

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

Samatha Marhaendra Putra
Rizky Alif Ramadhan
Daffa Bil Nadzary

TheLastDance

Poin Bahasan

01

Latar Belakang dan Tujuan

Gempa, gelombang, dan tujuan penelitian

03

Hasil dan Analisis

Preprocessing, EDA, validation

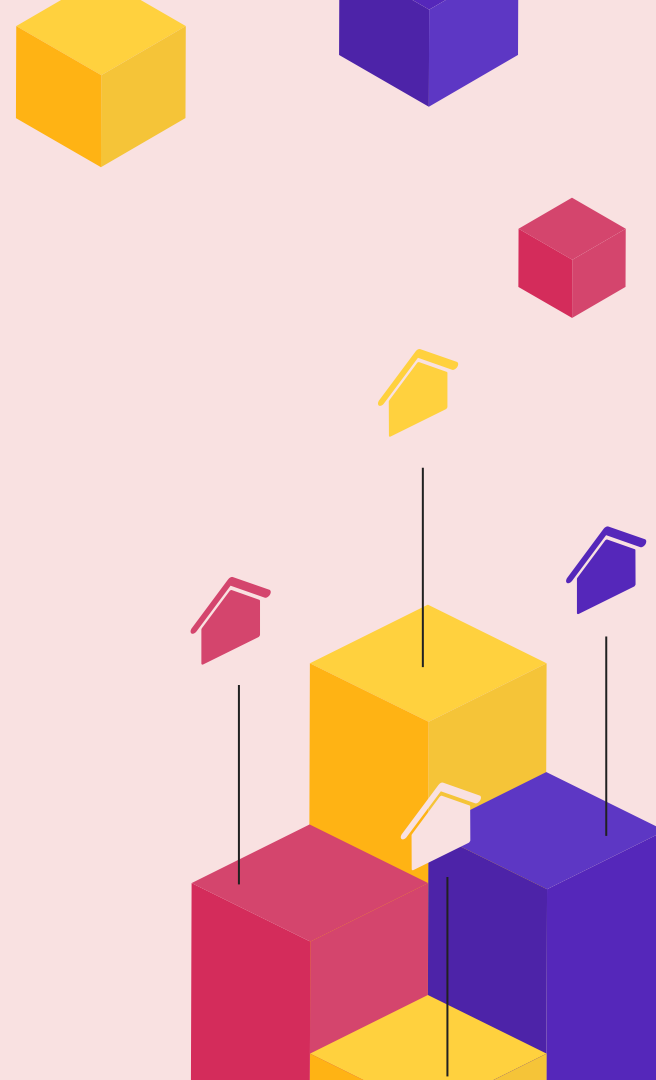
02

Metode

Dataset, skema sistem, *feature extraction*, model

04

Kesimpulan dan Saran



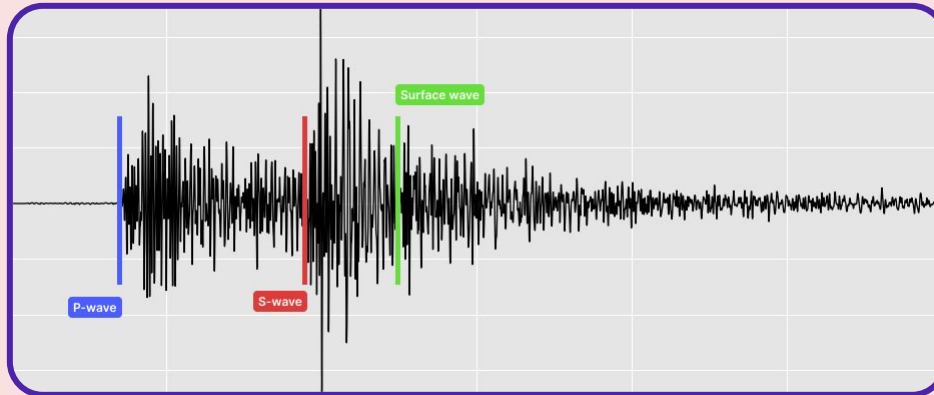


01

Latar Belakang dan Tujuan

Latar Belakang

- Gempa merupakan fenomena pergeseran lempeng bumi yang mengakibatkan terjadinya pelepasan energi secara spontan dan menghasilkan **gelombang seismik**
- Sinyal yang dapat ditangkap pada saat terjadinya gelombang seismik terdiri dari: **Primary Wave (P-wave)**, **Secondary Wave (S-wave)**, dan **Surface Wave**.



Contoh pengamatan 1 sinyal gempa selama 60 detik



Tujuan

Deteksi Dini



Membuat model klasifikasi dan regresi berbasis *machine learning* dan *deep learning* untuk melakukan deteksi dini secara *real time*.

Komparasi Performa

Melakukan komparasi performa antara model klasifikasi dan regresi dalam mendeteksi dini terjadinya gempa

Simulasi

Melakukan simulasi sistem deteksi gempa dini menggunakan model yang dibuat



02 Metode



Dataset

Journals & Magazines > IEEE Access > Volume: 7

Stanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI

Publisher: IEEE

Cite This

PDF

S. Mostafa Mousavi ; Yixiao Sheng ; Weiqiang Zhu ; Gregory C. Beroza [All Authors](#)

99

Paper

Citations

8654

Full

Text Views



Open Access



Comment(s)

Under a Creative Commons License

Abstract

Document Sections

I. Introduction

II. Properties of the Data Set

Abstract:

Seismology is a data rich and data-driven science. Application of machine learning for gaining new insights from seismic data is a rapidly evolving sub-field of seismology. The availability of a large amount of seismic data and computational resources, together with the development of advanced techniques can foster more robust models and algorithms to process and analyze seismic signals. Known examples or labeled data sets, are the essential requisite for building supervised models. Seismology has labeled data, but the reliability of those labels is highly variable, and the lack of high-quality labeled data sets to serve as

Stanford EArthquake Dataset (STEAD)

