

Assignment Test - Data Scientist

File lebih lengkapnya dapat diakses melalui file htm yang ada di [repository ini](#)

1. Hasil evaluasi kelengkapan data

a. Farms.csv

i. Raw data

Dataset statistics

Number of variables	4
Number of observations	551
Missing cells	165
Missing cells (%)	7.5%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	17.3 KiB
Average record size in memory	32.2 B

Variable types

Numeric	1
Categorical	2
Text	1

ii. After Cleansing

Dataset statistics

Number of variables	4
Number of observations	458
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	17.9 KiB
Average record size in memory	40.0 B

Variable types

Numeric	1
Categorical	2
Text	1

b. Ponds.csv

i. Raw Data

Dataset statistics

Number of variables	10
Number of observations	338
Missing cells	347
Missing cells (%)	10.3%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	26.5 KiB
Average record size in memory	80.4 B

Variable types

Numeric	6
DateTime	2
Text	1
Categorical	1

ii. After Cleansing

Dataset statistics

Number of variables	5
Number of observations	238
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	11.2 KiB
Average record size in memory	48.0 B

Variable types

Numeric	5
---------	---

c. **Cycles.csv**

i. Raw Data

Dataset statistics

Number of variables	24
Number of observations	2617
Missing cells	4983
Missing cells (%)	7.9%
Duplicate rows	49
Duplicate rows (%)	1.9%
Total size in memory	490.8 KiB
Average record size in memory	192.1 B

Variable types

Numeric	12
Categorical	4
DateTime	5
Text	3

ii. After Cleansing

Dataset statistics

Number of variables	6
Number of observations	2616
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	84
Duplicate rows (%)	3.2%
Total size in memory	143.1 KiB
Average record size in memory	56.0 B

Variable types

Numeric	4
DateTime	2

d. Feeds.csv

i. Raw Data

Dataset statistics

Number of variables	3
Number of observations	706908
Missing cells	42
Missing cells (%)	< 0.1%
Duplicate rows	21010
Duplicate rows (%)	3.0%
Total size in memory	16.2 MiB
Average record size in memory	24.0 B

Variable types

Numeric	2
DateTime	1

ii. After Cleansing

Dataset statistics

Number of variables	3
Number of observations	706887
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	21006
Duplicate rows (%)	3.0%
Total size in memory	21.6 MiB
Average record size in memory	32.0 B

Variable types

Numeric	2
DateTime	1

e. Fasting.csv

i. Raw Data

Dataset statistics

Number of variables	3
Number of observations	32379
Missing cells	40
Missing cells (%)	< 0.1%
Duplicate rows	150
Duplicate rows (%)	0.5%
Total size in memory	759.0 KiB
Average record size in memory	24.0 B

Variable types

DateTime	1
Numeric	1
Categorical	1

ii. After Cleansing

Dataset statistics

Number of variables	3
Number of observations	32339
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	148
Duplicate rows (%)	0.5%
Total size in memory	1010.6 KiB
Average record size in memory	32.0 B

Variable types

DateTime	1
Numeric	1
Categorical	1

f. Harvests.csv

i. Raw Data

Dataset statistics

Number of variables	9
Number of observations	8087
Missing cells	2056
Missing cells (%)	2.8%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	568.7 KiB
Average record size in memory	72.0 B

Variable types

Numeric	5
DateTime	3
Categorical	1

ii. After Cleansing

Dataset statistics

Number of variables	5
Number of observations	7824
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	106
Duplicate rows (%)	1.4%
Total size in memory	366.8 KiB
Average record size in memory	48.0 B

Variable types

Numeric	3
DateTime	1
Categorical	1

g. Sampling.csv

i. Raw Data

Dataset statistics

Number of variables	7
Number of observations	15032
Missing cells	13693
Missing cells (%)	13.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	822.2 KiB
Average record size in memory	56.0 B

Variable types

Numeric	3
DateTime	3
Text	1

ii. After Cleansing

Dataset statistics

Number of variables	3
Number of observations	15032
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	272
Duplicate rows (%)	1.8%
Total size in memory	352.4 KiB
Average record size in memory	24.0 B

Variable types

Numeric	2
DateTime	1

h. Measurements.csv

i. Raw Data

Dataset statistics

Number of variables	24
Number of observations	139050
Missing cells	2087077
Missing cells (%)	62.5%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	25.5 MiB
Average record size in memory	192.0 B

