

**Peta Jalan Kedaulatan Data Nasional: Analisis Kesenjangan Infrastruktur  
dan Klasterisasi Negara Global Menggunakan Pendekatan Data Science  
dalam Mendukung SDG 9 dan 13**



Jaenal Arifin(19240216)  
Erwin Gouw(19240336 )  
Rianti(19240211)

Alexhandra Dea Alfani(19241661)  
Belinda(19240446)  
Amelia Ulya Nashifa(19240192)

Afriza(19240293)

**PROGRAM STUDI SISTEM INFORMASI KAMPUS PONTIANAK  
BINA SARANA INFORMATIKA  
PONTIANAK  
2025**

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang Masalah

Ekosistem ekonomi digital Indonesia terus menunjukkan pertumbuhan yang resilien di tengah ketidakpastian global. Laporan terbaru *e-Cconomy SEA 2024* mencatat bahwa nilai *Gross Merchandise Value* (GMV) ekonomi digital Indonesia telah mencapai **\$90 miliar** pada tahun 2024, tumbuh 13% dibanding tahun sebelumnya, dan diproyeksikan akan terus meningkat signifikan hingga 2030 (Google, Temasek, & Bain, 2024). Pertumbuhan ini didukung oleh fondasi pengguna internet yang semakin kuat. Survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2025 melaporkan bahwa tingkat penetrasi internet nasional telah menembus angka **80,66%**, atau setara dengan 229 juta jiwa (APJII, 2025). Ledakan populasi digital ini menciptakan permintaan eksponensial terhadap infrastruktur penyimpanan dan pemrosesan data yang andal (*Data Center*).

Namun, ekspansi infrastruktur digital ini berhadapan langsung dengan krisis energi global. Pusat data kini menjadi salah satu konsumen listrik terbesar di dunia, terutama didorong oleh beban kerja *Artificial Intelligence* (AI). Laporan khusus *International Energy Agency* (IEA, 2025) memprediksi bahwa konsumsi listrik dari pusat data global akan berlipat ganda mencapai **945 TWh** pada tahun 2030. Lonjakan ini menjadikan efisiensi energi bukan lagi sekadar opsi, melainkan syarat mutlak operasional. Studi terbaru oleh Chidolue et al. (2024) dalam *Engineering Science & Technology Journal* menekankan bahwa transisi menuju *Green Data Center* melalui teknologi pendinginan canggih dan integrasi energi terbarukan adalah kunci untuk mencapai keberlanjutan infrastruktur IT sesuai target SDG 13.

Dalam konteks persaingan regional, Indonesia menghadapi tantangan *gap* infrastruktur yang serius. Meskipun memiliki basis pengguna internet terbesar di ASEAN, kapasitas daya (*power capacity*) pusat data di Indonesia masih tertinggal dibandingkan Singapura yang telah mapan sebagai *hub* regional. Analisis kesenjangan ini menjadi rumit karena data infrastruktur global seringkali tidak terstruktur. Pendekatan konvensional tidak lagi memadai untuk memetakan dinamika pasar yang bergerak cepat. Oleh karena itu, diperlukan implementasi *Business Intelligence* (BI) yang mampu mengolah *Big Data* menjadi wawasan strategis secara *real-time* untuk pengambilan keputusan yang presisi (Ranjan & Foropon, 2021).

Penelitian ini mengusulkan pengembangan *dashboard* analitik interaktif berbasis Python yang mengintegrasikan data infrastruktur global terbaru (2025). Sistem ini dirancang untuk memvisualisasikan posisi kompetitif Indonesia secara *head-to-head* dengan negara *Global Hub* dan *Regional Hub*, serta mengevaluasi kesiapan transisi energi hijau nasional. Kebaruan (*novelty*)

penelitian ini terletak pada penggunaan data terkini pasca-lonjakan AI dan visualisasi multidimensi yang menggabungkan parameter kapasitas teknis dengan indikator keberlanjutan lingkungan.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana posisi komparatif kapasitas daya (*power capacity*) Pusat Data Indonesia dibandingkan dengan *Regional Hub* (Singapura) dan *Global Leader* (Amerika Serikat) pasca-akselerasi AI tahun 2025?
2. Apakah ketersediaan infrastruktur fisik pusat data di Indonesia sudah sebanding dengan tingkat penetrasi internet yang telah mencapai 80% (analisis *supply vs demand*)?
3. Sejauh mana kesiapan infrastruktur pusat data Indonesia dalam mengadopsi energi terbarukan (*Renewable Energy*) dibandingkan tren global untuk memenuhi target SDG 13?
4. Bagaimana pemetaan kelompok negara (*clustering*) berdasarkan kemiripan karakteristik infrastruktur digital menggunakan pendekatan *Machine Learning*?