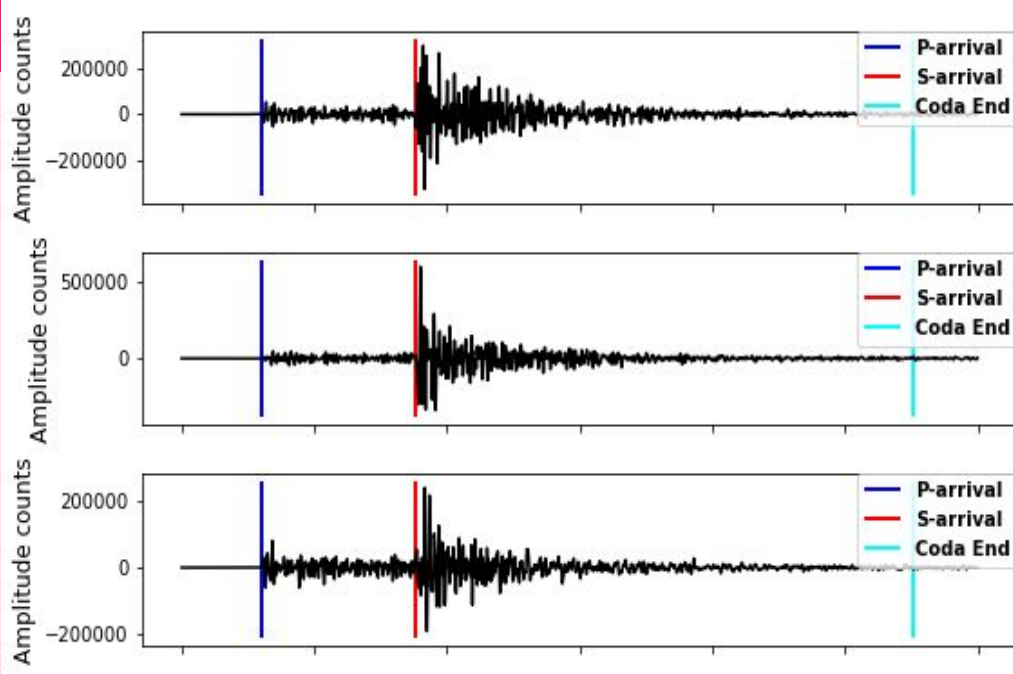
Dataset

1.200.000
Data Pengamatan

960.000
Gempa

240.000
Noise

Dataset



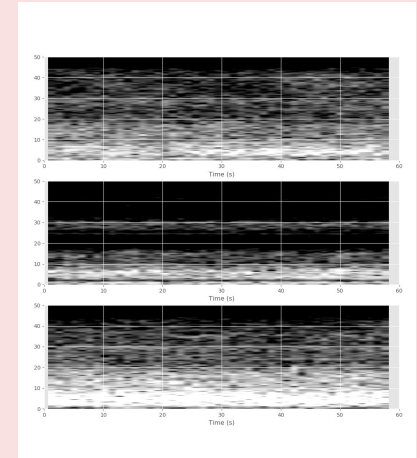
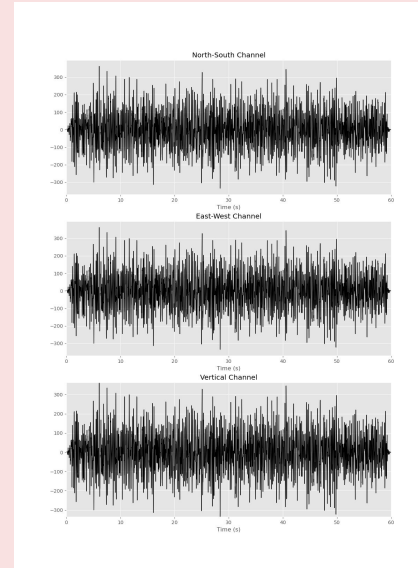
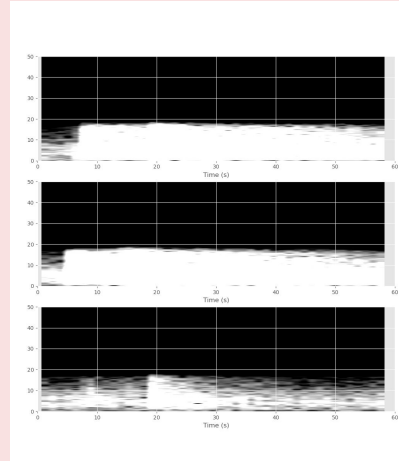
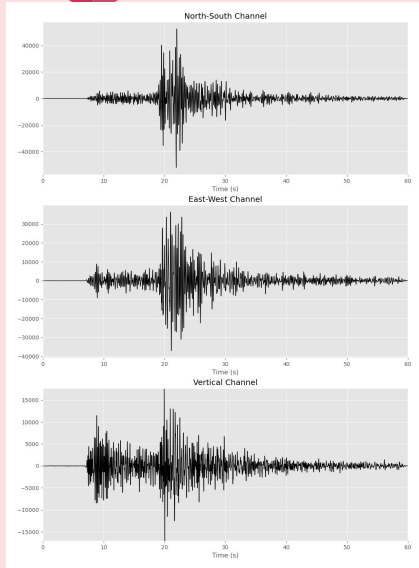
Gelombang Gempa

```
back_azimuth_deg 280.5
coda_end_sample 3114.0
network_code TA
p_arrival_sample 500.0
p_status manual
p_travel_sec 17.26
p_weight 0.5
receiver_code 109C
receiver_elevation_m 150.0
receiver_latitude 32.8889
receiver_longitude -117.1051
receiver_type BH
s_arrival_sample 1678.0
s_status manual
s_weight 0.5
snr_db [ 37.20000076 42. 38.59999847]
source_depth_km 12.66
source_depth_uncertainty_km None
source_distance_deg 0.92
source_distance_km 101.87
source_error_sec 0.8127
source_gap_deg 48.096
source_horizontal_uncertainty_km 3.0397
source_id 8940123
source_latitude 32.7253
source_longitude -116.0348
source_magnitude 3.6
source_magnitude_author None
source_magnitude_type ml
source_mechanism_strike_dip_rake None
source_origin_time 2006-11-03 16:12:12.44
source_origin_uncertainty_sec 0.27
trace_category earthquake_local
trace_name 109C.TA_20061103161223_EV
trace_start_time 2006-11-03 16:12:24.700000
```

Metadata

Dataset

2 kategori pengamatan: Gempa dan Noise
(Plot *Waveform* dan *Spectrogram*)



