

# **JOINTS DATA COMPETITION 2023**



**TIM DATTA**

**Adristi Marsalma Nectarine Amaranthi  
Devina Fitri Handayani  
Fitri 'Aliyah**

## LATAR BELAKANG

Gempa bumi adalah bencana alam yang dapat menyebabkan kerusakan pada bangunan. Oleh karena itu, untuk memahami bagaimana bangunan dapat bertahan dan merencanakan tindakan pemulihan pasca-gempa, diperlukan analisis tingkat kerusakan bangunan. Analisis ini berguna untuk menilai karakteristik bangunan yang dapat mempengaruhi tingkat kerusakan saat terjadi gempa, serta memprediksi tingkat kerusakan yang mungkin terjadi pada bangunan tersebut.

Salah satu fokus dalam analisis karakteristik bangunan yang memiliki tingkat kerusakan rendah adalah untuk menyelidiki penyebab kerusakan rendah pada bangunan saat terjadi gempa. Melalui analisis ini, dapat diidentifikasi faktor-faktor desain, konstruksi, dan material bangunan yang dapat mengurangi kerusakan pada bangunan, sehingga dapat diambil langkah-langkah yang tepat untuk meningkatkan ketahanan bangunan terhadap gempa bumi.

Dengan menggunakan data dari gempa-gempa sebelumnya, serta mengambil kriteria-kriteria tertentu seperti desain, konstruksi, dan material bangunan, analisis dapat digunakan untuk memprediksi tingkat kerusakan yang mungkin terjadi pada bangunan setelah terjadinya gempa. Hal ini dapat membantu dalam mengidentifikasi bangunan yang rentan terhadap kerusakan serta mengarahkan upaya pemulihan pada bangunan yang membutuhkan perhatian lebih.

## TUJUAN DAN MANFAAT

Tujuan dari analisis karakteristik bangunan yang memiliki tingkat kerusakan rendah serta dapat membuat prediksi tingkat kerusakan bangunan setelah terjadi gempa adalah sebagai berikut:

1. Meningkatkan Ketahanan Bangunan terhadap Gempa Bumi  
Analisis karakteristik bangunan yang dapat mengidentifikasi faktor-faktor desain, konstruksi, dan material bangunan yang dapat mengurangi kerusakan pada bangunan saat terjadi gempa, dapat membantu dalam meningkatkan ketahanan bangunan terhadap gempa bumi. Dengan pemahaman yang lebih baik tentang perilaku struktural bangunan selama gempa, desain dan konstruksi bangunan dapat ditingkatkan untuk mengurangi kerusakan dan risiko cedera akibat gempa.
2. Mengurangi Dampak Negatif dari Gempa Bumi  
Dengan memprediksi tingkat kerusakan bangunan setelah terjadi gempa, dapat membantu dalam mengidentifikasi bangunan yang rentan terhadap kerusakan dan mengarahkan upaya pemulihan pasca-gempa pada bangunan yang membutuhkan perhatian lebih. Hal ini dapat mengurangi dampak negatif dari gempa bumi, baik secara ekonomi maupun sosial.
3. Mendukung Pengambilan Keputusan yang Berbasis Data  
Analisis karakteristik bangunan yang didasarkan pada data dan metode analisis yang canggih dapat memberikan informasi yang akurat dan objektif bagi para pengambil keputusan, seperti insinyur struktural, pemilik bangunan, dan pemerintah, dalam merencanakan langkah-langkah mitigasi risiko gempa. Keputusan yang didasarkan pada analisis yang baik dapat mengarahkan sumber daya secara efektif dan efisien untuk memperkuat bangunan yang rentan terhadap gempa.

## METODE

### A. *Software*

Dalam melakukan analisis data, kami menggunakan *tools* Google Colab dengan bahasa pemrograman Python. Kami menggunakan *library-library* untuk melakukan analisis dan prediksi, seperti pandas, numpy, matplotlib, seaborn, dan sklearn.

### B. *Dataset*

Dataset yang kami gunakan adalah dataset Gempa yang disediakan oleh pihak panitia Joints. Train dataset memiliki jumlah kolom 25 dan baris 722.815. Sedangkan, Test dataset memiliki jumlah kolom 24 dan baris 242.082. Berikut adalah fitur-fitur yang terdapat dalam dataset.

