

# **DIAGNOSTIC ANALYTICS REPORT**

Disusun oleh :

Andira Yulianengtias 5026211038



**FINAL EXAM**

**Analitika Data dan Diagnostik**

**2023**

INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
FACULTY OF INTELLIGENT ELECTRICAL AND INFORMATICS TECHNOLOGY  
Department of Information Systems  
Bachelor of Computer Science in Information Systems Program

## Daftar Pustaka

Tentang Data .....	3
Deskripsi variabel .....	3
Tujuan .....	4
Summary Data.....	4
Finding Patterns .....	5
Identifying Relationship .....	7
Detecting Outliers:.....	9

## Tentang Data

Data yang digunakan dalam analitika diagnostik pada laporan ini adalah data Stunting Daerah yang diambil dari website Satu Data Indonesia <https://data.go.id/dashboard-stunting-daerah>. Provinsi yang digunakan adalah Jawa Timur, dimana terdapat data stunting anak pada 38 kabupaten/kota periode waktu 2021-2023. Terdapat 6 kolom pada table yaitu tahun, nama kota/kabupaten, jumlah balita, stunting pendek (anak), stunting sangat pendek (anak), dan prevalensi (%).

## Deskripsi variabel

- Stunting Pendek (Anak) :

Anak yang memiliki tinggi badan yang lebih rendah dari tinggi badan rata-rata yang diharapkan untuk usia mereka. Dimana hal ini menunjukkan bahwa pertumbuhan fisik anak tersebut terhambat, namun tidak dalam tingkat yang sangat parah.

- Stunting Pendek (Anak) :

Anak-anak yang memiliki tinggi badan yang sangat rendah, dan menunjukkan tingkat hambatan pertumbuhan yang lebih parah. Stunting sangat pendek dapat dianggap sebagai tingkat stunting yang lebih serius, dan kondisi ini dapat memberikan dampak kesehatan dan perkembangan yang lebih signifikan pada anak-anak.

- Prevalensi (%) :

Presentase prevalensi stunting dihitung berdasarkan persentase jumlah anak yang mengalami stunting dibandingkan dengan total jumlah balita dalam suatu kabupaten/kota pada suatu periode waktu tertentu.

$$\text{Prevalensi Stunting} = \left( \frac{\text{Jumlah Balita Stunting}}{\text{Total Jumlah Balita}} \right) \times 100\%$$

## Tujuan

Melihat korelasi antara jumlah balita dengan jumlah stunting anak menggunakan regresi

Variabel dependent (Y) :

- Stunting Pendek (Anak)
- Stunting Sangat Pendek (Anak)
- Prevalensi (%)

Variabel independent (X) :

- Jumlah Balita

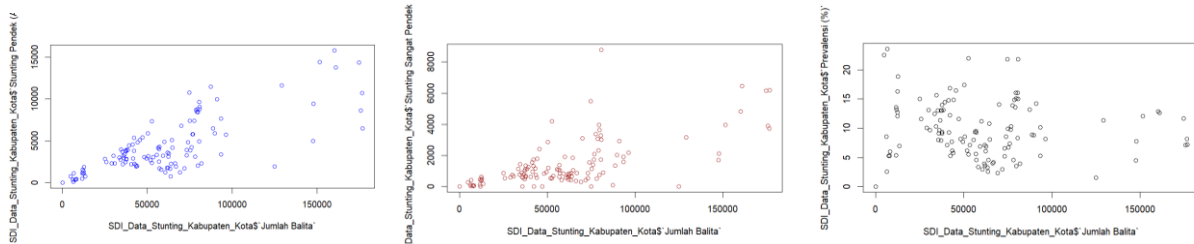
## Summary Data

Tahun		Kabupaten/Kota	Jumlah Balita
Min.	:2020-01-01 00:00:00.00	Length:122	Min. : 0
1st Qu.	:2021-01-01 00:00:00.00	Class :character	1st Qu.: 37252
Median	:2022-01-01 00:00:00.00	Mode :character	Median : 56959
Mean	:2021-10-21 01:34:25.56		Mean : 60666
3rd Qu.	:2023-01-01 00:00:00.00		3rd Qu.: 76617
Max.	:2023-01-01 00:00:00.00		Max. :176742
Stunting Pendek (Anak)		Stunting Sangat Pendek (Anak)	Prevalensi (%)
Min.	: 0	Min. : 0	Min. : 0.000
1st Qu.	: 2146	1st Qu.: 543	1st Qu.: 6.200
Median	: 3294	Median : 884	Median : 9.200
Mean	: 4292	Mean :1431	Mean : 9.762
3rd Qu.	: 5422	3rd Qu.:1729	3rd Qu.:13.000
Max.	:15807	Max. :8789	Max. :23.600

