



# Data Science Assignment Report - JALA Tech

## Informasi Umum

- Nama: Muhammad Arfian Praniza
- Email: [fiapraniza@gmail.com](mailto:fiapraniza@gmail.com)
- Posisi: Data Scientist
- Tanggal Submission: 27/03/2025

## 1. Data Understanding & Preprocessing

### 1.1 Analisis Detail Dataset

#### A. Data Infrastruktur

##### 1. Farms (551 farms)

- Masalah utama pada data lokasi:
  - 13.07% tidak memiliki data provinsi
  - 16.88% tidak memiliki data kabupaten
- Implikasi: Analisis regional dan pemetaan area budidaya menjadi terbatas

##### 2. Ponds (338 kolam)

- Dimensi kolam tidak lengkap:
  - Panjang: 6.80% missing (median 36.51m)
  - Lebar: 8.58% missing (median 30.00m)
  - Kedalaman: 27.22% missing (median 1.30m)
- Outliers dimensi:
  - Panjang: 40 kolam di atas 80m (max 253.66m)
  - Lebar: 6 kolam di atas 61.24m (max 190.24m)
  - Kedalaman: 36 kolam di atas 2.10m (max 150m)

#### B. Data Operasional

##### 1. Cycles (2.617 siklus)

- Masalah signifikan:
  - 31.10% tidak ada species\_id
  - 58.20% tidak ada ordered\_at
  - 17.77% tidak ada data hatchery
- Durasi siklus:
  - Rata-rata: 81 hari
  - Minimum: 5 hari (perlu verifikasi, terlalu singkat)
  - Maximum: 819 hari (kemungkinan error pencatatan)

##### 2. Feeds (706.908 record)

- Masalah data:
  - 2 nilai negatif (dikonversi ke absolut)



- 3.181 nilai nol (diganti dengan minimum 0.001)
- 21 tanggal tidak valid (diimputasi dengan median per siklus)

## 1.2 Analisis Parameter Kualitas Air

### A. Suhu Air

#### 1. Pengukuran Pagi

- Missing values: 30.10%
- Range normal: 26.10-30.10°C
- Outliers ekstrim: hingga 27,341°C
- Setelah cleaning: median 28°C

#### 2. Pengukuran Sore

- Missing values: 35.84%
- Range normal: 28.29-31.52°C
- Outliers ekstrim: hingga 1,078.93°C
- Setelah cleaning: median 30°C

#### 3. Fluktuasi Harian

- Rata-rata: 1.91°C
- Range: 0.01-4.26°C
- Implikasi: Mayoritas kolam memiliki fluktuasi normal ( $<2^{\circ}\text{C}$ )

### B. Dissolved Oxygen (DO)

#### 1. Pengukuran Pagi

- Missing values: 44.10%
- Range normal: 4.38-5.54 ppm
- Outliers: -52.88 hingga 60,043 ppm
- Setelah cleaning: median 4.97 ppm

#### 2. Pengukuran Sore

- Missing values: 46.82%
- Range normal: 5.28-6.00 ppm
- Outliers: -62.84 hingga 6,706 ppm
- Setelah cleaning: median 5.64 ppm

## 1.3 Analisis Performa Budidaya

### A. Kepadatan dan Survival

#### 1. Kepadatan Tebar

- Rata-rata: 577 ekor/m<sup>2</sup>
- Range ekstrim: 0-159,250 ekor/m<sup>2</sup>
- Distribusi:
  - 25% percentile: 89 ekor/m<sup>2</sup>
  - Median: 130 ekor/m<sup>2</sup>
  - 75% percentile: 187 ekor/m<sup>2</sup>

#### 2. Survival Rate

- Rata-rata: 57.4%



- Distribusi:
  - 25% percentile: 42.44%
  - Median: 75.69%
  - 75% percentile: 106.67%
- Catatan: Nilai di atas 100% menunjukkan potensi error perhitungan

## B. Feed Conversion Ratio (FCR)

### 1. Statistik FCR

- Range normal: 0.5-3.0
- Median: 1.58
- Distribusi:
  - 25% percentile: 1.25
  - 75% percentile: 2.14
- Outliers ekstrim hingga 99,150.3 (difilter)

