

Nama	: Anton Wijaya
Universitas	: Universitas Tidar

I. Deskripsi Dataset

- Sumber dataset asli: <https://www.kaggle.com/code/kacang/mlbb-eda-cleaning-web-scrapping/input>
- Google colab: https://colab.research.google.com/drive/13w78G2N4Yu_19aqUrWQMrXciRcPb58w5?usp=sharing
- Dataset yang digunakan pada tugas ini adalah dataset hero di MLBB yang berisi berbagai jenis data lengkap seperti role, HP, physical defense, dan lain- lain. Pada tugas ini, saya melakukan tranformasi data untuk membuat dataset hero mlbb untuk kepentingan *scene* kompetitif. Data yang ada pada dataset saya yaitu nama hero, role hero, pick rate, win rate, dan ban rate. Data- data tersebut dipilih karena data tersebutlah yang digunakan untuk kompetisi MLBB. Dataset telah diperbarui dengan menambahkan data hero.

Print Data yang asli

```
dataon1.head()
```

	hero_name	role	defense_overall	offense_overall	skill_effect_overall	difficulty_overall	movement_spd	magic_defense	mana	hp_regen	physical_atk	physical_defense	hp	attack_speed	mana_regen	win_rate	pick_rate	ban_rate
0	Terzla	fighter	7.0	8.0	6.0	6.0	255	10	430	54	129	19	2728	0.80	21.0	47.00	0.15	0.02
1	Martis	fighter	6.0	8.0	5.0	5.0	260	10	0	35	128	25	2738	0.86	0.0	45.00	0.15	0.02
2	Grock	tank	8.0	5.0	6.0	4.0	260	10	430	42	135	21	2819	0.81	12.0	47.69	0.14	0.02
3	Carmilla	support	5.0	5.0	9.0	5.0	255	10	430	39	126	25	2528	0.91	16.0	53.76	0.08	0.02
4	Iriethel	marksman	5.0	5.0	8.0	7.0	260	10	438	35	118	17	2540	0.82	15.0	51.52	0.38	0.03

Mengecek Nilai Kosong

Gambar 1. Dataset sebelum

```
print(databaru2) #print data baru
```

	nama_hero	role	win_rate	pick_rate	ban_rate	release_year
0	Terizla	fighter	47.00	0.15	0.02	2019
1	Martis	fighter	45.00	0.15	0.02	2018
2	Grock	tank	47.69	0.14	0.02	2017
3	Carmilla	support	53.76	0.08	0.02	2020
4	Iriethel	marksman	51.52	0.38	0.03	2017
..
98	Esmeralda	mage	49.14	1.25	42.29	2019
99	Benedetta	assassin	52.38	0.72	49.46	2020
100	Brody	marksman	51.57	0.71	60.05	2020
101	Natalia	assassin	47.76	0.41	61.29	2016
102	Silvanna	fighter	52.65	3.06	9.59	2019

[103 rows x 6 columns]

Gambar 2. Dataset yang diperbarui

- Dataset baru memiliki jumlah kolom yang lebih sedikit yaitu 6 dibandingkan dataset lama yaitu 19 kolom.

II. Analisis statistik

- Berikut ini nilai statistik pada dataset MLBB_Kompetitif:

```
print (databaru2.describe())
```

	win_rate	pick_rate	ban_rate	release_year
count	103.000000	103.000000	103.000000	103.000000
mean	50.183398	0.971068	4.461262	2017.660194
std	2.341717	0.794118	11.575990	1.397150
min	45.000000	0.050000	0.020000	2016.000000
25%	48.500000	0.350000	0.080000	2016.000000
50%	50.100000	0.750000	0.360000	2017.000000
75%	51.745000	1.430000	2.280000	2019.000000
max	56.490000	3.710000	61.290000	2021.000000

Gambar 3. Nilai statistik dataset MLBB_Kompetitif

III. Nilai yang hilang

- Tidak ditemukan nilai yang hilang pada dataset ini sehingga tidak perlu dilakukan pengisian nilai kosong.

```
Pengecekan data baru
```

```
print("jumlah baris dan kolom")
print(databaru2.shape)
print(sorted(dataori.isna().sum())) #mengecek nilai Kosong
```

```
[79]
```

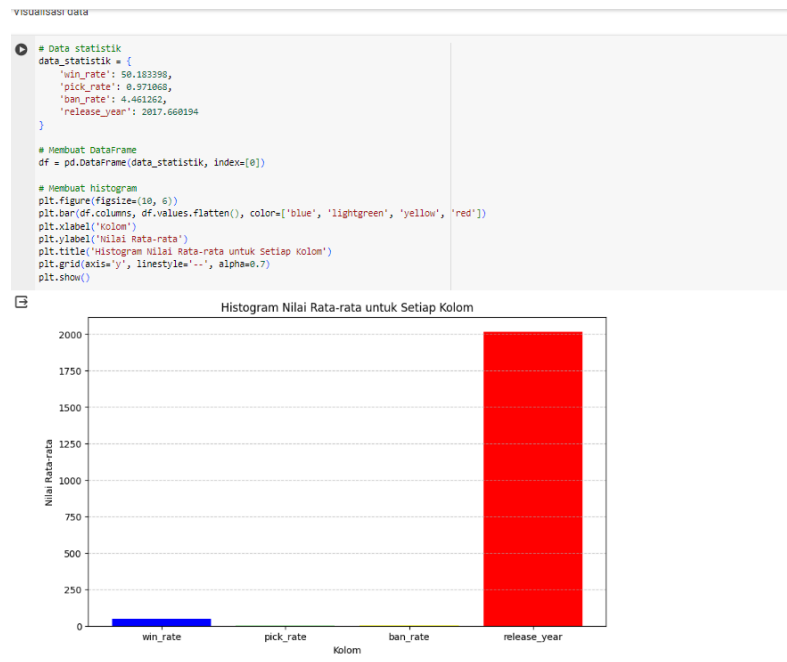
```
... jumlah baris dan kolom
(105, 6)
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

Gambar 4. Pengecekan data baru (MLBB_Kompetitif)

IV. Visualisasi data



Gambar 5. Visualisasi data untuk jumlah role hero



Gambar 6. Visualisasi data untuk nilai rata rata untuk setiap kolom numerik

- Berdasarkan Gambar 6, terdapat nilai outlier (Nilai yang terlalu besar atau mendominasi dalam data)

V. Feature Scaling

- Feature scaling adalah proses mengurangi rentang nilai suatu data. Hal ini bertujuan untuk menghilangkan dominasi suatu atribut yang mempengaruhi analisis data dan membuat analisis data menjadi lebih mudah dipahami. Feature scaling yang digunakan pada dataset MLBB_kompetitif adalah MinMax Scaling.

Feature Scaling Min Max