## Finding Patterns

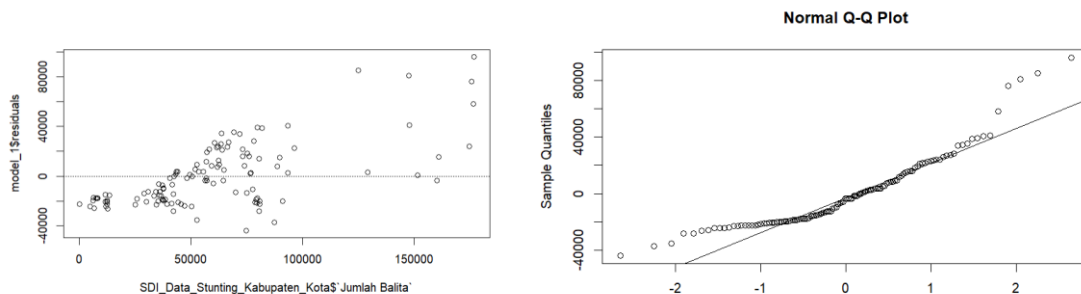
Hubungan antara jumlah balita dengan 3 variabel dependent lain

- **Biru** : korelasi jumlah balita dengan stunting pendek (anak)
- **Merah** : korelasi jumlah balita dengan stunting sangat pendek (anak)
- **Hitam** : korelasi jumlah balita dengan prevalensi (%)



### 1. Pola data jumlah balita dengan stunting pendek (anak)

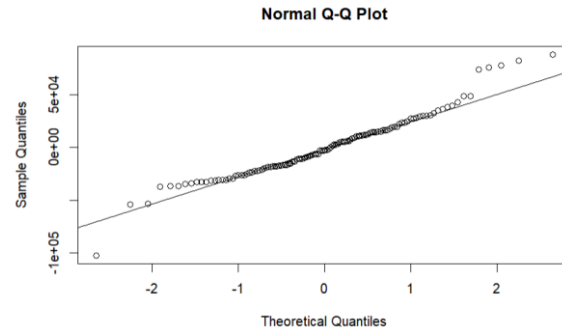
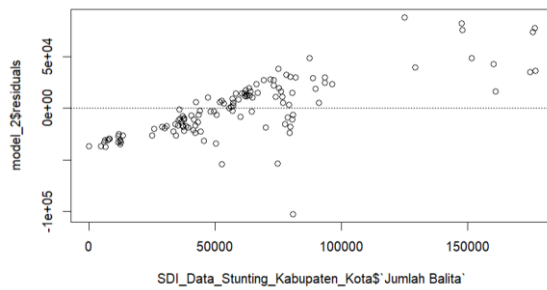
Pada penggunaan scatter plot di atas (**berwarna biru**), menunjukkan bahwa korelasi pola datanya adalah tren naik. Dibawah ini digambarkan juga dalam residual plot dan QQ plot.



Pada Residual plot, residu data terdistribusi dan tidak menunjukkan pola tertentu, yang menunjukkan bahwa model mungkin sesuai dengan data. Juga pada QQ plot dapat dilihat bahwa titik nya mengikuti garis diagonal dengan baik, sehingga menunjukkan bahwa residunya mendekati distribusi normal. Sedangkan di ujung-nya ekor itu menunjukkan keberadaan outlier.

### 2. Pola data jumlah balita dengan stunting sangat pendek (anak)

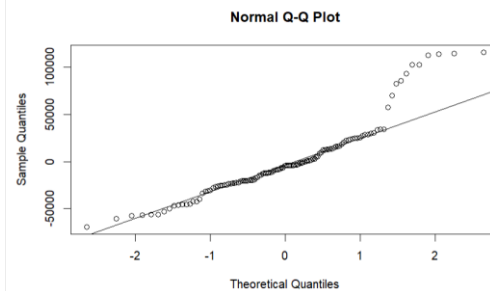
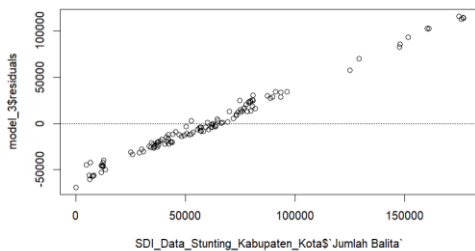
Pada penggunaan scatter plot di atas (**berwarna merah**), kurang menunjukkan pola, namun masih terdapat pola tren naik. Dibawah ini digambarkan juga dalam residual plot dan QQ plot.