## 1.3 Batasan Masalah

Agar penelitian lebih terarah, diberikan batasan masalah sebagai berikut:

1. **Sumber Data:** Dataset sekunder global tahun **2024–2025** yang telah divalidasi, mencakup variabel jumlah fasilitas, kapasitas MW, penetrasi internet, dan bauran energi.
2. **Metode:** Pengembangan sistem menggunakan bahasa Python dengan pustaka Pandas
3. (pengolahan data) dan Plotly/Streamlit (visualisasi interaktif).
4. **Lingkup:** Analisis difokuskan pada aspek fisik infrastruktur (Kapasitas & Kualitas Tier) dan lingkungan (Energi Hijau), tidak mencakup aspek keamanan siber teknis.

## 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun *dashboard* visualisasi data yang mampu memetakan daya saing infrastruktur digital Indonesia secara komprehensif.
2. Mengidentifikasi *Opportunity Gap* antara tingginya permintaan digital nasional dengan ketersediaan infrastruktur fisik.
3. Memberikan rekomendasi strategis berbasis data (*data-driven*) kepada pemerintah dan industri untuk percepatan pembangunan *Green Data Center*.
4. Menerapkan algoritma *K-Means Clustering* untuk melakukan segmentasi negara secara otomatis guna memvalidasi posisi strategis Indonesia secara statistik.

## 1.5 Manfaat Penelitian

1. Manfaat Akademis:

- a. Memperkaya literatur terkini mengenai penerapan Data Science dalam pemetaan infrastruktur digital strategis.
- 2. Manfaat Praktis:
  - a. Menjadi rujukan valid bagi regulator dalam merancang insentif investasi hijau dan bagi pelaku industri dalam menentukan strategi ekspansi kapasitas.

## BAB II

### TINJAUAN PUSTAKA

#### **2.1 Infrastruktur Digital dan Ketahanan Ekonomi**

Dalam lanskap ekonomi digital modern, Pusat Data (*Data Center*) bukan lagi sekadar fasilitas penyimpan data, melainkan aset strategis nasional. Laporan *e-Economy SEA 2024* menegaskan bahwa ketersediaan infrastruktur digital yang tangguh (*resilient*) berkorelasi positif dengan percepatan nilai ekonomi digital di Asia Tenggara, di mana Indonesia diproyeksikan tetap memimpin pasar dengan pertumbuhan GMV dua digit (Google, Temasek, & Bain, 2024).

Untuk mengukur kualitas infrastruktur ini, standar klasifikasi industri terus berevolusi. Uptime Institute (2024) dalam *Global Data Center Survey* terbaru menekankan bahwa di era *Artificial Intelligence* (AI), standar **Tier III** (*Concurrently Maintainable*) menjadi batas minimum operasional untuk menjamin ketersediaan layanan (*uptime*) 99,982% di tengah beban kerja komputasi yang semakin berat. Penelitian ini menggunakan distribusi Tier sebagai variabel proksi untuk mengukur kesiapan infrastruktur Indonesia dalam menangani beban kerja kritis dibandingkan negara maju.

#### **2.2 Dampak Generative AI dan Transformasi *Green Data Center***

Lonjakan adopsi *Generative AI* telah mengubah profil konsumsi energi pusat data secara drastis. *International Energy Agency* (IEA, 2025) memproyeksikan konsumsi listrik pusat data global akan mencapai rekor tertinggi pada 2026, didorong oleh kebutuhan pendinginan server AI yang intensif. Hal ini menempatkan target **SDG 13 (Penanganan Perubahan Iklim)** sebagai prioritas utama industri.

Konsep *Green Data Center* kini bergeser dari sekadar efisiensi daya (PUE) menjadi keberlanjutan menyeluruh. Chidolue et al. (2024) mendefinisikan *Next-Gen Green DC* sebagai fasilitas yang wajib mengadopsi teknologi pendinginan cair (*liquid cooling*) dan memiliki kontrak pembelian energi terbarukan (*Renewable Power Purchase Agreements*) minimal 50%. Kesenjangan adopsi teknologi ini antara negara berkembang dan negara maju menjadi fokus analisis dalam penelitian ini untuk memetakan urgensi transisi energi di Indonesia.