No	Nama Fitur	Penjelasan
1	floors before eq (total)	jumlah total lantai pada bangunan sebelum gempa terjadi
2	old building	usia bangunan (tahun)
3	plinth area (ft <sup>2</sup> )	luas bangunan (ft <sup>2</sup> )
4	height before eq (ft)	tinggi bangunan sebelum terjadi gempa (kaki)
5	land surface condition	kondisi permukaan tanah disaat bangunan akan dibangun
6	type of foundation	jenis fondasi yang dipakai untuk bangunan tsb

7	type of roof	jenis atap yang dipakai untuk bangunan tsb
8	type of ground floor	jenis lantai yang dipakai untuk <i>ground-floor</i>
9	type of other floor	jenis lantai yang dipakai untuk selain <i>ground-floor</i>
10	position	Posisi bangunan tersebut apakah bersisian secara langsung dengan bangunan lain
11	building_plan_configuration	konfigurasi bangunan yang berkaitan dengan bentuk, ukuran, dan penempatan struktur utama bangunan
12	technical_solution_proposed	Solusi yang ditawarkan untuk bangunan yang terdampak gempa
13	legal_ownership_status	Status kepemilikan bangunan
14	has_secondary_use	Keterangan apakah bangunan memiliki kegunaan sekunder
15	type of reinforcement concrete	Tipe beton bertulang/reinforcement concrete
16	residential_type	Tipe penggunaan sebagai hunian
17	no_family_residing	Jumlah keluarga yang tinggal dalam bangunan tersebut
18	public_place_type	Tipe penggunaan sebagai tempat umum
19	industrial_use_type	Tipe industri
20	governmental_use_type	Tipe penggunaan sebagai bangunan pemerintahan
21	flexible_superstructure	Keterangan penggunaan superstructure yang fleksibel
22	wall_binding	Material yang digunakan sebagai perekat bahan pembentuk dinding
23	wall_material	Material dasar sebagai pembangun dinding
24	damage_grade (variabel target)	Tingkat kerusakan yang disebabkan oleh gempa (1-5)

### C. Data Cleaning

Sebelum dilakukan *cleaning*, perlu diketahui karakteristik dari setiap data. Berikut adalah deskripsi data dari setiap kolom dalam dataset Train.

index	kolom	dataType	null	nullPct	uniqVal
0	Unnamed: 0	int64	0	0.0	520011
1	floors_before_eq (total)	object	332806	46.04	47
2	old_building	float64	239204	33.09	160
3	plinth_area (ft^2)	object	421208	58.27	930
4	height_before_eq (ft)	float64	332806	46.04	77
5	land_surface_condition	object	301606	41.73	3
6	type_of_foundation	object	239204	33.09	12
7	type_of_roof	object	421208	58.27	12
8	type_of_ground_floor	object	332806	46.04	15
9	type_of_other_floor	object	301606	41.73	11
10	position	object	312006	43.17	4
11	building_plan_configuration	object	301606	41.73	10
12	technical_solution_proposed	object	676014	93.53	4
13	legal_ownership_status	object	124802	17.27	13
14	has_secondary_use	float64	197604	27.34	2
15	type_of_reinforcement_concrete	float64	291206	40.29	4
16	residential_type	object	270404	37.41	6
17	no_family_residing	object	145602	20.14	11
18	public_place_type	object	0	0.0	13
19	industrial_use_type	object	114402	15.83	9
20	governmental_use_type	object	249604	34.53	3

21	flexible_superstructure	object	62400	8.63	2
22	wall_binding	float64	62400	8.63	6
23	wall_material	float64	228804	31.65	4
24	damage_grade	float64	0	0.0	5

Setelah itu, berikut tahapan kami dalam *cleaning data*

1. Mengganti nama kolom  
Pada tahap ini, kami mengganti kolom 'Unnamed: 0' pada data Train menjadi 'id'. Selain itu, mengganti kolom 'position' pada data Train dan Test menjadi 'no\_building\_attached'
2. Menangani data yang tidak akurat pada dataset Test dan Train  
Pada tahap ini, berikut adalah data yang kami tangani
  - 1) Pada kolom 'floors\_before\_eq (total)', kami mengganti nilai-nilai yang typo atau tidak konsisten, menjadi data numerik.
  - 2) Pada kolom 'plinth\_area (ft^2)', kami menghapus kata ' (ft^2)' di setiap value dan mengubah data menjadi numerik.
  - 3) Pada kolom 'type\_of\_foundation', kami melakukan merge data dan konsistensi data. Grouping dilakukan pada value yang sama, tetapi tertulis berbeda, seperti 'Reinforced brick concrete/rcc/rb' dan 'Reinforced Brick Slab/rcc/rbc', menjadi 'Reinforced Brick Concrete'
  - 4) Pada kolom 'type\_of\_ground\_floor', kami mengkapitalisasi dan mengkonsistensikan setiap value
  - 5) Pada kolom 'type\_of\_other\_floor', kami melakukan merge data dan konsistensi data. Grouping dilakukan pada value yang sama, tetapi tertulis berbeda, seperti 'Reinforced brick concrete/rcc/rbc' dan 'reinforced cement concrete/rb/rbc', menjadi 'RCC/RB/RBC'
  - 6) Pada kolom 'legal\_ownership\_status', kami mengganti nilai-nilai yang typo dan melakukan merge pada data-data yang memiliki karakteristik yang mirip.
  - 7) Pada kolom 'residential\_type', 'public\_place\_type', 'industrial\_use\_type', dan 'governmental\_use\_type', kami melakukan merge pada data-data yang memiliki karakteristik yang mirip.
  - 8) Pada kolom 'no\_building\_attached', kami mengganti data menjadi numerik.
3. Menyesuaikan tipe data  
Kami mengubah data pada kolom ['old\_building', 'floors\_before\_eq (total)', 'height\_before\_eq (ft)', 'no\_building\_attached', 'has\_secondary\_use', 'type\_of\_reinforcement\_concrete', 'wall\_binding', 'wall\_material', 'plinth\_area (ft^2)', 'no\_family\_residing'] menjadi data float.
4. Menghilangkan data yang duplikat  
Kami menghapus baris yang duplikat untuk meningkatkan kualitas data dan mencegah redundansi.
5. Menangani data *null*  
Kami menghapus kolom 'technical\_solution\_proposed' karena diketahui lebih dari 90% data tersebut adalah null. Setelah itu, kami melakukan drop ke semua baris yang memiliki nilai null.
6. Menangani *outlier*  
Kami hanya menangani *outlier* pada kolom 'old\_building' dengan hanya mengambil 0,995% dari data. Untuk kolom numerik lain, seperti floors\_before\_eq (total), plinth\_area (ft^2), dan height\_before\_eq (ft), kami tidak menghilangkan outlier karena data kami asumsikan normal dilihat dari minimum dan maksimum data tersebut.

