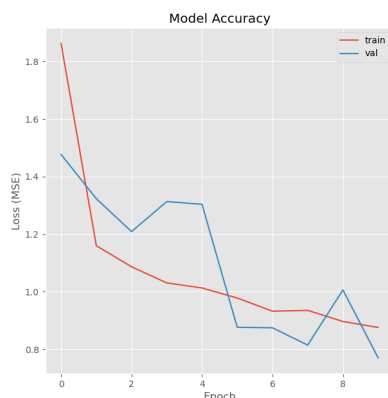


Gambar 4.6 Plot *confusion matrix* hasil regresi model XGBoost terhadap *test set*.

Sedangkan pada model CNN, proses *training* dilakukan pada *platform* Kaggle Notebook dengan menggunakan *Accelerator* GPU Nvidia T4 2x. Hasil proses *training* dan *validation* dari kedua model klasifikasi dan regresi dapat dilihat pada Gambar 4.7.



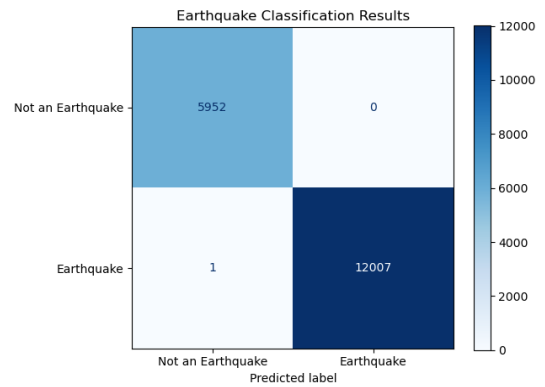
(a)



(b)

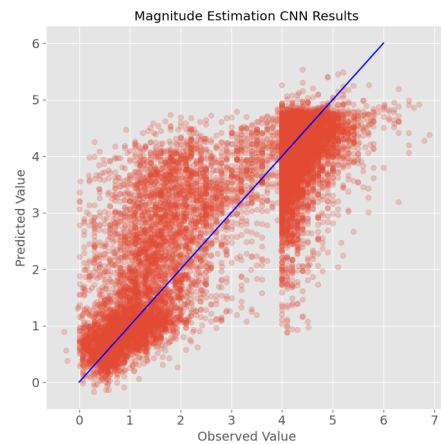
Gambar 4.7 Plot hasil proses *training* dan *validation* model CNN. (a) Model klasifikasi gempa, (b) Model estimasi magnitudo.

Model klasifikasi CNN di-*train* dengan jumlah *epoch* sebanyak 50, dan menghasilkan akurasi sebesar 0.99 pada *validation set* dengan **F1 score** sebesar **0.99**. Gambar 4.8 menunjukkan hasil plot *confusion matrix* prediksi model terhadap dataset *testing*.



Gambar 4.8 Plot *confusion matrix* hasil klasifikasi model CNN terhadap *testing set*.

Sedangkan pada model regresi CNN, model di-*train* dengan jumlah *epoch* sebesar 10, dan menghasilkan rerata *root mean squared error* (RMSE) sebesar **0.83**. Hasil *scatter plot* dari nilai terprediksi dengan *ground truth* terdapat pada Gambar 4.9.



Gambar 4.9 Plot *scatter* perbandingan hasil *ground truth* magnitudo dengan hasil prediksi pada model regresi CNN.

Perbandingan hasil kedua model XGBoost dan CNN dapat dilihat pada Tabel 2.