#### **2.3 Business Intelligence untuk Pengambilan Keputusan Strategis**

Kompleksitas ekosistem data center yang melibatkan variabel teknis (MW), ekonomi (pertumbuhan pasar), dan lingkungan (karbon), memerlukan pendekatan analitik yang canggih. Zhang & Chen (2024) dalam jurnal *Sustainability* menyoroti bahwa penggunaan *Big Data Analytics* dan visualisasi interaktif sangat efektif untuk memetakan disparitas infrastruktur antar-wilayah dan merumuskan kebijakan dekarbonisasi yang presisi.

Sistem *dashboard* yang dikembangkan dalam penelitian ini mengimplementasikan prinsip tersebut dengan mengubah data mentah global menjadi wawasan visual (*visual insights*). Pendekatan ini mengatasi keterbatasan metode konvensional dalam melakukan *benchmarking* multidimensi, memungkinkan identifikasi *opportunity gap* yang lebih akurat bagi pemangku kepentingan.

#### **2.4 Analisis Klasterisasi dengan K-Means**

*Machine Learning* tipe *Unsupervised Learning* digunakan untuk menemukan pola tersembunyi dalam data tanpa label. Algoritma **K-Means Clustering** adalah metode yang mempartisi data ke dalam  $k$  kelompok (klaster) di mana setiap pengamatan termasuk dalam klaster dengan rata-rata terdekat (centroid). Dalam konteks ekonomi, metode ini efektif untuk mengelompokkan negara berdasarkan indikator pembangunan yang kompleks (Jain, 2010).

## BAB III

### METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM

#### 3.1 Kerangka Kerja Penelitian

Penelitian ini menerapkan metodologi pengembangan perangkat lunak berbasis data (*Data-Driven Development*). Alur kerja penelitian dirancang secara sistematis melalui empat tahapan utama, yaitu:

1. **Akuisisi Data:** Pengumpulan data mentah dari sumber tepercaya.
2. **Pra-pemrosesan Data (*Data Preprocessing*):** Pembersihan dan normalisasi data.
3. **Analisis Eksploratif:** Pemahaman karakteristik data dan statistik deskriptif.
4. **Pengembangan Visualisasi Antarmuka (*Dashboarding*):** Implementasi visual ke dalam sistem interaktif.

Pendekatan ini memastikan bahwa data mentah yang heterogen dapat ditransformasi menjadi informasi strategis yang valid dan dapat ditindaklanjuti.

#### 3.2 Sumber Data

Data primer yang digunakan dalam penelitian ini bersumber dari dataset sekunder global terstruktur dengan format CSV (*Comma Separated Values*), yang dikompilasi dari berbagai laporan industri tahun 2024–2025. Dataset ini mencakup profil infrastruktur pusat data dari lebih dari 100 negara dengan atribut variabel sebagai berikut:

1. **Variabel Identitas:** country (Nama Negara).
2. **Variabel Infrastruktur Fisik:** total\_data\_centers (Unit) dan power\_capacity\_MW\_total (Megawatt).
3. **Variabel Pasar:** internet\_penetration\_percent (%) dan growth\_rate\_of\_data\_centers (%).
4. **Variabel Kualitas:** tier\_distribution (Teks deskriptif distribusi sertifikasi Tier I–IV).
5. **Variabel Lingkungan:** average\_renewable\_energy\_usage\_percent (Persentase adopsi energi hijau).

#### 3.3 Teknik Pra-pemrosesan Data (*Data Cleaning*)

Tantangan teknis utama dalam penelitian ini adalah ketidakkonsistenan format data (*dirty data*). Variabel numerik sering bercampur dengan karakter non-numerik (contoh: ~12,000+, Range 20–30, atau Tier III: 60%). Untuk mengatasi hal ini, dikembangkan sebuah mesin pembersih data (*Data Cleaning Engine*) menggunakan bahasa pemrograman Python dengan pustaka Pandas dan modul *Regular Expression* (Regex).

Algoritma pembersihan data yang diimplementasikan meliputi:

## **1. Algoritma Normalisasi Numerik (clean\_numeric)**

Fungsi ini dirancang untuk menangani variasi format angka. Algoritma bekerja dengan cara menghapus karakter non-digit dan menangani nilai rentang (range). Jika ditemukan format "A-B" (contoh: 25000-40000), algoritma akan menghitung nilai rata-rata (mean) untuk mendapatkan skalar tunggal. *Implementasi Logika (Python)*:

```
import re

def clean_numeric(value):    # Hapus koma dan
    spasi   value = str(value).lower().replace(',', " ")
    # Deteksi pola rentang angka (misal: 100-200)
    if '-' in value:
        parts = re.findall(r"\d\.", value)
        if len(parts) >= 2:
            return (float(parts[0]) + float(parts[1])) / 2
        # Deteksi angka tunggal   matches =
        re.findall(r"\d\.", value)   return
        float(matches[0]) if matches else 0
```

## **2. Algoritma Ekstraksi Kualitas Tier (clean\_tier\_score)**

Variabel tier\_distribution berisi teks tidak terstruktur. Algoritma parsing dikembangkan untuk mengekstrak persentase fasilitas yang memiliki sertifikasi Tier III dan Tier IV saja, sebagai indikator ketahanan infrastruktur (resilience).