Dataset Train akhir yang kami gunakan untuk melakukan prediksi memiliki jumlah baris sebanyak 301607 dan kolom sebanyak 24.

#### D. Algoritma Prediksi

Terdapat beberapa tahapan untuk melakukan prediksi Berikut penjelasan dari tahapan-tahapan yang kami lakukan.

1. Melakukan *label encoding*

Tahapan ini bertujuan untuk mengubah kolom kategorikal yang bertipe non-numerikal ke kategori numerikal sehingga sesuai dengan kebutuhan model yang akan digunakan. Kolom yang kami *encode* adalah kolom 'land\_surface\_condition', 'type\_of\_foundation', 'type\_of\_roof', 'type\_of\_ground\_floor', 'type\_of\_other\_floor', 'building\_plan\_configuration', 'legal\_ownership\_status', 'residential\_type', 'public\_place\_type', 'industrial\_use\_type', 'govermental\_use\_type', dan 'flexible\_superstructure'.

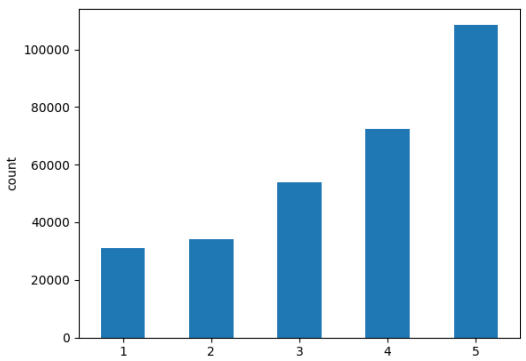
2. Memisahkan kolom 'damage\_grade' pada dataset 'train\_raw'  
Tahapan ini bertujuan untuk memisahkan variabel bebas dan variabel terikat ('damage\_grade'). Tahapan ini menghasilkan dua dataset, yaitu X dan y. X berisi semua kolom kecuali 'damage\_grade', sedangkan y berisi kolom 'damage\_grade'.
3. Membagi dataset untuk *training* dan *testing*  
Tahapan ini bertujuan untuk memisahkan data yang digunakan membangun model prediksi dan data yang akan diprediksi oleh model yang sudah dibuat. Tahapan tersebut dilakukan terhadap dua dataset X dan y. Tahapan ini menghasilkan empat dataset baru yaitu X\_train, X\_test, y\_train, y\_test.
4. Menginisiasi model prediksi dan melakukan prediksi terhadap data *testing*  
Model prediksi yang dibuat akan dibangun berdasarkan dataset X\_train dan y\_train. Kemudian model tersebut akan digunakan untuk memprediksi nilai 'damage\_grade' berdasarkan input dari X\_test, dan memberikan output berupa hasil prediksi yang disimpan dalam dataset rf\_pred.
5. Memvisualisasikan *confusion matrix* dari hasil prediksi  
*Confusion matrix* akan dibuat menggunakan rf\_pred dan y\_test. y\_test merupakan nilai aktual dari 'damage\_grade' yang diprediksi, sedangkan rf\_pred merupakan nilai prediksi dari hasil prediksi tersebut. *Confusion matrix* akan memetakan nilai aktual dan prediksi untuk setiap nilai unik pada kolom 'damage\_grade', yaitu 1, 2, 3, 4, 5.
6. Mengevaluasi performa model  
Tahapan ini bertujuan untuk mengevaluasi performa model. Model tersebut dievaluasi menggunakan f1-score (macro). Berikut rumus dari f1-score macro.