Variable types

Numeric	23
DateTime	1

ii. After Cleansing

Dataset statistics

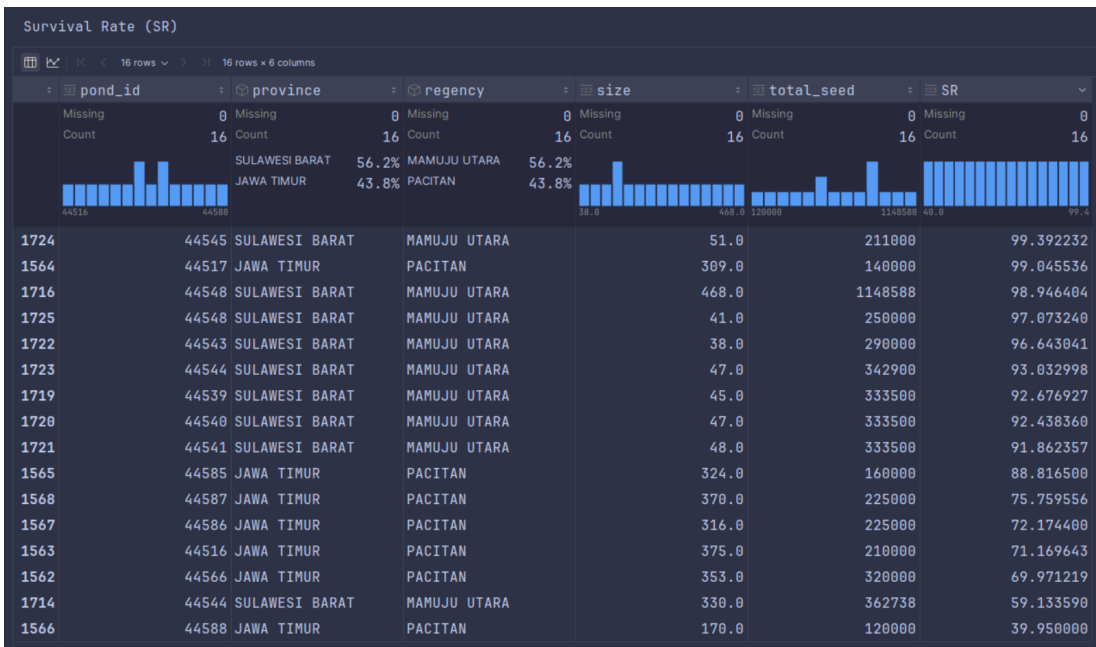
Number of variables	22
Number of observations	139050
Missing cells	1948193
Missing cells (%)	63.7%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	23.3 MiB
Average record size in memory	176.0 B