```

from sklearn.preprocessing import MinMaxScaler

data_Featurescaling = databaru2 # Data yang akan di-scaling
# Memilih fitur numerik yang akan di-scaling
numerical_features = ['win_rate', 'pick_rate', 'ban_rate', 'release_year']
# Min-Max Scaling
scaler = MinMaxScaler()
data_Featurescaling[numerical_features] = scaler.fit_transform(data_Featurescaling[numerical_features])
print(data_Featurescaling)

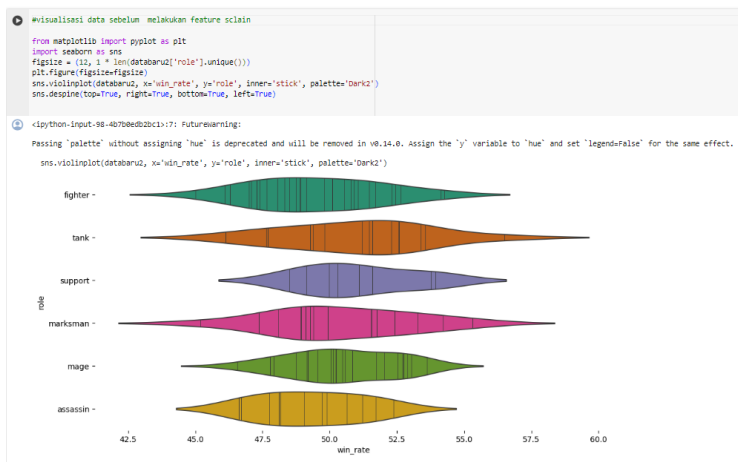
```

	nama_hero	role	win_rate	pick_rate	ban_rate	release_year
0	Terizla	fighter	0.174064	0.027322	0.000000	0.6
1	Martis	fighter	0.000000	0.027322	0.000000	0.4
2	Grock	tank	0.234117	0.024590	0.000000	0.2
3	Carmilla	support	0.762402	0.008197	0.000000	0.8
4	Irithel	marksman	0.567450	0.090164	0.000163	0.2
..
98	Esmeralda	mage	0.360313	0.327869	0.689897	0.6
99	Benedetta	assassin	0.642298	0.183060	0.806920	0.8
100	Brody	marksman	0.571802	0.180328	0.979762	0.8
101	Natalia	assassin	0.240209	0.098361	1.000000	0.0
102	Silvana	fighter	0.665796	0.822404	0.156194	0.6

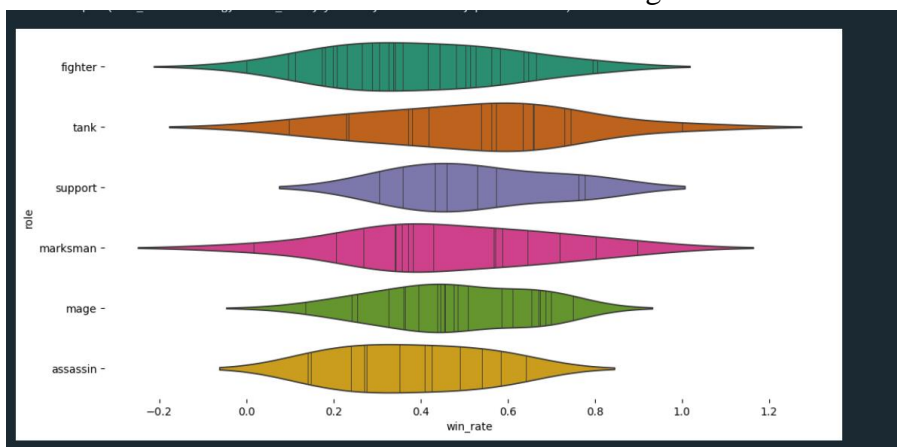
[103 rows x 6 columns]

Gambar 7. Feature Scaling min max

- Berdasarkan Gambar 7. Data telah disamakan dari rentang 0 hingga 1.



Gambar 8. Visualisasi data sebelum feature scaling



Gambar 9. Visualisasi data setelah feature scaling

- Berdasarkan Gambar 9 dan 8, skala win_rate telah diperkecil sehingga feature scaling yang dilakukan telah berhasil