Pada Residual plot, residu data sedikit menunjukkan pola tertentu yaitu garis diagonal, yang menunjukkan bahwa model kurang menangkap pola dalam data. Juga pada QQ plot dapat dilihat bahwa titik nya mengikuti garis diagonal dengan baik, sehingga menunjukkan bahwa residunya mendekati distribusi normal. Sedangkan di ujung-nya ekor itu menunjukkan keberadaan outlier.

### 3. Pola data jumlah balita dengan prevalensi (%)

Pada penggunaan scatter plot di atas (berwarna hitam), tidak menunjukkan pola, yang berarti sebaran datanya tidak konsisten. Dibawah ini digambarkan juga dalam residual plot dan QQ plot.



Pada Residual plot, residu data pola tertentu yaitu garis diagonal, yang menunjukkan bahwa model tidak menangkap pola dalam data. Juga pada QQ plot dapat dilihat bahwa titik nya mengikuti garis diagonal dengan baik, sehingga menunjukkan bahwa residunya mendekati distribusi normal. Sedangkan di ujung-nya ekor itu menunjukkan keberadaan outlier.

## Identifying Relationship

Untuk mengidentifikasi hubungan antar variabel, digunakan model regresi linear berganda, dan didapatkan hasil seperti dibawah ini:

```
Call:
lm(formula = `Jumlah Balita` ~ `Stunting Pendek (Anak)` + `Stunting Sangat Pendek (Anak)` +
  `Prevalensi (%)`, data = SDI_Data_Stunting_Kabupaten_Kota)

Residuals:
    Min       1Q   Median       3Q      Max
-56306  -6413  -1231    3716   58840

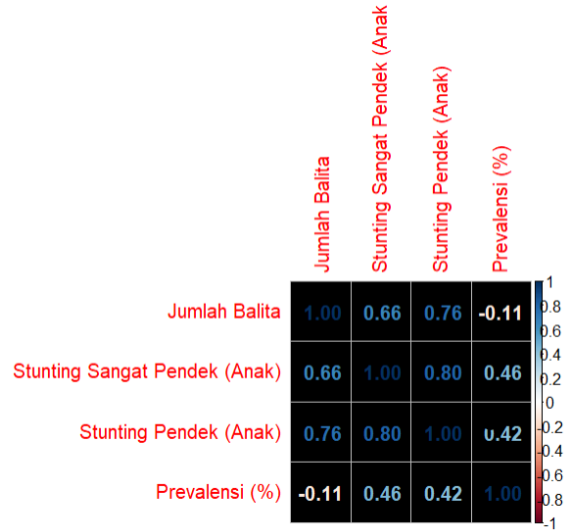
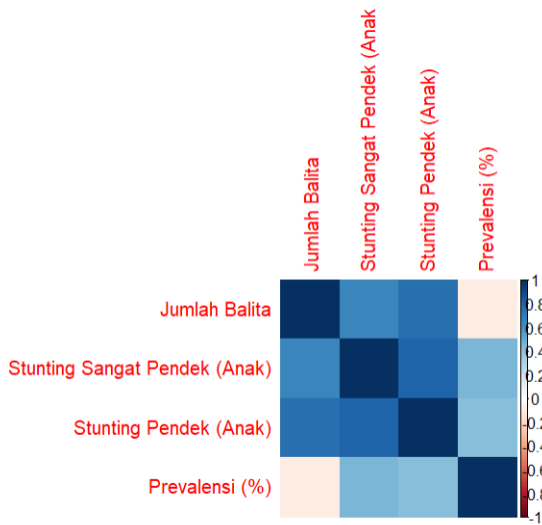
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      56306.4039   3465.6830   16.247 < 2e-16 ***
`Stunting Pendek (Anak)`      8.6350     0.7463   11.570 < 2e-16 ***
`Stunting Sangat Pendek (Anak)`  8.4534     1.6331    5.176 9.4e-07 ***
`Prevalensi (%)`     -4588.8983    342.3896  -13.403 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15990 on 118 degrees of freedom
Multiple R-squared:  0.8329,    Adjusted R-squared:  0.8287
F-statistic: 196.1 on 3 and 118 DF,  p-value: < 2.2e-16
```