Variable types

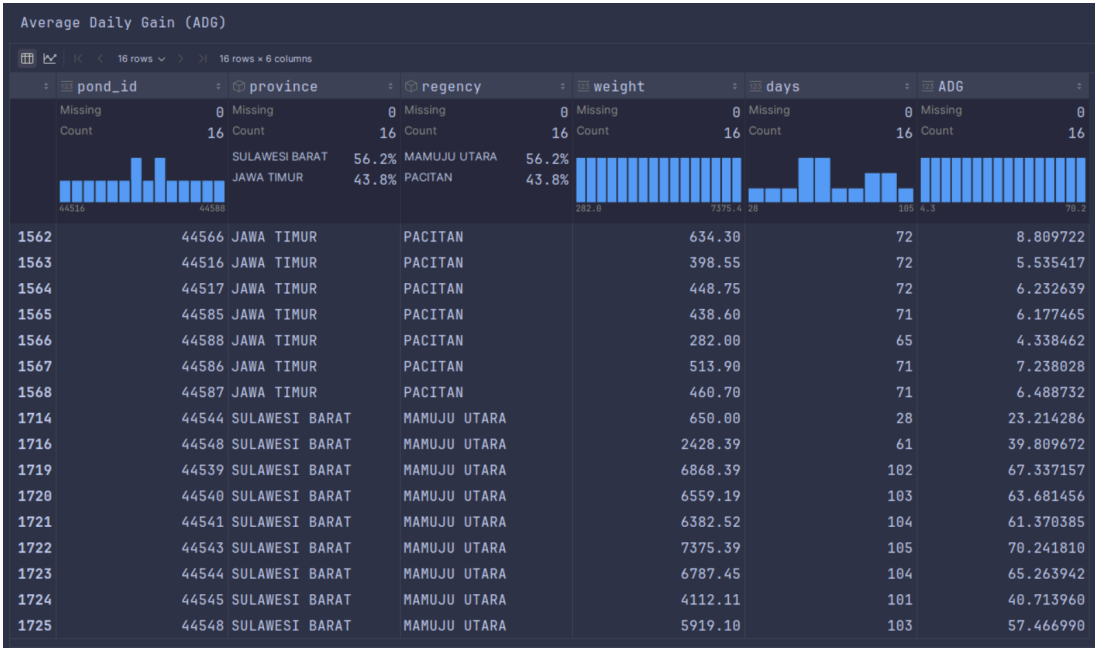
Numeric	21
DateTime	1

2. Hitung Survival Rate (SR) dan Average Daily Gain (ADG)

a. SR



b. ADG



3. Hasil Prediksi Model

	cycle_id ▾	predicted_sr ▾	predicted_abw ▾	predicted_biomass ▾	predicted_revenue ▾
1	24775.0	74.9	3.56	1003903.32	14252089339917.49
2	17624.0	78.78	6.18	326778.93	6245285628216.62
3	17405.0	84.54	3.92	1521652.58	21864856468821.16
4	18751.0	53.91	35.98	342896.85	7153267436760.2
5	17879.0	64.37	3.23	159928.53	1055568821696.09
6	17719.0	73.62	4.52	695934.45	8827582231519.71
7	9097.0	93.87	6.69	2346666.13	37507944193338.19
8	23317.0	57.8	5.63	190959.67	3894217475001.14
9	21280.0	68.73	3.52	53883985.05	3928098130524.28
10	12067.0	36.74	5.46	480874.33	2731870309101.48
11	27656.0	89.51	3.7	9464376.92	139777832392116.66
12	27413.0	90.0	17.06	1012540.45	61712078601193.88
13	19294.0	89.09	1.55	2502154.15	45290921598217.01
14	27556.0	99.98	3.9	3499165.0	1079769861.54
15	18360.0	31.35	0.07	5091.52	4073794788.61
16	27414.0	43.4	6.22	488275.29	22359298437829.07
17	19615.0	91.0	4.79	2911892.96	86934153772248.5
18	26178.0	59.91	58.95	215772.69	31726863621165.98
19	15768.0	39.59	1.85	170531.14	2599603180291.9
20	21280.0	68.73	3.52	53883985.05	3928098130524.28
21	27559.0	92.83	8.42	421246.36	10248328412893.49
22	9665.0	61.25	0.41	26533.73	44715309616.0
23	12131.0	43.67	2.96	70925.84	2975518225544.01
24	27903.0	66.91	1.7	138225.05	1306839647879.55
25	23737.0	89.72	15.76	1021783.03	8326588654118.62
26	26684.0	85.11	21.6	16544780.8	17885274701115.38

4. Kesimpulan dan Penjelasan Terkait Fitur/Variabel yang Penting

Tabel hasil menunjukkan pentingnya setiap fitur untuk memprediksi variabel target yang berbeda: Survival Rate (SR), Average Body Weight (ABW), Biomass, dan Revenue. Nilai pentingnya menunjukkan seberapa besar kontribusi setiap fitur terhadap prediksi model. Berikut adalah penjelasan rinci dari hasil tersebut:

- **Pentingnya Fitur untuk Survival Rate (SR)**

- total_seed: Fitur ini memiliki kepentingan tertinggi (0.318583) untuk memprediksi Survival Rate. Ini menunjukkan bahwa jumlah benih yang digunakan dalam siklus sangat mempengaruhi tingkat kelangsungan hidup.
- pond_id: Fitur kedua terpenting (0.255979) untuk SR, menunjukkan bahwa kolam spesifik tempat siklus berlangsung juga memainkan peran penting.
- cycle_id: Fitur ini (0.247982) juga penting, menunjukkan bahwa pengenalan unik untuk setiap siklus memiliki dampak signifikan pada tingkat kelangsungan hidup.
- days: Jumlah hari (0.175927) siklus berlangsung juga merupakan faktor yang signifikan.
- farm_id, regency, province: Fitur-fitur ini memiliki kepentingan minimal, menunjukkan bahwa lokasi dan pengenalan peternakan memiliki sedikit dampak pada tingkat kelangsungan hidup.