*Implementasi Logika (Python)*:

```
def clean_tier_score(value):
    # Regex mencari angka persen setelah kata kunci 'III' atau 'IV'
    match3 = re.search(r'III[:\s]*(\d+)%', str(value))   match4 =
    re.search(r'IV[:\s]*(\d+)%', str(value))

    tier3 = float(match3.group(1)) if match3 else 0
    tier4 = float(match4.group(1)) if match4 else 0

    return tier3 + tier4 # Total Skor High Quality
```

### 3.4 Desain Visualisasi dan Antarmuka Sistem

Sistem dibangun menggunakan kerangka kerja *Streamlit* untuk menghasilkan antarmuka web yang interaktif. Strategi visualisasi data dirancang menggunakan pustaka *Plotly* untuk menjawab rumusan masalah melalui empat modul analisis:

1. Modul Skala Infrastruktur (SDG 9):

Menggunakan Grouped Bar Chart dengan skala logaritmik (Log Scale). Penggunaan skala logaritmik krusial untuk memvisualisasikan perbandingan antara negara dengan kapasitas ekstrem (seperti Amerika Serikat) dan negara berkembang (Indonesia) dalam satu bidang pandang yang proporsional.

2. Modul Posisi Pasar (Market Matrix):

Menggunakan Scatter Plot Quadrant Analysis. Sumbu X merepresentasikan Demand (Penetrasi Internet) dan Sumbu Y merepresentasikan Supply (Kapasitas Daya). Visualisasi ini memetakan negara ke dalam kuadran strategis: Mature, Saturated, Lagging, atau Opportunity Gap.

3. Modul Analisis Kompetitif (Radar Chart):

Menggunakan Spider/Radar Chart untuk perbandingan head-to-head multidimensi (5 variabel) antara Indonesia dan negara pembanding (Singapura). Data dinormalisasi ke skala 0-1 untuk perbandingan yang adil.

4. Modul Keberlanjutan (SDG 13):

Menggunakan Lollipop Chart untuk memeringkat negara berdasarkan persentase penggunaan energi terbarukan, memberikan visualisasi yang bersih mengenai kesenjangan transisi energi hijau.

### 3.5 Lingkungan Pengembangan

Spesifikasi perangkat lunak dan pustaka yang digunakan dalam pengembangan sistem adalah sebagai berikut:

1. **Bahasa Pemrograman:** Python 3.10 atau lebih baru.
2. **Manipulasi Data:** Pandas (*Python Data Analysis Library*) v2.0+.
3. **Visualisasi Data:** Plotly Express & Graph Objects v5.0+.
4. **Antarmuka Web:** Streamlit v1.30+.
5. **Pola Pencarian Teks:** Modul re (*Regular Expression*) bawaan Python.

### 3.6 Implementasi Algoritma Machine Learning

Sistem ini mengintegrasikan modul kecerdasan buatan sederhana menggunakan pustaka *Scikit-Learn*. Proses klasterisasi dilakukan dengan tahapan sebagai berikut:

1. **Seleksi Fitur (*Feature Selection*):** Lima variabel kunci dipilih sebagai input model, yaitu: Jumlah DC, Kapasitas Daya, Penetrasi Internet, Laju Pertumbuhan, dan Bauran Energi.
2. **Normalisasi Data (*Feature Scaling*):** Mengingat satuan data yang berbeda (Megawatt vs Persentase), diterapkan algoritma **Standard Scaler** (

$$z = \frac{x - \mu}{\sigma}$$

) agar setiap variabel memiliki bobot yang setara dalam perhitungan jarak.

3. **Pemodelan K-Means:** Algoritma dikonfigurasi untuk membagi negara menjadi 4 klaster strategis:
  1. *Top Tier (Global Hub)*
  2. *Mature Market*
  3. *High Growth / Emerging*
  4. *Early Stage*
4. **Pelabelan Otomatis:** Sistem secara otomatis mendeteksi karakteristik *centroid* (titik tengah) setiap klaster untuk memberikan label yang sesuai.