Gelombang Gempa

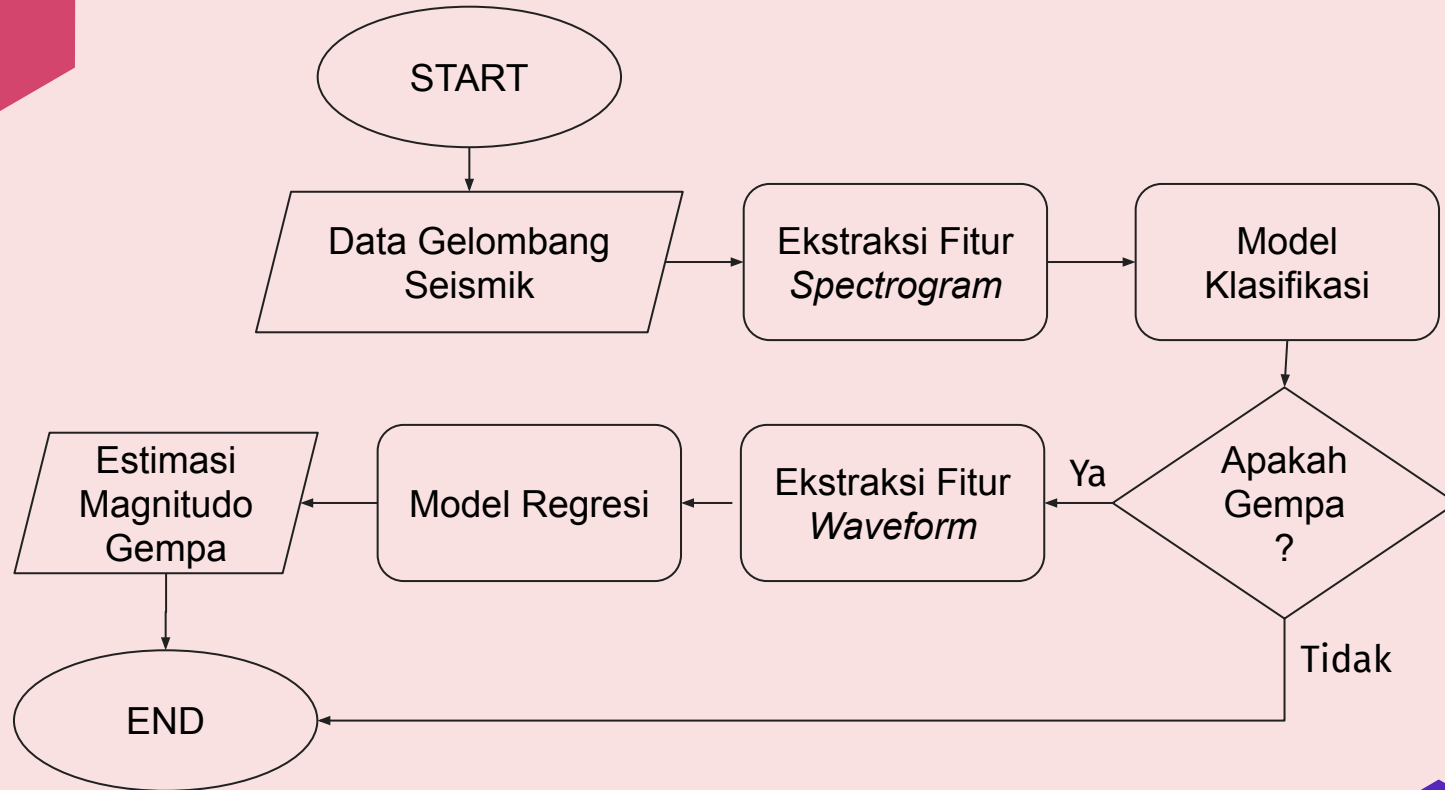
Noise

Skala Gempa

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

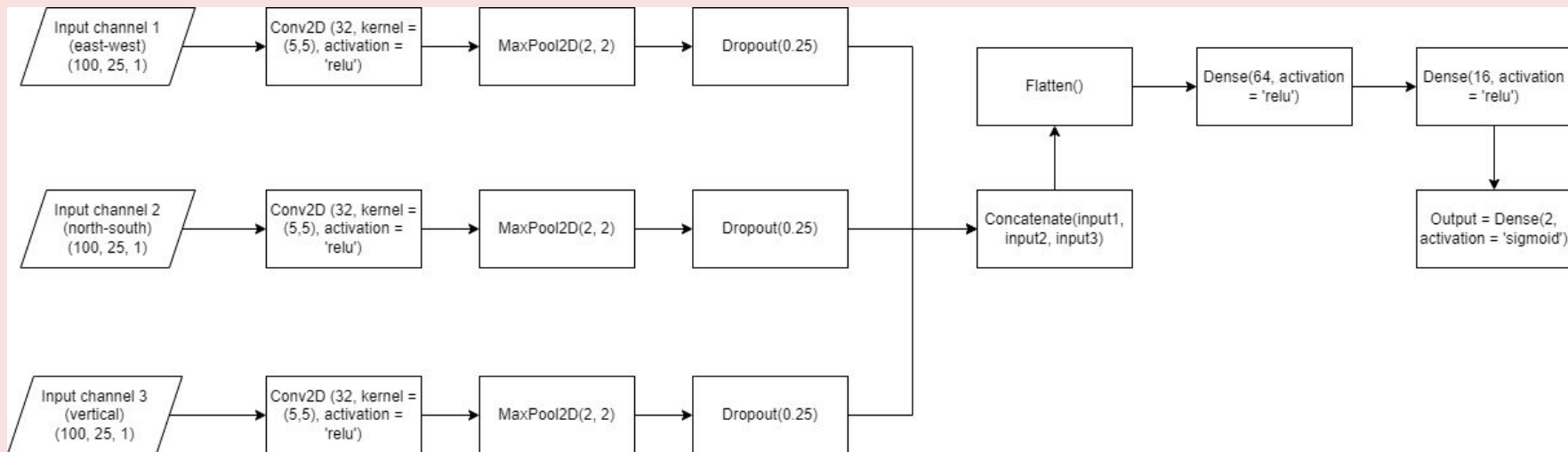
Tabel Skala MMI beserta dengan Skala Richter (SR) dari Sebuah Gempa