Pentingnya Fitur untuk Average Body Weight (ABW)

- pond_id: Fitur ini memiliki kepentingan tertinggi (0.388858) untuk memprediksi ABW, menunjukkan bahwa kondisi kolam sangat mempengaruhi berat badan rata-rata ikan.
- cycle_id: Fitur kedua terpenting (0.296254) untuk ABW, menunjukkan bahwa siklus spesifik memiliki dampak yang cukup besar.
- total_seed: Fitur ini (0.168109) juga memainkan peran signifikan dalam menentukan berat badan rata-rata.
- days: Jumlah hari (0.145545) siklus berlangsung juga penting.
- regency: Fitur ini (0.001227) memiliki kepentingan yang sangat rendah, menunjukkan bahwa lokasi regional memiliki dampak minimal pada ABW.
- farm_id, province: Fitur-fitur ini memiliki kepentingan yang dapat diabaikan untuk memprediksi ABW.

Pentingnya Fitur untuk Biomass

- total_seed: Fitur ini memiliki kepentingan yang sangat tinggi (0.9948448) untuk memprediksi Biomass, menunjukkan bahwa jumlah benih adalah faktor paling kritis.
- pond_id: Fitur ini (0.004020129) memiliki kepentingan yang sangat rendah, menunjukkan bahwa kondisi kolam memiliki dampak minimal pada biomassa.
- cycle_id, days: Fitur-fitur ini memiliki kepentingan yang dapat diabaikan untuk memprediksi biomassa.
- farm_id, regency, province: Fitur-fitur ini hampir tidak memiliki dampak pada biomassa, seperti yang ditunjukkan oleh nilai kepentingannya yang hampir nol.

Pentingnya Fitur untuk Revenue

- **total_seed**: Fitur ini memiliki kepentingan tertinggi (0.419901) untuk memprediksi Revenue, menunjukkan bahwa jumlah benih sangat mempengaruhi pendapatan yang dihasilkan.
- **cycle_id**: Fitur kedua terpenting (0.242432) untuk pendapatan, menunjukkan bahwa siklus spesifik memiliki dampak yang cukup besar.
- **pond_id**: Fitur ini (0.191069) juga memainkan peran signifikan dalam menentukan pendapatan.
- **days**: Jumlah hari (0.146158) siklus berlangsung juga penting.
- **regency, province, farm_id**: Fitur-fitur ini memiliki kepentingan minimal, menunjukkan bahwa lokasi dan pengenalan peternak memiliki sedikit dampak pada pendapatan.

Ringkasan

Secara keseluruhan, **total_seed** adalah fitur yang paling berpengaruh untuk memprediksi Biomass dan Revenue, sementara **pond_id** dan **cycle_id** juga penting untuk memprediksi SR dan ABW. Fitur-fitur lain seperti **farm_id**, **regency**, dan **province** memiliki dampak minimal pada prediksi. Analisis ini membantu dalam memahami faktor-faktor mana yang paling kritis untuk mengoptimalkan hasil dalam siklus akuakultur.

5. Rekomendasi Mengenai Budidaya Udang Perlu Dilakukan

Tingkatkan Jumlah Benih:

Fitur **total_seed** memiliki kepentingan tertinggi untuk memprediksi Biomass dan Revenue. Meningkatkan jumlah benih yang digunakan dalam setiap siklus dapat secara signifikan meningkatkan biomassa dan pendapatan.

Optimalkan Kondisi Kolam:

Fitur **pond_id** sangat penting untuk memprediksi Average Body Weight (ABW) dan Survival Rate (SR). Pastikan kondisi kolam optimal, termasuk kualitas air, suhu, dan aerasi, untuk meningkatkan kesehatan dan pertumbuhan udang secara keseluruhan.

Pantau dan Sesuaikan Durasi Siklus:

Fitur **cycle_id** dan **days** penting untuk memprediksi SR, ABW, dan Revenue. Pantau secara teratur durasi setiap siklus dan lakukan penyesuaian berdasarkan pertumbuhan dan kesehatan udang untuk memaksimalkan tingkat kelangsungan hidup dan pendapatan.

Analisis Data Secara Berkala:

Kumpulkan dan analisis data secara terus-menerus untuk mengidentifikasi tren dan membuat keputusan berdasarkan data. Gunakan wawasan ini untuk menyempurnakan praktik budidaya dan meningkatkan hasil.

Dengan fokus pada area kunci ini, petani udang dapat mengoptimalkan operasi mereka untuk mencapai tingkat kelangsungan hidup yang lebih baik, berat badan rata-rata yang lebih tinggi, peningkatan biomassa, dan pendapatan yang lebih besar.