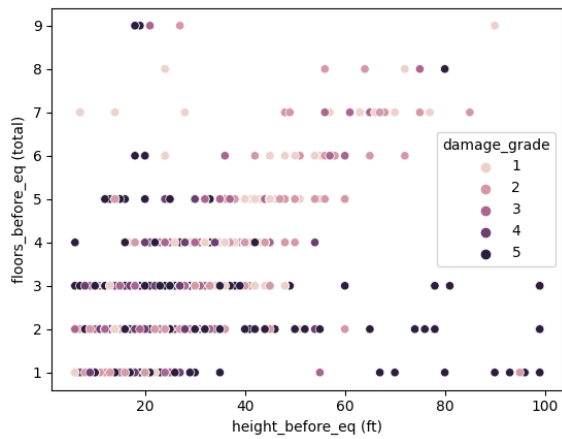
$$precision = \frac{\sum True\ Positive}{\sum True\ Positive + \sum False\ Positive}$$

$$recall = \frac{\sum True\ Positive}{\sum True\ Positive + \sum False\ Negative}$$

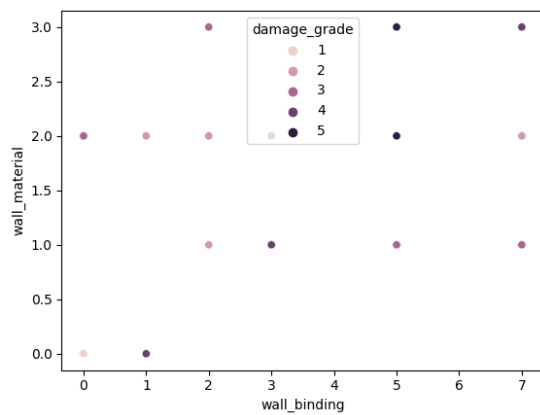
$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

## ANALISIS DISERTAI EDA

No	EDA	Hasil Analisis
1	<p>Persebaran damage_grade</p> 	<p>Modus dari damage_grade berada pada level 5, dengan jumlah lebih dari 100000. Nilai yang paling jarang ditemui pada damage_grade adalah 1. Nilai damage grade cenderung memiliki negative skew.</p>
2	<p>Analisis hubungan tinggi bangunan dan jumlah lantai dengan damage grade</p>	<p>Secara logika, semakin tinggi bangunan, semakin banyak jumlah lantai bangunan tersebut. Maka dari itu, kami hanya melihat data yang berbanding lurus. Oleh karena itu, diketahui bahwa semakin tinggi bangunan dan semakin banyak lantainya, damage grade dari bangunan tersebut semakin kecil.</p>