Skema Sistem



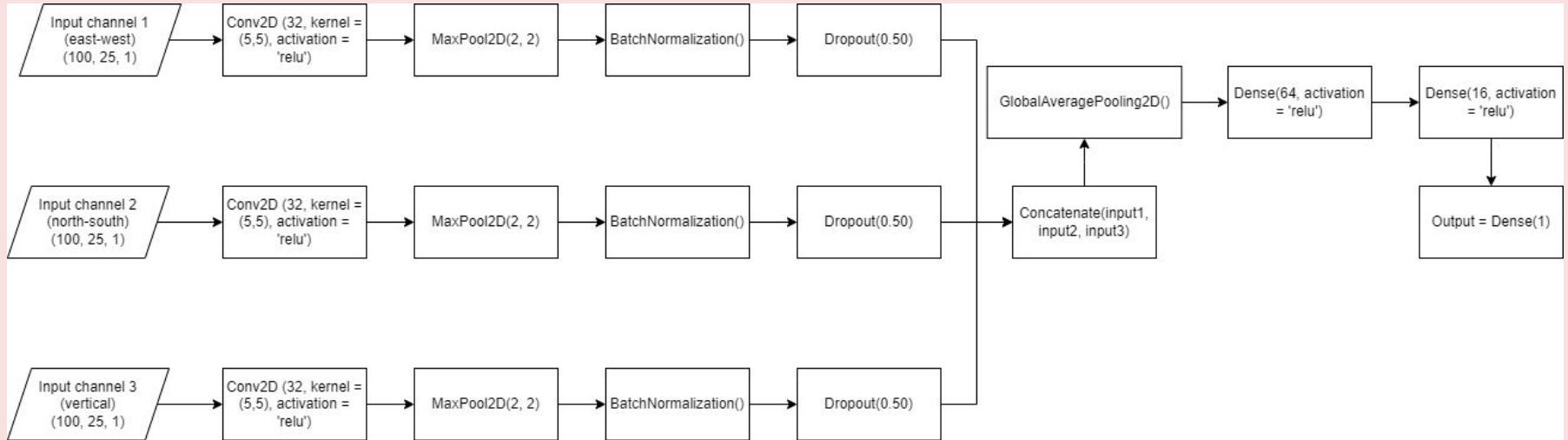
Skema Deteksi Dini Gempa

Model Klasifikasi: CNN



Arsitektur Model CNN (Klasifikasi)

Model Regresi: CNN



Arsitektur Model CNN (Regresi)

Waveform Feature Extraction

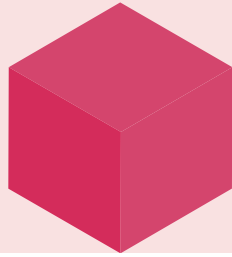
Mean	Max	Peak	Peak to Peak
$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$	$\max(x_i)$	$P_m = \max(x_i)$	$P_k = \max(x_i) - \min(x_i)$

RMS	Variance	Standard Deviation	Power
$\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	$\sum_{i=1}^N (x_i - \bar{x})^2 / N - 1$	$\sigma = \sqrt{\sum_{i=1}^n (x_i - \mu)^2 / n - 1}$	$\frac{1}{N} \sum_{i=1}^N x_i^2$

Crest Factor	Form Factor	Pulse Indicator	Margin	Kurtosis	Skewness
$\frac{P_m}{RMS}$	$\frac{RMS}{\bar{x}}$	$\frac{P_m}{\bar{x}}$	$\frac{P_m}{ \frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } ^2}$	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$	$\frac{N \sum (x_i - \bar{x})^3}{(N-1)(N-2)\sigma^3}$

Domain Waktu

Waveform Feature Extraction



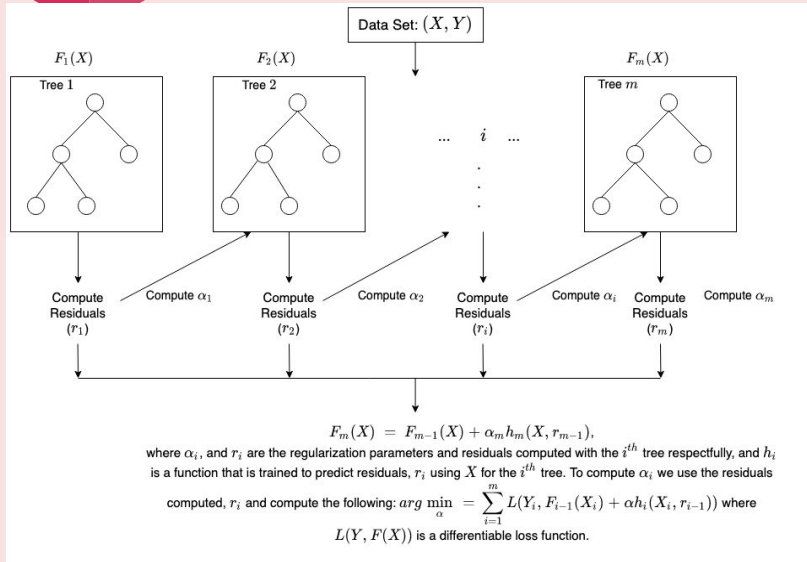
Mean of Band Power Spectrum	Max of Band Power Spectrum	Sum of total Band Power	Variance of Band Power
$S_{\mu} = \frac{1}{n} \sum_{i=1}^n S(f)_i$	$S_{MAX} = \max (S(f)_i)$	$S_{SBP} = \sum_{i=1}^n S(f)_i$	$S_V = \frac{\sum_{i=1}^n (S(f)_i - S_{\mu})^2}{n - 1}$

Skewness of Band Power	Kurtosis of Band Power	Relative Spectral Peak per Band
$S_s = \frac{1}{n} \frac{\sum_{i=1}^n (S(f)_i - S_{\mu})^3}{S_V^{3/2}}$	$S_K = \frac{1}{n} \frac{\sum_{i=1}^n (S(f)_i - S_{\mu})^4}{S_V^{4/2}}$	$S_{RSPPB} = \frac{\max (S(f)_i)}{\frac{1}{n} \sum_{i=1}^n S(f)_i}$