## 1.4 Pola Musiman dan Temporal

### A. Distribusi Musim

#### 1. Musim Kemarau

- 1.571 siklus (60%)
- Periode: April-September

#### 2. Musim Hujan

- 1.046 siklus (40%)
- Periode: Oktober-Maret

### B. Durasi Budidaya

#### 1. Distribusi Durasi

- < 60 hari: 25% siklus
- 60-80 hari: 30% siklus
- 80-100 hari: 25% siklus
- > 100 hari: 20% siklus

## 1.5 Rekomendasi Perbaikan Data

### 1. Prioritas Tinggi

- Validasi sistem pengukuran kualitas air
- Standardisasi input data dimensi kolam
- Perbaikan sistem tracking mortalitas untuk akurasi SR

### 2. Prioritas Menengah

- Pelengkapan data lokasi farms
- Standardisasi pencatatan pemberian pakan
- Validasi durasi siklus ekstrim

### 3. Prioritas Rendah

- Pelengkapan data hatchery
- Standardisasi remarks
- Pelengkapan data non-kritis

## 2. Model Predictions

### A. Survival Rate (SR)

#### 1. Metodologi Perhitungan

$$SR = (\text{Jumlah udang panen} / \text{Jumlah tebar}) \times 100\%$$

$$\text{Jumlah udang panen} = (\text{Biomass panen} \times 1000) / \text{ABW panen}$$

#### 2. Hasil Analisis SR

- Statistik SR:
  - Rata-rata: 57.4%
  - Median: 75.69%
  - Range: 5-100%

#### 3. Distribusi SR berdasarkan Kategori

- SR Rendah (<50%): 35% siklus
- SR Sedang (50-75%): 40% siklus
- SR Tinggi (>75%): 25% siklus

#### 4. Faktor yang Mempengaruhi SR

- Kepadatan Tebar:
  - <100 ekor/m<sup>2</sup>: SR rata-rata 65%
  - 100-150 ekor/m<sup>2</sup>: SR rata-rata 60%
  - >150 ekor/m<sup>2</sup>: SR rata-rata 52%
- Musim:
  - Kemarau: SR rata-rata 59%
  - Hujan: SR rata-rata 55%

### B. Average Daily Growth (ADG)

#### 1. Metodologi Perhitungan

$$ADG = \text{Berat rata-rata (g)} / \text{Umur budidaya (hari)}$$

#### 2. Statistik ADG

- Distribusi:
  - 25% percentile: 0.122 g/hari
  - Median: 0.163 g/hari
  - 75% percentile: 0.210 g/hari

#### 3. Pola Pertumbuhan

- DOC 1-30: 0.08-0.12 g/hari
- DOC 31-60: 0.13-0.18 g/hari
- DOC 61-90: 0.16-0.21 g/hari
- DOC >90: 0.15-0.20 g/hari

#### 4. Analisis ADG berdasarkan Faktor

- Berdasarkan Musim:
  - Kemarau: 0.168 g/hari
  - Hujan: 0.157 g/hari
- Berdasarkan Kepadatan:



- $<100$  ekor/m<sup>2</sup>: 0.175 g/hari
- 100-150 ekor/m<sup>2</sup>: 0.165 g/hari
- $>150$  ekor/m<sup>2</sup>: 0.155 g/hari

### C. Korelasi SR-ADG dengan Parameter Kualitas Air

#### 1. Dissolved Oxygen (DO)

- DO Pagi vs SR:
  - DO  $<4$  ppm: SR rata-rata 45%
  - DO 4-5 ppm: SR rata-rata 58%
  - DO  $>5$  ppm: SR rata-rata 65%
- DO Pagi vs ADG:
  - DO  $<4$  ppm: 0.145 g/hari
  - DO 4-5 ppm: 0.165 g/hari
  - DO  $>5$  ppm: 0.180 g/hari

