

Laporan Praktikum Minggu 2: Data Wrangling & Exploratory Data Analysis (EDA)

Nama : Raki Raihan

NIM : 2411070095

URL W&B Project : <https://wandb.ai/han-dev321-stikomelrahma/titanic-eda-2026?nw=nwuserhandev321>

1. Pendahuluan

Praktikum ini tujuannya adalah untuk memahami bagaimana proses dari data Wrangling & Exploratory Data Analysis (EDA).

Data wrangling adalah suatu proses data dimana mentransformasikan data mentah menjadi data bersih serta menyiapkan data agar dapat digunakan dengan baik.

Dataset Titanic digunakan karena disiapkan untuk pengenalan serta pelatihan analisis dalam machine learning.

2. Analisis Data Mentah (Inspeksi/Scanning)

Berdasarkan hasil analisis dari inspeksi data menggunakan query(`df.info`, `df.isnull`, `.sum`), maka ditemukanlah kolom yang memiliki missing values

- **Age** : 177 data kosong.
- **Embarked** : 2 data kosong.
- **Cabin** : 687 data kosong.

3. Strategi Pembersihan Data

- Imputasi *Age* menggunakan *Median* karena data *Age* banyak yang kosong dan tidak merata, sehingga menggunakan *median* lebih stabil serta lebih kebal terhadap nilai ekstrem
 - Imputasi *Embarked* menggunakan *Modus* karena merupakan data kategori, sehingga bisa diambil dari nilai yang sering muncul dan dianggap paling mewakili
 - Kolom *Cabin* di buang karena banyak data yang kosong (+70% data kosong), sehingga apabila tetap diisi justru akan menciptakan informasi palsu yang merusak model dan akan berakibat fatal pada kualitas model
-

4. Visualisasi dan Temuan Utama

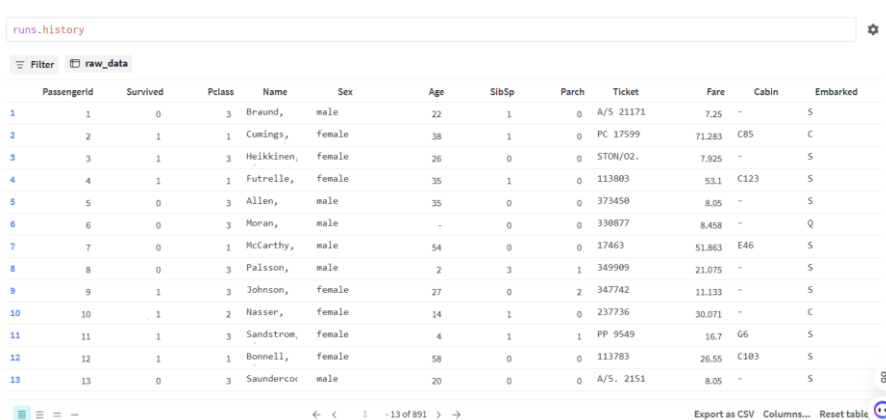
- **Survival Rate by Gender** : Berdasarkan hasil grafik/visualisasi, penumpang perempuan lah yang memiliki tingkat keselamatan paling tinggi dibandingkan dengan laki-laki, karena penyelamatan lebih diutamakan kepada perempuan dan anak-anak.
- **Distribusi Umur** : Berdasarkan hasil grafik (Visualisasi), Titanic didominasi oleh penumpang orang dewasa muda **usia 20 – 30**, serta dengan kehadiran anak-anak yang cukup menonjol, akan tetapi sangat sedikit penumpang yang berusia lanjut

5. Feature Engineering

- **Fitur baru FamilySize di buat dengan rumus:**
Formula: $\text{FamilySize} = \text{SibSp} + \text{Parch} + 1$
- **Analisis** : berdasarkan hasil dari visualisasi menunjukan bahwa jumlah keluarga berpengaruh terhadap peluang selamat.
- **Encoding** : proses encoding dilakukan pada data yang kategorikal, di mana dalam kasus ini ada pada (“Sex”) dengan cara mengubah nilai “text” menjadi nilai “Biner/Angka (0/1)” agar dapat di proses oleh algoritma yang ada di dalam machine learning.

6. Integrasi Weights & Biases (W&B)

1. Before



	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	0	3	Braund,	male	22	1	0	A/5 21171	7.25	-	S
2	2	1	1	Cummings,	female	38	1	0	PC 17599	71.283	C85	C
3	3	1	3	Heikkinen,	female	26	0	0	STON/O2.	7.925	-	S
4	4	1	1	Futrelle,	female	35	1	0	113803	53.1	C123	S
5	5	0	3	Allen,	male	35	0	0	373450	8.05	-	S
6	6	0	3	Moran,	male	-	0	0	330877	8.458	-	Q
7	7	0	1	McCarthy,	male	54	0	0	17463	51.863	E46	S
8	8	0	3	Palsson,	male	2	3	1	349909	21.075	-	S
9	9	1	3	Johnson,	female	27	0	2	347742	11.133	-	S
10	10	1	2	Nasser,	female	14	1	0	237736	30.071	-	C
11	11	1	3	Sandstrom,	female	4	1	1	PP 9549	16.7	G6	S
12	12	1	1	Bonnell,	female	58	0	0	113783	26.55	C103	S
13	13	0	3	Saunderscox	male	20	0	0	A/5. 2151	8.05	-	S

2. After

runs.history

Filter final_processed_data

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	FamilySize
1	1	0	3	Braund,	0	22	1	0	A/5 21171	7.25	2	2
2	2	1	1	Cumings,	1	38	1	0	PC 17599	71.283	0	2
3	3	1	3	Heikkinen,	1	26	0	0	STON/O2.	7.925	2	1
4	4	1	1	Futrelle,	1	35	1	0	113803	53.1	2	2
5	5	0	3	Allen,	0	35	0	0	373450	8.05	2	1
6	6	0	3	Moran,	0	28	0	0	330877	8.458	1	1
7	7	0	1	McCarthy,	0	54	0	0	17463	51.863	2	1
8	8	0	3	Palsson,	0	2	3	1	349909	21.075	2	5
9	9	1	3	Johnson,	1	27	0	2	347742	11.133	2	3
10	10	1	2	Nasser,	1	14	1	0	237736	30.071	0	2
11	11	1	3	Sandstrom,	1	4	1	1	PP 9549	16.7	2	3
12	12	1	1	Bonnell,	1	58	0	0	113783	26.55	2	1
13	13	0	3	Saunders,	0	20	0	0	A/5. 2151	8.05	2	1

Pada hasil layout W&B Tables, ditampilkan 2 tabel data yaitu table data mentah (raw_data) dan table data yang sudah di proses (final_processed_data). Dari table-table ini, penganalisis dapat dengan mudah melihat perubahan data setelah di proses inpeksi/scanning, pembersihan/imputasi, dan encoding.

Fitur filtering dan sorting memudahkan dalam membantu mengelompokkan dan mencari data dengan lebih cepat, sehingga menganalisis data menjadi lebih mudah dan efisien disbanding hanya dengan menggunakan tampilah '.head()' yang ada pada notebook.

7. Kesimpulan & Refleksi

Tantangan tersulit dalam kesempatan praktikum kali ini adalah memahami (EDA) serta memilih strategi imputasi dan feature engineering. Praktikum kali ini menggambarkan tahap EDA sangat penting untuk memahami karakteristik data sehingga model yang akan dibangun selanjutnya akan lebih akurat