Domain Frekuensi



Model Regresi + Klasifikasi: XGBoost + Bayesian Opt



Cara kerja XGBoost

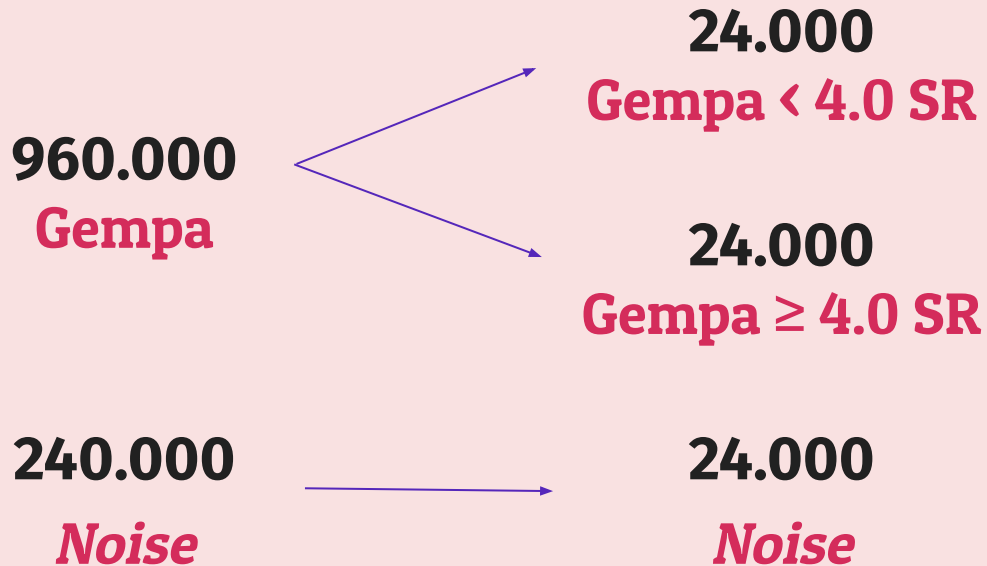
- **XGBoost** adalah **library** peningkatan gradien terdistribusi yang teroptimisasi dengan tujuan untuk pelatihan model *machine learning* yang lebih efisien dan bersifat **scalable**.
- XGBoost banyak digunakan dikarenakan kemampuannya untuk **menangani kumpulan data besar** dengan kinerja yang baik.
- **Bayesian Optimization** sering digunakan dalam *applied machine learning* untuk menyetel **hyperparameter** dari model berperforma baik tertentu pada kumpulan data validasi.

The background features several 3D cubes in purple, yellow, and pink, scattered across the slide. A large purple cube is in the top left, a small yellow cube is in the top right, a medium purple cube is on the right side, a small yellow cube is in the bottom right, a medium pink cube is in the bottom left, and a small yellow cube is on the far left edge. A pink cube is also partially visible in the top right corner.

03

Hasil dan Analisis

Preprocessing: Undersample

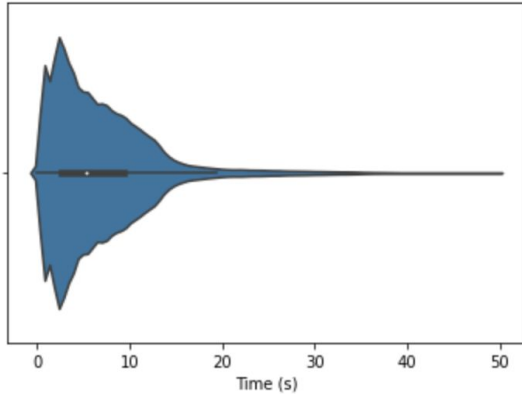


Penentuan nilai *threshold* magnitudo untuk *sampling* didasarkan pada skala MMI.

Exploratory Data Analysis

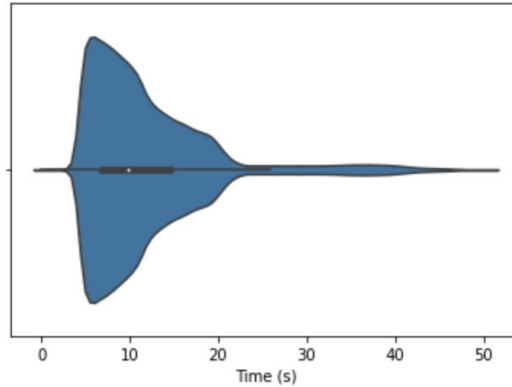
Sebaran waktu interval antar jenis gelombang

Time Interval Between P-wave and S-wave



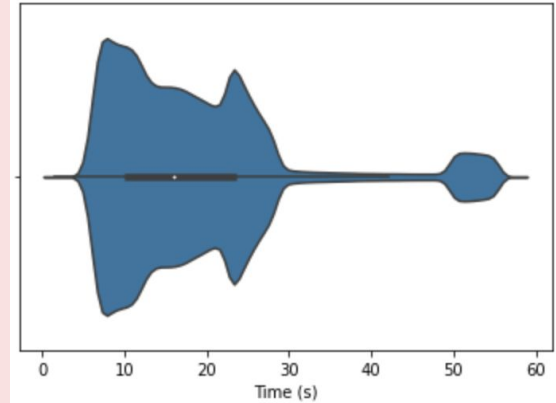
**Interval antara P-wave
dengan S-wave
(rata-rata: 6.7 detik)**

Time Interval Between Coda-end and S-wave



**Interval antara awal S-wave
hingga akhir gelombang
(rata-rata: 11.8 detik)**

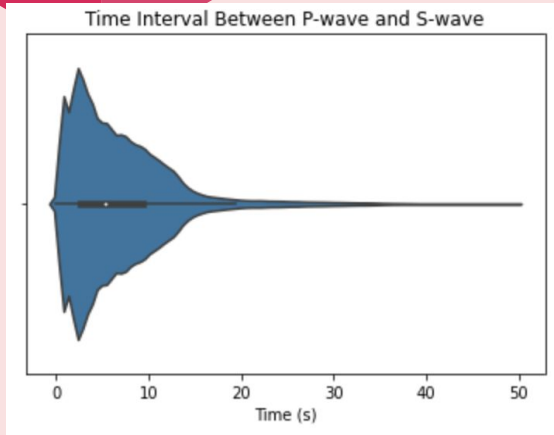
Time Interval Between Coda-end and P-wave



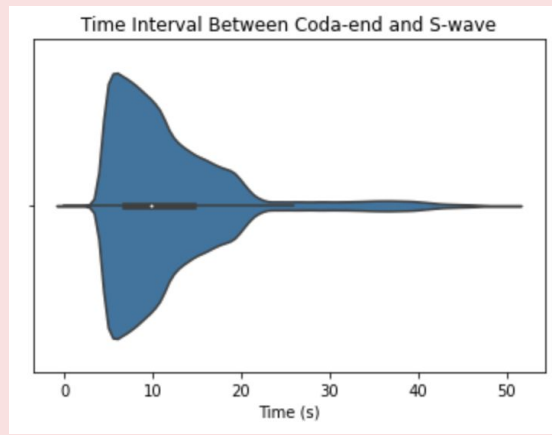
**Interval antara P-wave hingga
akhir gelombang
(rata-rata: 18.6 detik)**