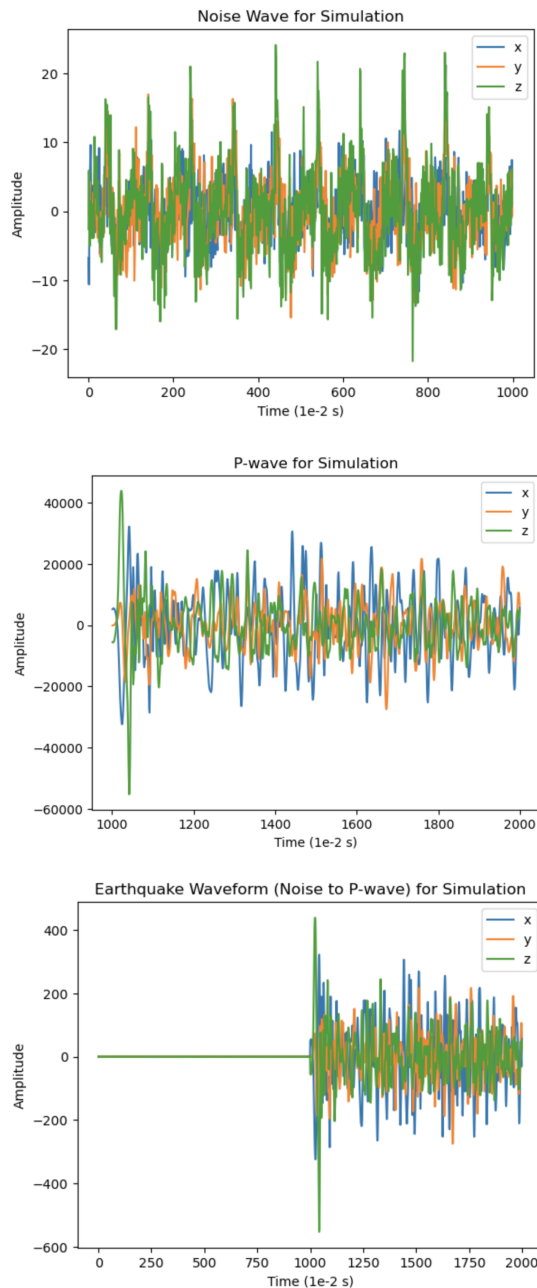
TABEL 2. TABEL PERBANDINGAN METRIK HASIL MODEL CNN DAN XGBOOST.

Model	Jenis	F1 Score	RMSE
CNN	Klasifikasi Gempa	0.999	-
	Regresi Estimasi Magnitudo	-	0.83
XGBoost	Klasifikasi Gempa	0.951	-
	Regresi Estimasi Magnitudo	-	0.368

F. Simulation

Setelah membuat model, penulis melakukan simulasi menggunakan model yang telah dibuat untuk kasus *real-time*. Tujuan dari simulasi ini adalah untuk melihat waktu *running* rata-rata dari sistem yang telah dibuat dan mengetahui berapa waktu yang dibutuhkan untuk mendeteksi gempa sejak P-wave pertama kali tercatat pada seismograf. Simulasi dilakukan pada Kaggle Notebook dengan menggunakan *Accelerator* GPU T4 2x.

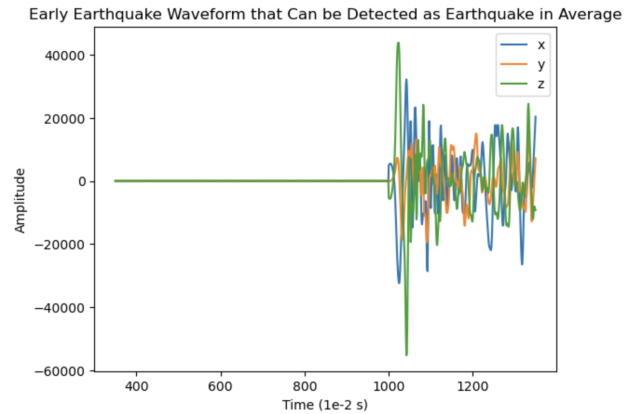
Untuk melihat waktu running rata-rata, penulis menggunakan 1000 sampel data dan didapat **running time untuk setiap waveformnya adalah sekitar ~0.03 detik**. Kemudian untuk *Waveform* yang digunakan pada simulasi kedua yaitu untuk mengetahui berapa waktu yang dibutuhkan untuk mendeteksi gempa adalah **penggabungan antara noise dan P-wave gempa magnitude > 4**.



Gambar 4.10 Penggabungan waveform untuk simulasi

Gelombang tersebut akan berjalan dan ditangkap pada sistem untuk setiap 0.01 detiknya. Sistem akan menangkap gelombang untuk setiap iterasinya dengan formula $i:1000+i$, i adalah angka dari 0 sampai 1000. Dengan demikian, sistem akan membaca gelombang dari index 0:1000 sampai 1000:2000. Simulasi ini dilakukan pada 100 gelombang yang berbeda. Setelah simulasi dilakukan, didapat bahwa **sistem membutuhkan waktu rata-rata 3.5 detik untuk mendeteksi gempa magnitude ≥ 4 setelah P-wave pertama**. Karena **terdapat delay antara data dan model (0.01 detik dibanding 0.03)** selama 0.02 detik, maka sistem akan

benar-benar dapat mendeteksi gempa setelah 7 detik dari P-wave pertama.