## BAB IV

# HASIL DAN PEMBAHASAN

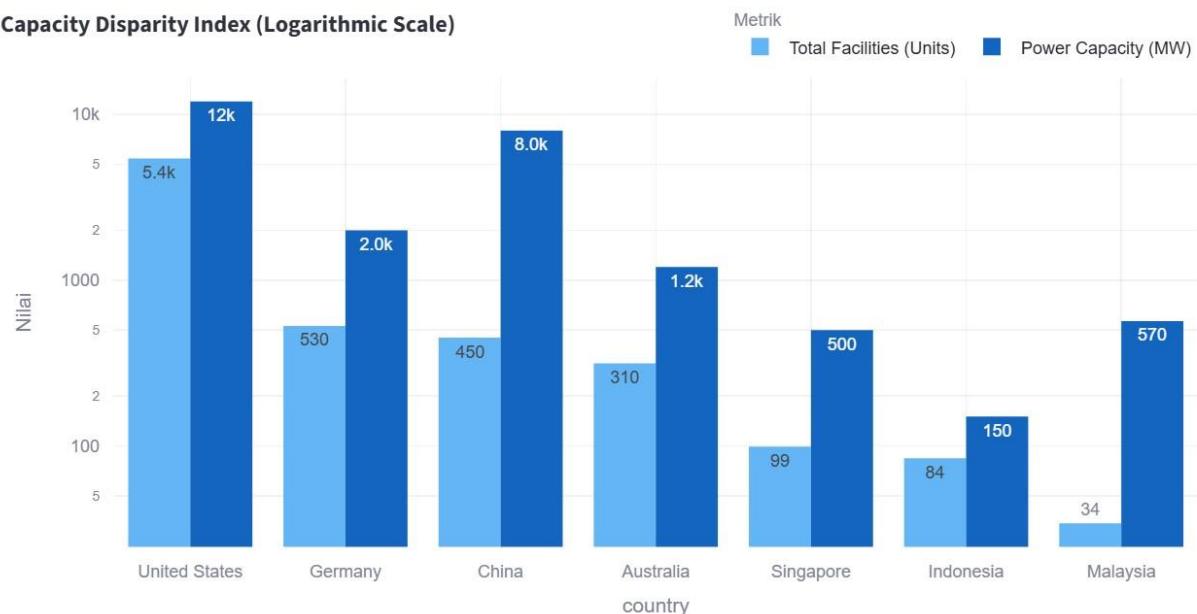
### 4.1 Implementasi Sistem Dashboard

Sistem *Business Intelligence* yang dikembangkan telah berhasil diimplementasikan menggunakan kerangka kerja *Streamlit*. Antarmuka utama (*User Interface*) dirancang dengan pendekatan *user-centric*, menyajikan empat panel analisis strategis: (1) Skala Infrastruktur, (2) Posisi Pasar, (3) Profil Kompetitif, dan (4) Keberlanjutan Lingkungan.

Sistem mampu melakukan pembersihan data (*data cleaning*) secara *real-time* terhadap dataset global yang heterogen, mengubah format teks tidak terstruktur (seperti distribusi Tier dan rentang kapasitas) menjadi variabel numerik yang siap dianalisis. Berikut adalah pembahasan mendalam mengenai temuan data yang dihasilkan oleh sistem.

### 4.2 Analisis Disparitas Kapasitas Infrastruktur (SDG 9)

Berdasarkan visualisasi modul *Infrastructure Scale* (Gambar 4.1), ditemukan adanya disparitas struktural yang signifikan antara Indonesia dan *Regional Hub* (Singapura).



Temuan Data:

Meskipun secara kuantitas fisik jumlah fasilitas pusat data di Indonesia cukup kompetitif (bersaing dengan jumlah fasilitas di negara tetangga), terdapat kesenjangan ekstrem pada total kapasitas daya (Power Capacity).

Pembahasan:

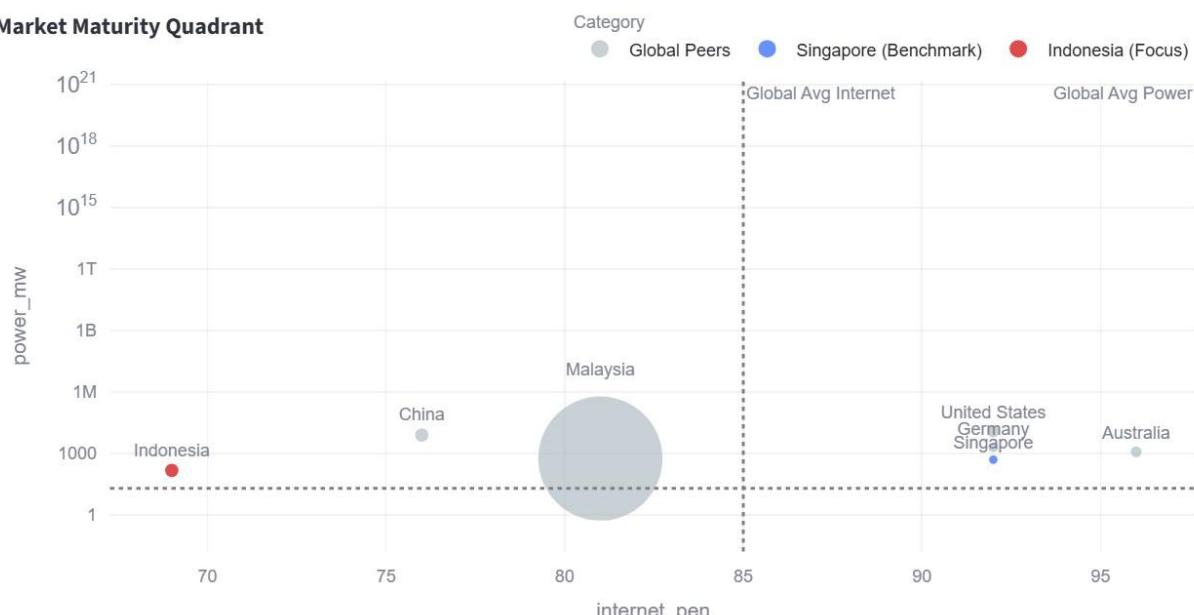
Grafik batang logaritmik menunjukkan bahwa kapasitas daya total Singapura jauh melampaui Indonesia. Fenomena ini mengindikasikan bahwa ekosistem pusat data di Indonesia masih didominasi oleh fasilitas berskala kecil hingga menengah (Retail Colocation), sedangkan Singapura telah didominasi oleh fasilitas berskala raksasa (Hyperscale).

Implikasi:

Indonesia perlu memprioritaskan peningkatan densitas daya per fasilitas, bukan sekadar memperbanyak jumlah gedung, untuk mendukung beban kerja komputasi berat seperti Artificial Intelligence.

#### 4.3 Pemetaan Posisi Pasar: Analisis Kuadran

Modul analisis *Market Matrix* memetakan posisi negara berdasarkan dua sumbu: Penetrasi Internet (Permintaan) dan Kapasitas Infrastruktur (Penawaran). Hasil visualisasi *Scatter Plot* (Gambar 4.2) menempatkan Indonesia pada posisi strategis.



Posisi Kuadran:

Indonesia teridentifikasi berada di Kuadran Kanan Bawah (Opportunity Gap).

**Sumbu X (Demand):** Penetrasi internet tinggi (>79%), setara dengan negara-negara maju.

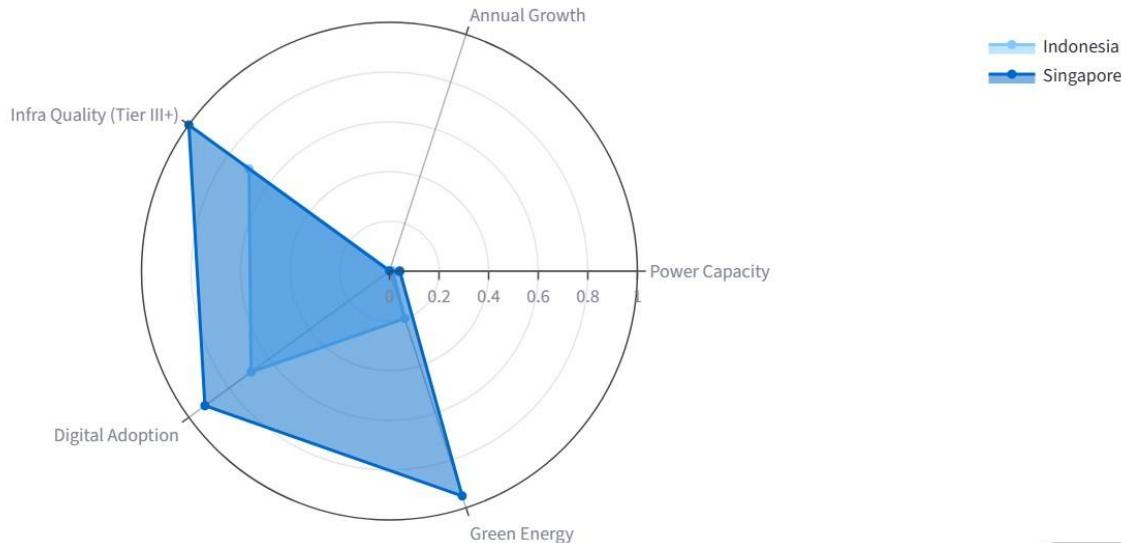
**Sumbu Y (Supply):** Ketersediaan infrastruktur fisik (jumlah dan daya DC) masih relatif rendah.