Exploratory Data Analysis

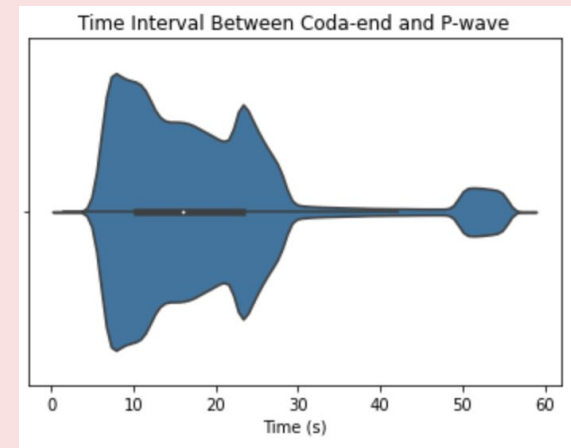
Sebaran waktu interval antar jenis gelombang



**Interval antara P-wave
dengan S-wave
(rata-rata: 6.7 detik)**



**Interval antara awal S-wave
hingga akhir gelombang
(rata-rata: 11.8 detik)**

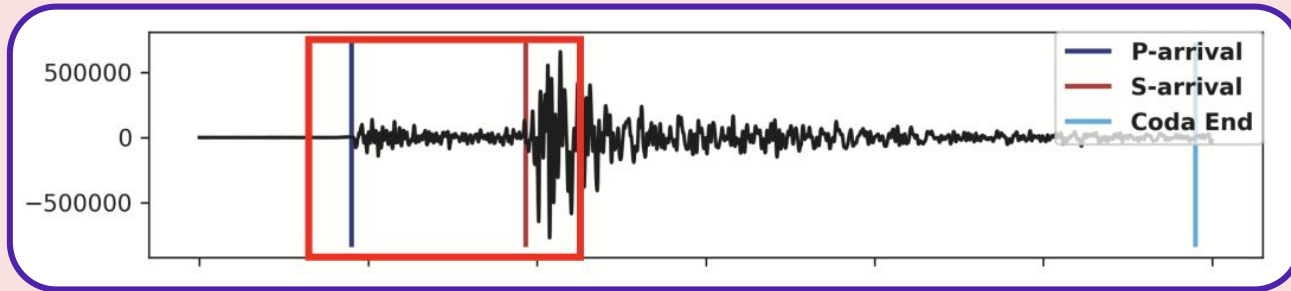


**Interval antara P-wave hingga
akhir gelombang
(rata-rata: 18.6 detik)**

Sistem yang dikembangkan harus mampu mendeteksi gempa 18.6 detik setelah P-wave pertama

Preprocessing (cont.): Capturing Sample



- Memfokuskan pengolahan data hanya pada **bagian P-wave saja**
- Diambil data dalam interval waktu sejak **P-wave arrival time** hingga **10 detik setelahnya**.



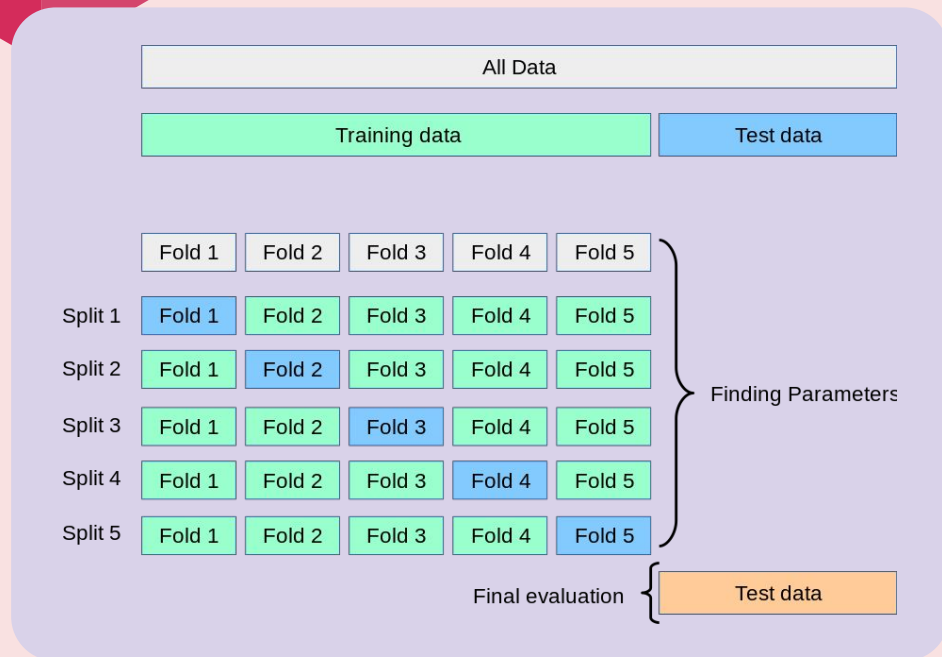
Contoh cuplikan P-wave pada gelombang gempa



Feature Engineering

- Pada pemodelan menggunakan **CNN**, **data gelombang gempa akan diubah menjadi plot *spectrogram***. Selanjutnya, dengan memasukkan sekumpulan gambar tersebut, proses ekstraksi fitur akan dilakukan oleh CNN dengan sendirinya.
 - Pada pemodelan menggunakan **XGBoost**, setiap kanal gelombang akan dilakukan **ekstraksi fitur** menggunakan sekumpulan rumus perhitungan yang telah ditunjukkan sebelumnya. Lalu, dilakukan **merge kolom** sehingga menjadi satu set data baru yang kemudian dipakai di tahap pemodelan.
- 
- 