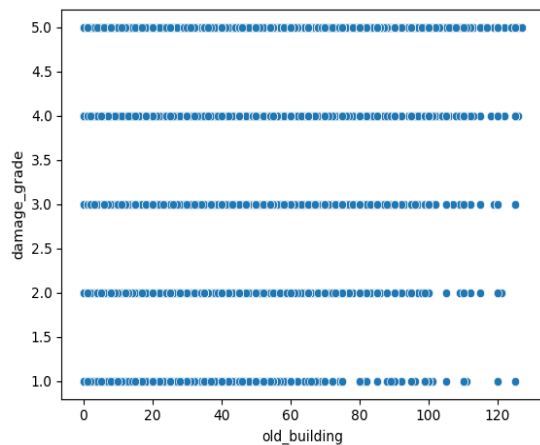


3 Analisis hubungan wall material dan wall binding dengan damage grade

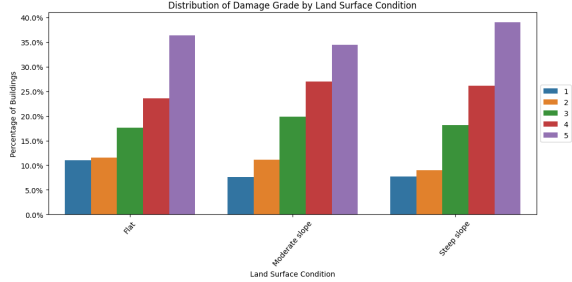
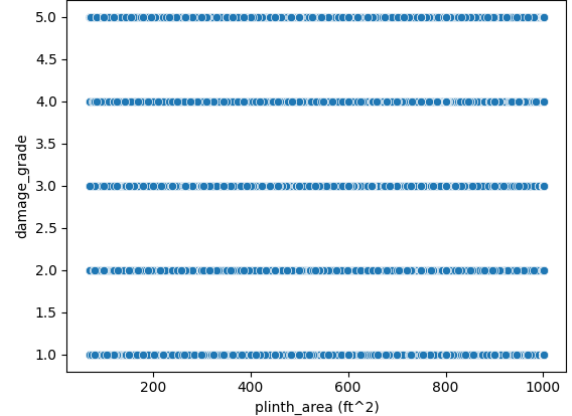
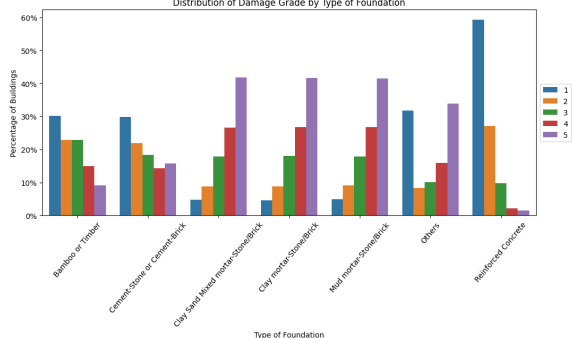
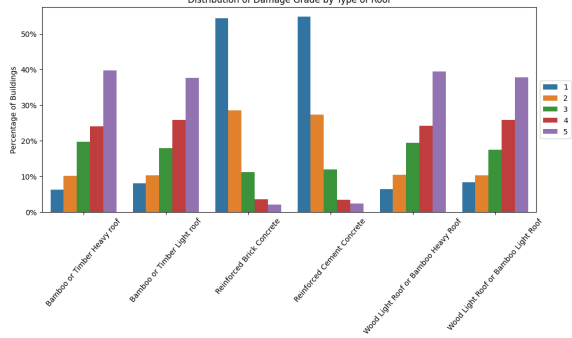


Diketahui bahwa wall\_binding bertipe mortar cement dan wall\_material bertipe stone bricks memiliki damage grade yang cukup rendah.

4 Analisis hubungan old\_building dengan damage grade



Secara keseluruhan, hubungan kedua variabel dengan umur bangunan dari 0 - 60 tahun memiliki tingkat damage grade yang beragam. Namun, semakin tua umur bangunan semakin besar kemungkinan memiliki damage grade yang tinggi.

5	<p>Analisis hubungan land_surface_condition dengan damage grade</p> 	<p>Berdasarkan diagram di samping, dapat dilihat karakter land_surface_condition dengan tingkat kerusakan paling rendah adalah moderate slope (kecuraman menengah). Bangunan yang tahan gempa adalah bangunan yang dibangun pada kecuraman menengah.</p>
6	<p>Analisis hubungan plinth_area dengan damage grade</p> 	<p>Hubungan kedua variabel antara plinth_area (ft<sup>2</sup>) dengan damage grade terlihat tidak saling mempengaruhi atau dalam kata lain besar plinth area tidak mengakibatkan damage grade semakin tinggi atau semakin kecil karena plot tersebar sangat merata</p>
7	<p>Distribution of Damage Grade by Type of Foundation</p> 	<p>Berdasarkan diagram di samping, bangunan yang memiliki tingkat kerusakan rendah memiliki tipe pondasi mayoritas berupa reinforced concrete. Selain itu, juga bamboo or timber dan cement-stone or cement-brick.</p>
8	<p>Distribution of Damage Grade by Type of Roof</p> 	<p>Berdasarkan diagram di samping, bangunan yang memiliki tingkat kerusakan rendah memiliki tipe atap mayoritas berupa reinforced brick concrete dan reinforced cement concrete</p>