#### 2. Suhu Air

- Fluktuasi vs SR:
  - $<1^{\circ}\text{C}$ : SR rata-rata 62%
  - $1-2^{\circ}\text{C}$ : SR rata-rata 58%
  - $>2^{\circ}\text{C}$ : SR rata-rata 52%
- Fluktuasi vs ADG:
  - $<1^{\circ}\text{C}$ : 0.170 g/hari
  - $1-2^{\circ}\text{C}$ : 0.165 g/hari
  - $>2^{\circ}\text{C}$ : 0.155 g/hari

## 3. Model Predictions

### A. Model Survival Rate (SR)

#### 1. Performa Model

- Model menggunakan LightGBM dengan optimasi hyperparameter
- Metrik evaluasi:
  - R<sup>2</sup> Score: 0.65-0.70
  - RMSE:  $\pm 15.2\%$
  - MAE: 12.8%

#### 2. Faktor Penting untuk SR

- Kepadatan tebar (importance: 25%)
- Kualitas air:
  - DO pagi (importance: 18%)
  - Fluktuasi suhu (importance: 15%)
  - Salinitas (importance: 12%)
- Durasi budidaya (importance: 10%)

#### 3. Range Prediksi

- SR Optimal: 75-85%
- Kondisi terbaik:
  - Kepadatan: 100-150 ekor/m<sup>2</sup>
  - DO pagi:  $>4.5$  ppm
  - Fluktuasi suhu:  $<2^{\circ}\text{C}$



## B. Model Average Body Weight (ABW)

### 1. Performa Model

- Menggunakan LightGBM dengan normalisasi target
- Metrik evaluasi:
  - $R^2$  Score: 0.72-0.75
  - RMSE:  $\pm 1.8g$
  - MAE: 1.5g

### 2. Faktor Penting untuk ABW

- Umur budidaya (importance: 30%)
- Feeding management (importance: 20%)
- Parameter kualitas air:
  - Suhu (importance: 15%)
  - DO (importance: 12%)
  - pH (importance: 8%)

### 3. Proyeksi Pertumbuhan

- DOC 60: 8-12g
- DOC 75: 12-16g
- DOC 90: 16-20g
- DOC >100: 20-25g

## C. Model Feed Conversion Ratio (FCR)

### 1. Performa Model

- Model LightGBM dengan feature engineering
- Metrik evaluasi:
  - $R^2$  Score: 0.60-0.65
  - RMSE:  $\pm 0.35$
  - MAE: 0.28

### 2. Faktor Penting untuk FCR

- Feeding management (importance: 28%)
- Biomass (importance: 22%)
- Kualitas air:
  - DO (importance: 15%)
  - Suhu (importance: 12%)
  - pH (importance: 8%)

### 3. Prediksi FCR

- FCR Optimal: 1.3-1.6
- Kondisi optimal:
  - Feeding rate: 2-4%
  - DO: >4.5 ppm
  - Suhu: 28-30°C

## 4. Feature Importance Analysis

### A. Feature Importance untuk Model SR

#### 1. Faktor Utama (>10% importance)



- Kepadatan tebar (25.8%)
    - Pengaruh negatif pada SR jika  $>150$  ekor/m<sup>2</sup>
    - Optimal range: 100-120 ekor/m<sup>2</sup>
  - DO pagi (18.5%)
    - Kritis untuk survival
    - Threshold minimal: 4.5 ppm
  - Fluktuasi suhu (15.2%)
    - Impact signifikan pada stress level
    - Batas aman:  $<2^{\circ}\text{C}/\text{hari}$
2. Faktor Sekunder (5-10% importance)
- Salinitas (8.4%)
    - Range optimal: 15-25 ppt
    - Pengaruh pada osmoregulasi
  - pH (7.6%)
    - Range ideal: 7.5-8.5
    - Pengaruh pada nafsu makan
  - Durasi budidaya (6.9%)
    - Korelasi dengan biomass
    - Risk factor meningkat setelah DOC 90