Pembahasan:

Posisi ini mengkonfirmasi adanya ketimpangan antara konsumsi data masyarakat yang masif dengan ketersediaan infrastruktur lokal. Namun, indikator Growth Rate (dивизуализирован с помощью красного кольца) menunjukkan bahwa Indonesia memiliki laju pertumbuhan tercepat di kawasan ini untuk menutup kesenjangan tersebut.

#### 4.4 Analisis Lanskap Kompetitif (*Head-to-Head*)

Melalui visualisasi *Radar Chart* (Gambar 4.3), dilakukan perbandingan multidimensi antara profil infrastruktur Indonesia dan Singapura.



##### Kekuatan Indonesia:

Indonesia unggul mutlak dalam aspek Laju Pertumbuhan Tahunan (Growth Rate) dan potensi pasar pengguna (Digital Adoption).

##### Kelemahan:

Indonesia tertinggal dalam aspek Kapasitas Daya dan Energi Hijau.

##### Kualitas Infrastruktur:

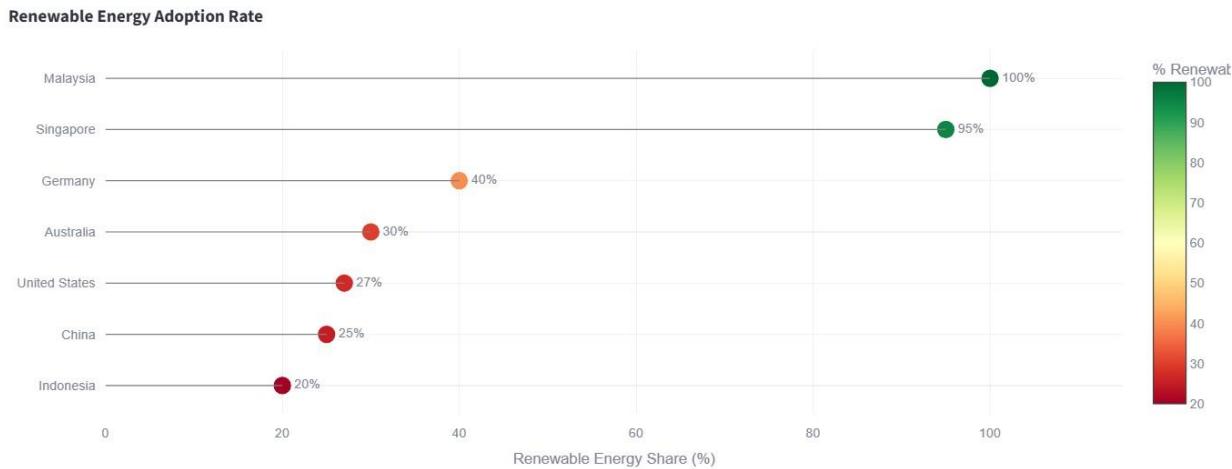
Temuan menarik menunjukkan bahwa dari segi kualitas ketahanan (Resilience), Indonesia sudah cukup kompetitif. Mayoritas fasilitas baru telah memenuhi standar sertifikasi Tier III (Concurrently Maintainable), yang membantah stigma bahwa kualitas infrastruktur negara berkembang selalu rendah. Hal ini sejalan dengan standar industri yang dirujuk oleh Uptime Institute (2024).

#### 4.5 Evaluasi Kesiapan Keberlanjutan Energi (SDG 13)

Modul analisis *Sustainability* (Gambar 4.4) menyoroti tantangan terbesar Indonesia dalam persaingan global.

## Green Energy Transition (SDG 13)

Peringkat negara berdasarkan persentase penggunaan energi terbarukan pada fasilitas Data Center.