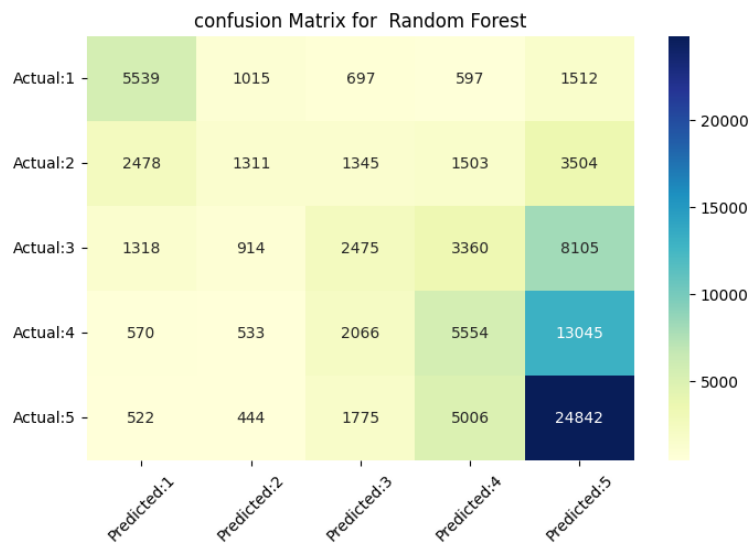
9	<p>Distribution of Damage Grade by Type of Ground Floor</p> <p>Type of Ground Floor</p>	<p>Berdasarkan diagram di samping, bangunan yang memiliki tingkat kerusakan rendah memiliki tipe ground floor mayoritas berupa reinforced concrete</p>
10	<p>Distribution of Damage Grade by Type of Other Floor</p> <p>Type of Other Floor</p>	<p>Berdasarkan diagram di samping, bangunan yang memiliki tingkat kerusakan rendah memiliki tipe floor selain ground floor mayoritas berupa RCC/RB/RBC</p>
11	<p>Distribution of Damage Grade by Number of Building Attached</p> <p>Number of Building Attached</p>	<p>Berdasarkan diagram di samping, dapat dilihat nilai number_of_building_attached dengan tingkat kerusakan paling rendah adalah 3. Dengan demikian, bangunan yang berhimpitan dengan tiga bangunan lain adalah bangunan yang paling tahan gempa.</p>
12	<p>Distribution of Damage Grade by Legal Ownership Status</p> <p>Legal Ownership Status</p>	<p>Berdasarkan diagram di samping, bangunan yang memiliki tingkat kerusakan rendah memiliki legal ownership status public</p>
13	<p>Distribution of Damage Grade by Type of Reinforcement Concrete</p> <p>Type of Reinforcement Concrete</p>	<p>Berdasarkan diagram di samping, bangunan yang memiliki tingkat kerusakan rendah memiliki tipe reinforcement concrete 3, yaitu bangunan yang reinforcement concretenya gabungan antara non-engineered reinforcement dan engineered reinforcement. Jika dilihat dari gambar di samping pula, bangunan yang memiliki tipe 2 atau yang engineered reinforcement juga memiliki tingkat kerusakan yang rendah.</p>