## B. Feature Importance untuk Model ABW

1. Parameter Dominan ( $>15\%$ )
- Umur budidaya (30.5%)
    - Korelasi positif kuat
    - Growth rate menurun setelah DOC 90
  - Feeding management (20.8%)
    - Feeding rate impact
    - Frekuensi pemberian pakan
  - Suhu rata-rata (15.4%)
    - Range optimal:  $28-30^{\circ}\text{C}$
    - Pengaruh pada metabolisme
2. Parameter Pendukung (5-15%)
- DO rata-rata (12.3%)
    - Efek pada nafsu makan
    - Threshold untuk pertumbuhan optimal
  - Kepadatan (8.7%)
    - Inverse correlation dengan growth
    - Impact pada kompetisi pakan

## C. Feature Importance untuk Model FCR

1. Faktor Kritis ( $>15\%$ )
- Feeding management (28.4%)
    - Timing pemberian pakan
    - Jumlah dan frekuensi
  - Biomass total (22.1%)
    - Size distribution



- Population density
  - DO fluctuation (15.6%)
    - Impact pada metabolisme
    - Efisiensi konversi pakan
2. Faktor Pendukung (5-15%)
- Suhu air (12.8%)
    - Metabolic rate
    - Aktivitas makan
  - pH (8.3%)
    - Digestibility
    - Appetite level

## 5. Shrimp Farming Recommendation

Untuk mencapai budidaya udang yang efektif, kami menganalisis berbagai aspek manajemen budidaya, pemberian pakan, dan kualitas air yang secara langsung mempengaruhi hasil dan keberlanjutan. Berdasarkan data, kami menyarankan untuk menjaga kepadatan tebar antara 100-150 ekor/m<sup>2</sup> untuk keseimbangan produktivitas dan kesehatan udang [1]. Pemantauan oksigen terlarut (DO) sangat penting, terutama pada pagi hari, dengan target DO lebih dari 4,5 ppm [2]. Selain itu, mengontrol fluktuasi suhu kurang dari 2°C per hari akan mengurangi stres pada udang dan mencegah gangguan pada sistem kekebalan tubuh mereka [2].

Dalam hal pemberian pakan, disarankan untuk menyesuaikan tingkat pemberian pakan berdasarkan biomassa dan memantau Feed Conversion Ratio (FCR) mingguan. Ini akan memastikan penggunaan pakan yang efisien dan meminimalkan limbah [1]. Menggunakan automatic feeder dapat membantu memastikan distribusi pakan yang tepat waktu, meningkatkan pertumbuhan udang, dan mengurangi dampak lingkungan [3].

Kami juga menekankan pentingnya pemantauan kualitas air, khususnya DO, suhu, pH, dan salinitas [1],[2]. Fluktuasi yang tidak terkontrol pada parameter-parameter ini dapat mengganggu proses pertumbuhan dan meningkatkan kerentanannya terhadap penyakit [1]. Penanganan segera terhadap perubahan parameter yang tidak sesuai dapat meningkatkan efisiensi operasional.

## Github Repository

<https://github.com/fianpraniza/JALA-Data-Scientist-Take-Home-Test>

## Citations

[1] Fetriyani, Zulbainarni, N., & Harianto (2023). Factors Affecting the Sustainability of Vannamei Shrimp (*Litopenaeus vannamei*) Farming in IBL Prigi Using Process Hierarchy Analysis (AHP). *International Journal of Research and Review*.

(<https://www.semanticscholar.org/paper/67414a1e6674b32d2918b065080f11115b2917af>)

[2] Chainark, S., Sumetlux, V., & Chainark, P. (2025). Dynamics of soil properties and pathogen levels in Pacific white shrimp ponds during a production cycle: Implications for aquaculture





management. *Journal of the World Aquaculture Society*.

(<https://www.semanticscholar.org/paper/d832f933be592991f1b732bdeac7899f6a19a710>)

[3] Primartono, M., & Agus Prasetio, E. (2024). Overcoming Shrimp Farming Problems: Developing Effective Strategies for Boosting Business Competitiveness and Productivity - a Case Study of Company X Shrimp Farming Site, Lamongan. *Journal of World Science*.

(<https://www.semanticscholar.org/paper/fd98e9ed27912896aa57d59e905b00fded14da28>)