Validation: XGBoost



5-fold Cross Validation

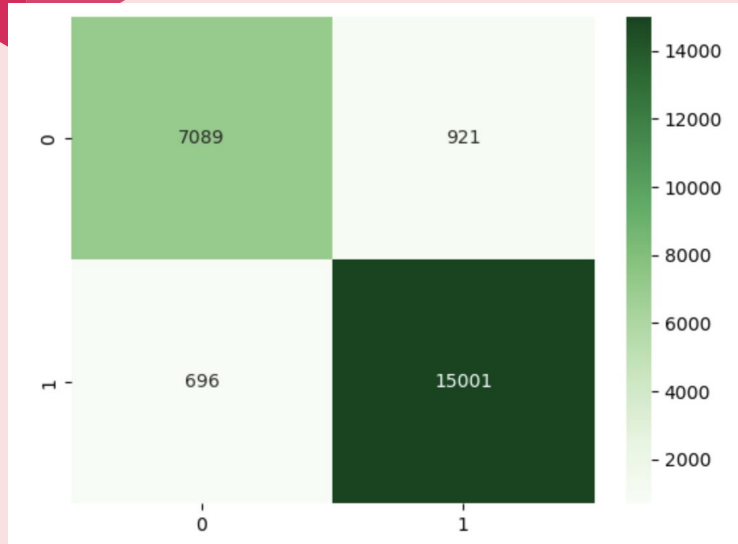
F1 Score
Model klasifikasi:

0.951

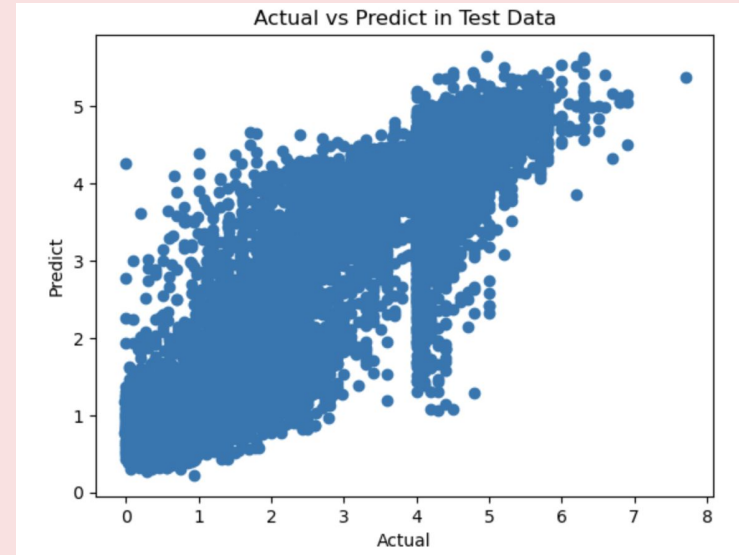
Skor RMSE
Model Regresi

0.368

Validation: XGBoost

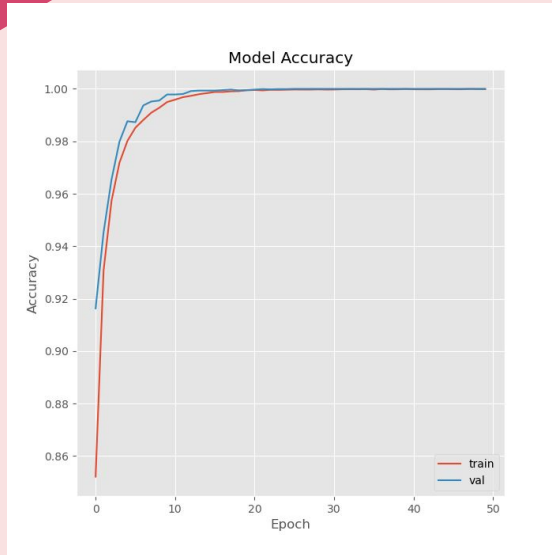


Confusion Matrix
Model Klasifikasi

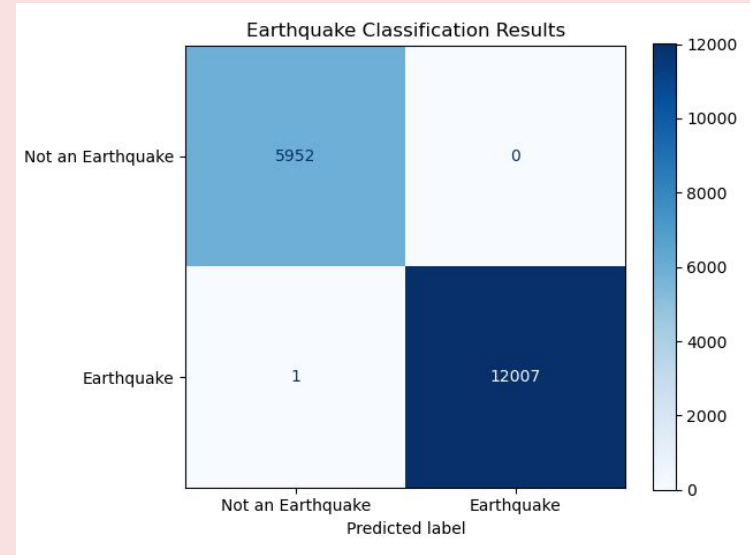


Prediction vs Actual Plot
Model Regresi

Validation: Model Klasifikasi CNN



Model Performance

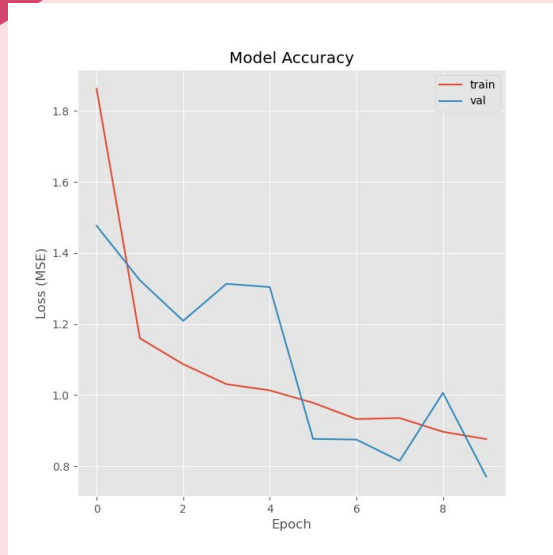


Confusion Matrix

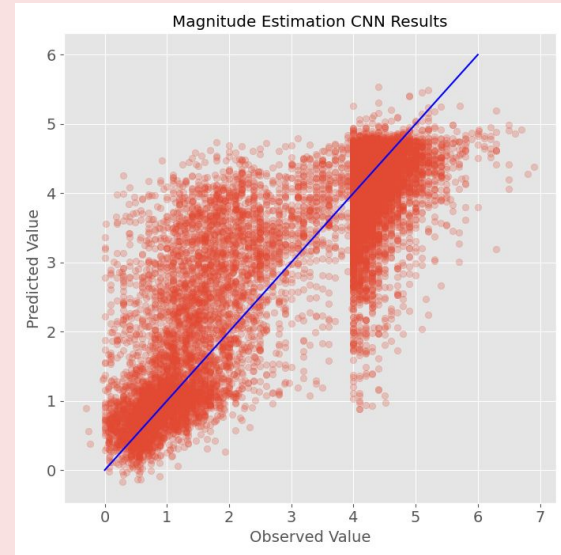
F1 Score
Model klasifikasi:

0.999

Validation: Model Regresi CNN



Model Performance



**Prediction vs Actual
Plot Results**

Skor RMSE
Model Regresi:

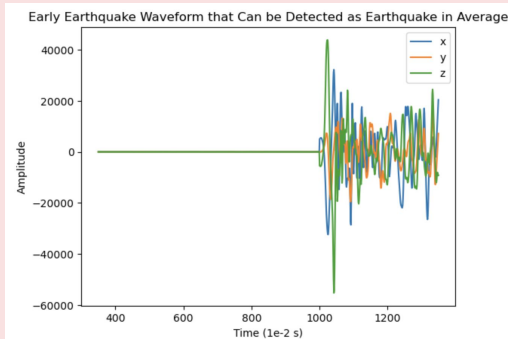
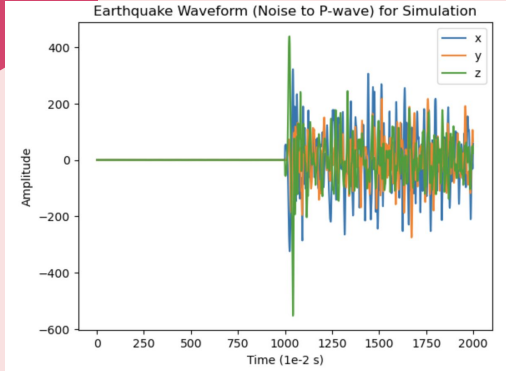
0.83

Validation: Tabel Perbandingan

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

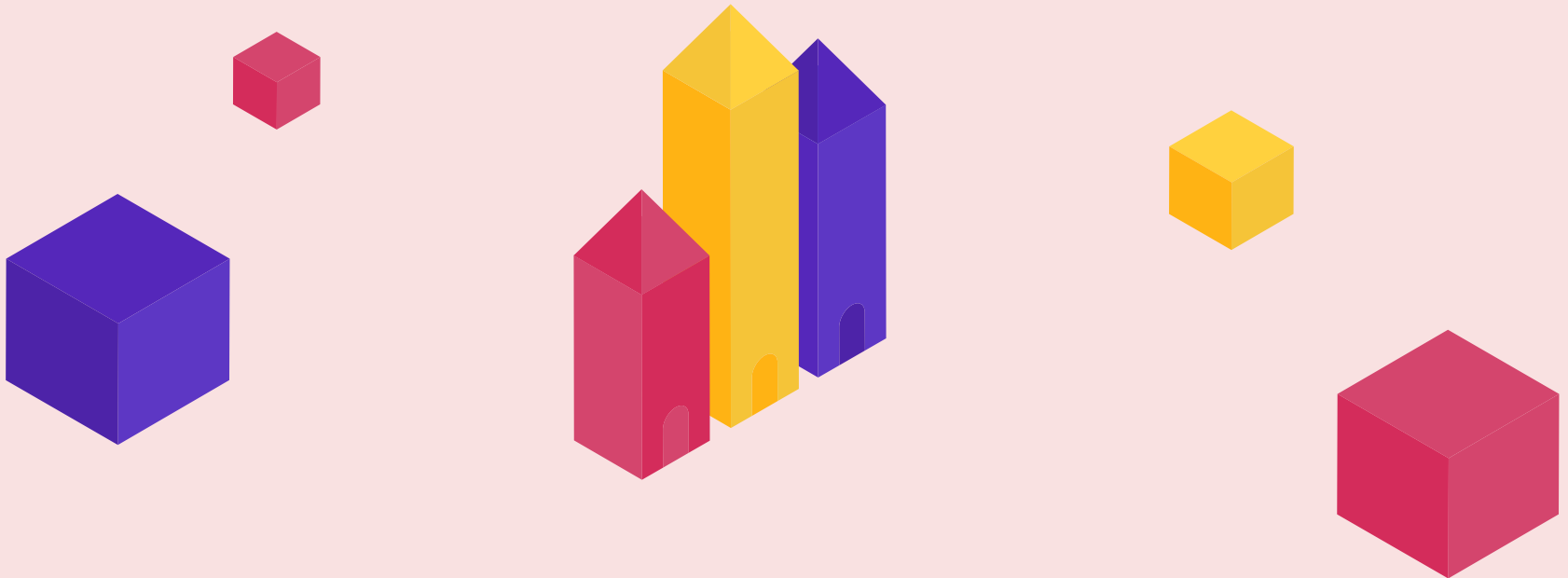
**Tabel Performa Hasil Prediksi Model
CNN dan XGBoost**

Simulation



- Dilakukan simulasi terhadap 1000 sampel data untuk melihat **running time**.
- Simulasi bekerja dengan waktu iterasi setiap 0.01 detik.
- Didapat bahwa sistem membutuhkan waktu rata-rata **3.5 detik** untuk mendeteksi gempa magnitudo ≥ 4 setelah P-wave pertama.
- Dengan memperhitungkan **delay**, sistem dapat benar-benar mendeteksi gempa setelah **7 detik dari P-wave pertama**.

04 Kesimpulan dan Saran



Kesimpulan dan Saran

1

Pemodelan sistem **early warning** bagi gempa dapat dilakukan dengan **mendeteksi adanya P-wave** yang dapat terdeteksi oleh sensor terlebih dahulu.

2

Jarak rata-rata antara P-wave dengan S-wave adalah **6.7 detik**. Pemilihan jumlah index cuplikan didasari untuk **memaksimalkan** antara performa dengan ukuran dataset.

3

Model **klasifikasi CNN** mampu menghasilkan **skor F1** yang lebih tinggi. Sebaliknya, model **regresi XGBoost** mampu menghasilkan skor **RMSE** yang lebih kecil. Saran ke depan adalah penerapan teknik **ensembling**.

4

Sistem yang dibuat **dianggap baik** karena **mampu mendeteksi gempa < 18.6 detik** yang merupakan rata-rata waktu interval antara P-wave pertama sampai akhir S-wave.

**Terima
Kasih**