14	<p>Distribution of Damage Grade by Residential Type</p> <table><thead><tr><th>Residential Type</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr></thead><tbody><tr><td>Non-residential</td><td>10.0%</td><td>11.0%</td><td>18.0%</td><td>24.0%</td><td>35.0%</td></tr><tr><td>Other</td><td>34.0%</td><td>25.0%</td><td>19.0%</td><td>10.0%</td><td>12.0%</td></tr><tr><td>Residential</td><td>34.0%</td><td>21.0%</td><td>20.0%</td><td>11.0%</td><td>14.0%</td></tr></tbody></table>	Residential Type	1	2	3	4	5	Non-residential	10.0%	11.0%	18.0%	24.0%	35.0%	Other	34.0%	25.0%	19.0%	10.0%	12.0%	Residential	34.0%	21.0%	20.0%	11.0%	14.0%	<p>Berdasarkan diagram di samping, dapat dilihat nilai residential_type dengan tingkat kerusakan paling rendah adalah residential. Dengan demikian, bangunan hunian cenderung lebih tahan gempa.</p>																																																						
Residential Type	1	2	3	4	5																																																																											
Non-residential	10.0%	11.0%	18.0%	24.0%	35.0%																																																																											
Other	34.0%	25.0%	19.0%	10.0%	12.0%																																																																											
Residential	34.0%	21.0%	20.0%	11.0%	14.0%																																																																											
15	<p>Distribution of Damage Grade by Number of Family Residing</p> <table><thead><tr><th>Number of Family Residing</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr></thead><tbody><tr><td>0.0</td><td>15%</td><td>10%</td><td>10%</td><td>20%</td><td>25%</td></tr><tr><td>1.0</td><td>10%</td><td>10%</td><td>15%</td><td>20%</td><td>35%</td></tr><tr><td>2.0</td><td>10%</td><td>10%</td><td>15%</td><td>20%</td><td>35%</td></tr><tr><td>3.0</td><td>10%</td><td>10%</td><td>15%</td><td>20%</td><td>35%</td></tr><tr><td>4.0</td><td>10%</td><td>10%</td><td>15%</td><td>20%</td><td>35%</td></tr><tr><td>5.0</td><td>10%</td><td>10%</td><td>15%</td><td>20%</td><td>35%</td></tr><tr><td>6.0</td><td>10%</td><td>10%</td><td>15%</td><td>20%</td><td>35%</td></tr><tr><td>7.0</td><td>10%</td><td>10%</td><td>15%</td><td>20%</td><td>35%</td></tr><tr><td>8.0</td><td>80%</td><td>10%</td><td>10%</td><td>10%</td><td>10%</td></tr><tr><td>9.0</td><td>20%</td><td>10%</td><td>10%</td><td>10%</td><td>10%</td></tr><tr><td>10.0</td><td>10%</td><td>10%</td><td>10%</td><td>10%</td><td>10%</td></tr><tr><td>11.0</td><td>10%</td><td>10%</td><td>10%</td><td>10%</td><td>10%</td></tr></tbody></table>	Number of Family Residing	1	2	3	4	5	0.0	15%	10%	10%	20%	25%	1.0	10%	10%	15%	20%	35%	2.0	10%	10%	15%	20%	35%	3.0	10%	10%	15%	20%	35%	4.0	10%	10%	15%	20%	35%	5.0	10%	10%	15%	20%	35%	6.0	10%	10%	15%	20%	35%	7.0	10%	10%	15%	20%	35%	8.0	80%	10%	10%	10%	10%	9.0	20%	10%	10%	10%	10%	10.0	10%	10%	10%	10%	10%	11.0	10%	10%	10%	10%	10%	<p>Berdasarkan diagram di samping, bangunan yang memiliki tingkat kerusakan rendah memiliki jumlah keluarga yang tinggal dibangunan tersebut sekitar 8. Selain itu, dapat dilihat bahwa semakin tinggi jumlah keluarga, cenderung semakin kecil tingkat kerusakannya, tetapi maksimal hanya 9 keluarga.</p>
Number of Family Residing	1	2	3	4	5																																																																											
0.0	15%	10%	10%	20%	25%																																																																											
1.0	10%	10%	15%	20%	35%																																																																											
2.0	10%	10%	15%	20%	35%																																																																											
3.0	10%	10%	15%	20%	35%																																																																											
4.0	10%	10%	15%	20%	35%																																																																											
5.0	10%	10%	15%	20%	35%																																																																											
6.0	10%	10%	15%	20%	35%																																																																											
7.0	10%	10%	15%	20%	35%																																																																											
8.0	80%	10%	10%	10%	10%																																																																											
9.0	20%	10%	10%	10%	10%																																																																											
10.0	10%	10%	10%	10%	10%																																																																											
11.0	10%	10%	10%	10%	10%																																																																											
16	<p>Distribution of Damage Grade by Public Place Type</p> <table><thead><tr><th>Public Place Type</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr></thead><tbody><tr><td>Non-public</td><td>30.0%</td><td>30.0%</td><td>20.0%</td><td>25.0%</td><td>35.0%</td></tr><tr><td>Public</td><td>40.0%</td><td>25.0%</td><td>15.0%</td><td>10.0%</td><td>10.0%</td></tr></tbody></table>	Public Place Type	1	2	3	4	5	Non-public	30.0%	30.0%	20.0%	25.0%	35.0%	Public	40.0%	25.0%	15.0%	10.0%	10.0%	<p>Berdasarkan diagram di samping, dapat dilihat nilai public_place_type dengan tingkat kerusakan paling rendah adalah public. Dengan demikian, fasilitas publik cenderung lebih tahan gempa.</p>																																																												
Public Place Type	1	2	3	4	5																																																																											
Non-public	30.0%	30.0%	20.0%	25.0%	35.0%																																																																											
Public	40.0%	25.0%	15.0%	10.0%	10.0%																																																																											
17	<p>Distribution of Damage Grade by Industrial Use Type</p> <table><thead><tr><th>Industrial Use Type</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr></thead><tbody><tr><td>Industrial</td><td>5.0%</td><td>10.0%</td><td>20.0%</td><td>30.0%</td><td>35.0%</td></tr><tr><td>Non-industrial</td><td>10.0%</td><td>10.0%</td><td>15.0%</td><td>20.0%</td><td>35.0%</td></tr></tbody></table>	Industrial Use Type	1	2	3	4	5	Industrial	5.0%	10.0%	20.0%	30.0%	35.0%	Non-industrial	10.0%	10.0%	15.0%	20.0%	35.0%	<p>Berdasarkan diagram di samping, dapat dilihat nilai industrial_use_type dengan tingkat kerusakan paling rendah adalah industrial. Dengan demikian, bangunan industri cenderung lebih tahan gempa.</p>																																																												
Industrial Use Type	1	2	3	4	5																																																																											
Industrial	5.0%	10.0%	20.0%	30.0%	35.0%																																																																											
Non-industrial	10.0%	10.0%	15.0%	20.0%	35.0%																																																																											
18	<p>Distribution of Damage Grade by Governmental Use Type</p> <table><thead><tr><th>Governmental Use Type</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr></thead><tbody><tr><td>Governmental</td><td>25.0%</td><td>20.0%</td><td>15.0%</td><td>15.0%</td><td>25.0%</td></tr><tr><td>Non-governmental</td><td>10.0%</td><td>10.0%</td><td>15.0%</td><td>20.0%</td><td>35.0%</td></tr></tbody></table>	Governmental Use Type	1	2	3	4	5	Governmental	25.0%	20.0%	15.0%	15.0%	25.0%	Non-governmental	10.0%	10.0%	15.0%	20.0%	35.0%	<p>Berdasarkan diagram di samping, dapat dilihat nilai governmental_use_type dengan tingkat kerusakan paling rendah adalah governmental. Dengan demikian, fasilitas pemerintahan cenderung lebih tahan gempa.</p>																																																												
Governmental Use Type	1	2	3	4	5																																																																											
Governmental	25.0%	20.0%	15.0%	15.0%	25.0%																																																																											
Non-governmental	10.0%	10.0%	15.0%	20.0%	35.0%																																																																											
19	<p>Distribution of Damage Grade by Flexible Superstructure</p> <table><thead><tr><th>Flexible Superstructure</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr></thead><tbody><tr><td>available</td><td>15.0%</td><td>15.0%</td><td>20.0%</td><td>20.0%</td><td>20.0%</td></tr><tr><td>unavailable</td><td>10.0%</td><td>10.0%</td><td>15.0%</td><td>20.0%</td><td>35.0%</td></tr></tbody></table>	Flexible Superstructure	1	2	3	4	5	available	15.0%	15.0%	20.0%	20.0%	20.0%	unavailable	10.0%	10.0%	15.0%	20.0%	35.0%	<p>Berdasarkan diagram di samping, bangunan yang memiliki tingkat kerusakan rendah cenderung memiliki super struktur yang fleksibel.</p>																																																												
Flexible Superstructure	1	2	3	4	5																																																																											
available	15.0%	15.0%	20.0%	20.0%	20.0%																																																																											
unavailable	10.0%	10.0%	15.0%	20.0%	35.0%																																																																											