- Coefficient :  
Stunting Pendek (Anak), dan Stunting Sangat Pendek (Anak) memiliki koefisien positif yang menunjukkan bahwa kedua variabel ini berkorelasi positif dengan jumlah balita, sedangkan prevalensi memiliki koefisien negative yang menunjukkan bahwa prevalensi memiliki korelasi negative dengan jumlah balita. Artinya semakin tinggi jumlah balita, prevalensi stuntingnya cenderung lebih rendah.
- Multiple R-Squared :  
Nilai R-squarednya adalah 0,8329 dimana hal ini menunjukkan bahwa sekitar 83,29% variasi dapat dijelaskan oleh 3 variabel dependent lain
- p-value :  
Semua variabel memiliki nilai p-value yang sangat kecil yaitu mendekati 0, hal ini menunjukkan semua variabel berpengaruh.

Karena masih belum jelas, variabel mana yang paling baik untuk menjadi predictor, maka digunakan corrplot heatmap untuk melihat hubungan korelasi antar variabel. Berikut adalah korelasi variabel jumlah balita dengan variabel lain.

	Jumlah Balita Stunting Sangat Pendek (Anak)	
Jumlah Balita	1.0000000	0.6569323
Stunting Sangat Pendek (Anak)	0.6569323	1.0000000
Stunting Pendek (Anak)	0.7556550	0.8002844
Prevalensi (%)	-0.1094775	0.4566239
	Stunting Pendek (Anak) Prevalensi (%)	
Jumlah Balita	0.7556550	-0.1094775
Stunting Sangat Pendek (Anak)	0.8002844	0.4566239
Stunting Pendek (Anak)	1.0000000	0.4227053
Prevalensi (%)	0.4227053	1.0000000



Dari Heatmap diatas didapatkan korelasi tertinggi yaitu Jumlah balita dengan Stunting Pendek (Anak), dengan nilai 0,76. Sehingga banyaknya jumlah balita juga mempengaruhi banyanya jumlah stunting pendek (anak) pada data.



## Detecting Outliers:

Karena variabel yang memiliki korelasi dengan Jumlah Balita adalah Stunting Pendek (Anak) , sehingga saya memilih untuk mencari outlier pada korelasi pada dua variabel ini :

Pertama, mencari angka korelasinya , dan didapatkan hasil yaitu : 0,75565

```
##{r}
correlation <- cor(SDI_Data_Stunting_Kabupaten_Kota$`Jumlah Balita`,
SDI_Data_Stunting_Kabupaten_Kota$`Stunting Pendek (Anak)`)
```

correlation	0.755654979971816
-------------	-------------------

Kemudian mencari, data mana yang merupakan outliers pada data stunting anak di Jawa Timur menggunakan kode di bawah ini :