Gambar 4.11 Bentuk gelombang minimal yang dapat dideteksi sebagai gempa

V. KESIMPULAN DAN SARAN

1) Sinyal yang tertangkap pada saat terjadinya gempa dapat dibagi menjadi tiga: P-wave, S-wave serta *surface wave*. **Pemodelan sistem *early warning* bagi gempa dapat dilakukan dengan mendeteksi adanya P-wave yang terdeteksi oleh sensor terlebih dahulu.**

2) Proses pemilihan data gempa dapat disederhanakan menjadi tiga kategori: *noise*, magnitudo < 4.0 serta magnitudo ≥ 4.0 . **Gempa dengan magnitudo ≥ 4.0 memiliki dampak yang dapat dirasakan secara umum**, sehingga pemodelan gempa yang didasarkan pada data ini akan lebih membantu.

3) Dari hasil analisis, didapat bahwa waktu **rata-rata jarak antara tibanya P-wave dengan S-wave adalah sekitar 6.7 detik**. Untuk memaksimalkan antara performa model dalam memprediksi tibanya gempa tanpa harus mengorbankan ukuran dari dataset *training* (yang dapat menyebabkan *underfitting*), **dipilih jumlah index sinyal cuplikan sebesar 1000 index (10 detik)**.

4) Dari hasil prediksi, diperoleh bahwa **model klasifikasi CNN mampu menghasilkan skor F1 yang lebih tinggi** dibandingkan dengan XGBoost dengan perbandingan **0.99 dibanding 0.95**. Sebaliknya, model XGBoost lebih baik pada kasus regresi untuk mengestimasi magnitudo dari gempa yang akan terjadi dengan perbandingan **0.368 dan 0.83**. Oleh karena itu, **penerapan metode *ensembling* pada kedua model klasifikasi dan regresi, untuk menghasilkan hasil yang terbaik**.

5) Hasil simulasi menunjukkan bahwa model dapat memprediksi terjadinya gempa setelah **7 detik dari saat P-wave pertama ditangkap**. **Sistem dianggap baik karena mampu mendeteksi gempa < 18.6 detik yang merupakan rata-rata waktu interval antara P-wave pertama sampai akhir dari S-wave**.

REFERENCES

- [1] W. H. K. Lee, International Association of Seismology and Physics of the Earth's Interior, and International Association for Earthquake Engineering, Eds., International handbook of earthquake and engineering seismology. Amsterdam ; Boston: Academic Press, 2002.
- [2] D. Zhang, J. Fu, Z. Li, L. Wang, J. Li, and J. Wang, "A Synchronous Magnitude Estimation with P-Wave Phases' Detection Used in

- Earthquake Early Warning System,” *Sensors*, vol. 22, no. 12, p. 4534, Jun. 2022, doi: 10.3390/s22124534.
- [3] R. A. Sugondo and C. Machbub, “P-Wave detection using deep learning in time and frequency domain for imbalanced dataset,” *Heliyon*, vol. 7, no. 12, p. e08605, Dec. 2021, doi: 10.1016/j.heliyon.2021.e08605.
 - [4] A. Geron, “Hands-On Machine Learning with Scikit-Learn & Tensorflow,” O’Reilly, 2017.
 - [5] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, “Stanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI,” *IEEE Access*, vol. 7, pp. 179464–179476, 2019, doi: 10.1109/ACCESS.2019.2947848.
 - [6] E. Sejdic, I. Djurovic, and Lj. Stankovic, “Quantitative Performance Analysis of Scalogram as Instantaneous Frequency Estimator,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3837–3845, Aug. 2008, doi: 10.1109/TSP.2008.924856.
 - [7] “Earthquake Facts and Statistics,” *GNS Science*, 29-Nov-2012. [Online]. Available: <https://www.gns.cri.nz/our-science/natural-hazards-and-risks/earthquakes/earthquake-monitoring/>. [Accessed: 18-Feb-2023].
 - [8] “How to extract Features from Signals,” *Medium*, 16-Feb-2021. [Online]. Available: <https://m-gambara.medium.com/how-to-extract-features-from-signals-15e7db225c15>. [Accessed: 18-Feb-2023].
 - [9] M. V. Valueva, N. N. Nagornov, P. A. Lyakhov, G. V. Valuev, and N. I. Chervyakov, “Application of the residue number system to reduce hardware costs of the convolutional neural network implementation,” *Mathematics and Computers in Simulation*, vol. 177, pp. 232–243, Nov. 2020, doi: 10.1016/j.matcom.2020.04.031.
 - [10] “XGBoost,” *Geeksforgeeks* 06-Feb-2023. [Online]. Available: <https://www.geeksforgeeks.org/xgboost/>. [Accessed: 18-Feb-2023].
 - [11] “How to Implement Bayesian Optimization from Scratch in Python,” *Geeksforgeeks* 09-Oct-2019. [Online]. Available: <https://machinelearningmastery.com/what-is-bayesian-optimization/>. [Accessed: 18-Feb-2023].