## HASIL DAN PEMBAHASAN

Berikut confusion matrix dari salah satu prediksi yang kami lakukan.



Berdasarkan confusion matrix tersebut, modus dari hasil pemetaan nilai prediksi dan nilai aktual ‘damage\_grade’ ditemukan pada nilai aktual = 5 dan prediksi = 5, yaitu sebanyak 24842 baris. Modus dari hasil nilai prediksi adalah 5, dengan hasil prediksi sebanyak 51008 yang berarti mayoritas damage\_grade diprediksi sebagai 5. Modus dari nilai aktual ‘damage\_grade’ adalah 5, yaitu sebesar 32589. Modus dari kesalahan hasil prediksi ditemukan pada nilai prediksi 5, yaitu dengan jumlah kesalahan prediksi sebanyak 26166. Selain menggunakan *confusion matrix*, performa model juga kami nilai berdasarkan f1-score macro. Nilai f1-score yang kami dapatkan dari beberapa percobaan prediksi berkisar pada 36. Berdasarkan kedua hasil evaluasi, maka dapat dikatakan bahwa performa model tidak cukup baik, karena nilai f1-score masih dibawah 50 (*What Is a Good F1 Score?*, 2022).

Setelah dianalisis lebih lanjut, ditemukan bahwa banyak variabel kategorikal yang tidak seimbang (*imbalanced*) yang tidak kami tangani. Kondisi *imbalanced* terjadi ketika frekuensi dari sebuah nilai jauh lebih tinggi atau jauh lebih sedikit dari nilai yang lainnya. Hal tersebut akan mengurangi performa model dikarenakan model tidak cukup belajar untuk nilai dengan frekuensi yang sangat sedikit, sehingga menyebabkan bias saat prediksi dilakukan.

## KESIMPULAN

Kasus ini menuntut kami untuk memprediksi tingkat kerusakan bangunan berdasarkan karakteristik yang telah diberikan pada dataset. Setelah dilakukan prediksi menggunakan model Random Forest melalui tahapan - tahapan yang telah dipaparkan di atas, didapatkan hasil f1 score mencapai 0.36 yang berarti model tidak cukup baik untuk digunakan sebagai prediksi. Berdasarkan analisa EDA, bangunan yang cenderung lebih tahan gempa memiliki karakteristik bangunan yang tinggi dan jumlah lantai yang banyak; memiliki material perekat dinding mortar dan cement, serta menggunakan material dasar batu bata; umur bangunan cenderung di bawah 60 tahun, dibangun di atas tanah yang memiliki kecuraman menengah; memiliki tipe pondasi *reinforcement concrete*; memiliki tipe atap *reinforced brick concrete* atau *reinforced cement concrete*; memiliki tipe lantai, baik itu ground floor maupun bukan, berupa *reinforcement concrete*; berhimpitan dengan 3 bangunan lain; memiliki kepemilikan legal publik; memiliki reinforcement concrete type 2 dan 3, yaitu terdapat *engineered reinforcement*; merupakan bangunan hunian; memiliki jumlah keluarga yang tinggal bangunan sekitar 8 keluarga; merupakan tempat publik; merupakan tempat industri; merupakan bangunan pemerintahan; dan memiliki superstruktur yang fleksibel.

## REFERENSI

*What is a good F1 score? Simply explained* (2022). (2022, April 20). Stephen Allwright. <https://stephenallwright.com/good-f1-score/>