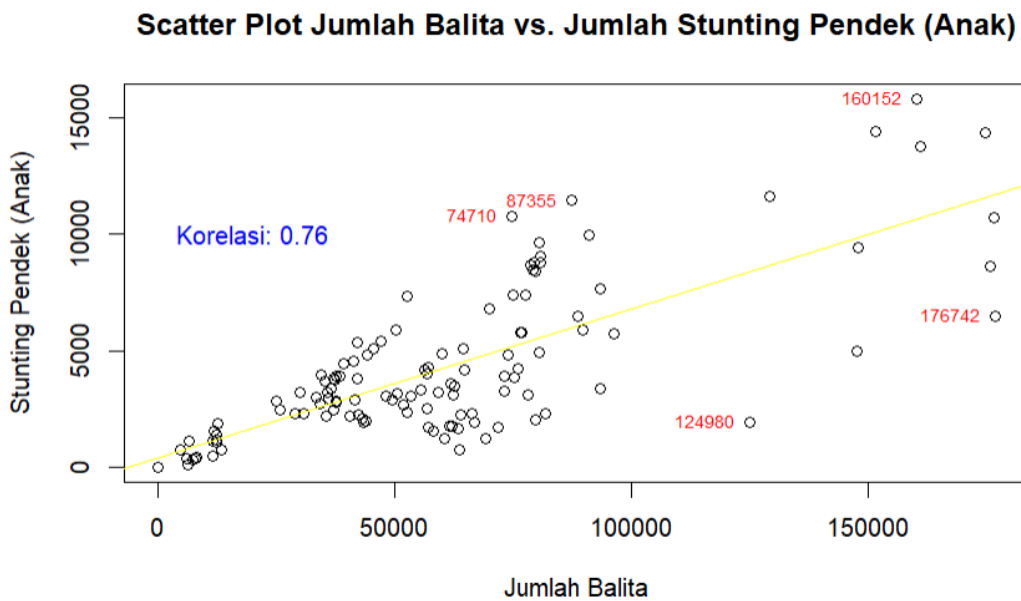
```
##{r}
plot(SDI_Data_Stunting_Kabupaten_Kota$`Jumlah Balita`,
SDI_Data_Stunting_Kabupaten_Kota$`Stunting Pendek (Anak)` ,
      main = "Scatter Plot Jumlah Balita vs. Jumlah Stunting Pendek (Anak)",
      xlab = "Jumlah Balita",
      ylab = "Stunting Pendek (Anak)")

#Menambahkan garis tren
abline(lm(SDI_Data_Stunting_Kabupaten_Kota$`Stunting Pendek (Anak)` ~
SDI_Data_Stunting_Kabupaten_Kota$`Jumlah Balita`), col = "yellow")

#Menampilkan besar korelasi
text(20000, 10000, labels = paste("Korelasi:", round(correlation, 2)), col = "blue")

# Menambahkan teks untuk variabel yang merupakan outlier
text(SDI_Data_Stunting_Kabupaten_Kota$`Jumlah Balita`,
SDI_Data_Stunting_Kabupaten_Kota$`Stunting Pendek (Anak)` ,
      labels = ifelse(abs(residuals) > outlier_limit,
as.character(SDI_Data_Stunting_Kabupaten_Kota$`Jumlah Balita`), ""),
      pos = 2, col = "red", cex = 0.7)
```

Terlihat bahwa korelasi ini memiliki pola tren positif naik (pada garis kuning), dimana semakin banyak jumlah balita maka semakin besar juga angka anak yang mengalami stunting pendek



Sehingga dapat dilihat pada gambar, bahwa outliers nya ada pada nilai Jumlah Balita : 74710, 87355, 124980, 160152, 176742.

Outliers ini terjadi pada :

1. Kabupaten/Kota Pasuruan tahun 2020 , dimana jumlah balitanya berada pada range sedang namun memiliki jumlah anak stunting pendek yang paling tinggi setelah kediri

Tahun	Kabupaten/Kota	Jumlah Balita	Stunting Pendek (Anak)	Stunting Sangat Pendek (Anak)	Prevalensi (%)
01-01-2020 00:00:00	Pasuruan	74710	10789	5493	21.79999924

2. Kota Kediri tahun 2020, dimana jumlah balitanya berada pada range sedang namun memiliki jumlah anak stunting pendek yang paling tinggi

Tahun	Kabupaten/Kota	Jumlah Balita	Stunting Pendek (Anak)	Stunting Sangat Pendek (Anak)	Prevalensi (%)
01-01-2020 00:00:00	Kediri	87355	11438	112	13.19999981

3. Kota Surabaya tahun 2023, dimana jumlah balitanya termasuk ke dalam range tinggi namun memiliki jumlah anak stunting yang sangat rendah

Tahun	Kabupaten/Kota	Jumlah Balita	Stunting Pendek (Anak)	Stunting Sangat Pendek (Anak)	Prevalensi (%)
01-01-2023 00:00:00	Kota Surabaya	124980	1936	0	1.5

4. Kabupaten/Kota Jember tahun 2020, dimana jumlah balitanya termasuk ke dalam range tinggi namun memiliki jumlah anak stunting yang paling tinggi dibanding kabupaten/kota lain yang berada pada jumlah balita tinggi

Tahun	Kabupaten/Kota	Jumlah Balita	Stunting Pendek (Anak)	Stunting Sangat Pendek (Anak)	Prevalensi (%)
01-01-2020 00:00:00	Jember	160152	15807	4815	12.89999962

5. Kota Surabaya tahun 2021, dimana memiliki jumlah balita paling tinggi namun memiliki jumlah anak stunting pendek yang rendah

Tahun	Kabupaten/Kota	Jumlah Balita	Stunting Pendek (Anak)	Stunting Sangat Pendek (Anak)	Prevalensi (%)
01-01-2021 00:00:00	Kota Surabaya	176742	6508	6205	7.199999809