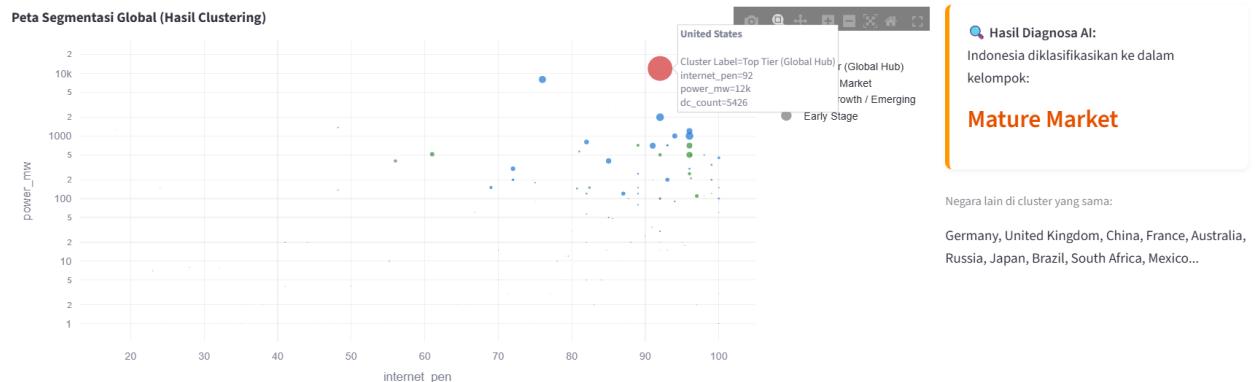
### Temuan Data:

1. Grafik Lollipop Chart menunjukkan bahwa persentase penggunaan energi terbarukan pada pusat data di Indonesia masih berada di kisaran rendah (~20%), jauh tertinggal dibandingkan negara-negara Nordik atau bahkan negara tetangga yang sudah mencapai di atas 50-90%.
2. Analisis Hambatan:  
Berdasarkan data kualitatif yang diekstrak sistem, hambatan utama bukan hanya pada teknologi, melainkan pada ketersediaan pasokan listrik hijau dari jaringan utilitas nasional (PLN) di kawasan industri utama.
3. Urgensi:  
Mengingat investor teknologi global (Hyperscalers) memiliki target internal Net-Zero Emissions, rendahnya bauran energi hijau ini menjadi hambatan non-teknis utama bagi masuknya investasi asing ke Indonesia.

## 4.6 Segmentasi Pasar Berbasis Artificial Intelligence

Berdasarkan hasil komputasi algoritma K-Means, sistem secara otomatis mengklasifikasikan Indonesia ke dalam klaster "High Growth / Emerging".

## ④ Segmentasi Negara dengan Unsupervised Learning (K-Means)



Interpretasi Hasil ML: Algoritma memisahkan Indonesia dari Singapura yang masuk dalam klaster "*Mature Market*".

Validasi Statistik: Hal ini memvalidasi hipotesis awal secara statistik bahwa meskipun pasar Indonesia besar, karakteristik infrastrukturnya masih memiliki kemiripan pola dengan negara berkembang lain (seperti Vietnam atau India), bukan dengan negara maju. Posisi ini menegaskan status Indonesia sebagai pasar dengan potensi pertumbuhan tinggi namun belum matang.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil pengembangan sistem *Business Intelligence Dashboard* dan analisis data mendalam terhadap ekosistem pusat data global, dapat ditarik beberapa kesimpulan strategis sebagai berikut:

##### 1. Keberhasilan Sistem Analitik:

Penelitian ini berhasil mengembangkan perangkat lunak visualisasi berbasis Python yang mampu mengatasi masalah fragmentasi dan ketidakkonsistenan data infrastruktur global. Algoritma pembersihan data (Data Cleaning Engine) berbasis Regular Expression (Regex) terbukti efektif dalam mengekstraksi variabel kualitas (Tier) dan normalisasi data kapasitas, sehingga memungkinkan analisis komparatif yang presisi.

##### 2. Status "Opportunity Gap" (Kesenjangan Peluang):

Visualisasi Market Matrix menempatkan Indonesia pada kuadran strategis High Demand - Low Supply. Tingkat penetrasi internet Indonesia yang telah mencapai level maturitas (setara negara maju) belum diimbangi oleh kapasitas infrastruktur fisik yang memadai. Hal ini mengindikasikan adanya risiko latensi layanan digital nasional, namun sekaligus menandakan peluang investasi yang sangat besar bagi pengembangan fasilitas baru.

##### 3. Disparitas Kualitas Infrastruktur:

Terdapat perbedaan karakteristik yang mencolok antara Indonesia dan Singapura. Singapura telah mencapai tahap Hyperscale Hub dengan densitas daya yang sangat tinggi, sementara Indonesia masih didominasi oleh fasilitas Retail Colocation dengan kapasitas daya per gedung yang relatif kecil. Namun, dari sisi reliabilitas, mayoritas pusat data baru di Indonesia telah memenuhi standar Tier III, menunjukkan kesiapan teknis untuk beban kerja kritis.

##### 4. Tantangan Keberlanjutan (SDG 13):

Adopsi energi terbarukan pada sektor pusat data di Indonesia masih tertinggal signifikan dibandingkan standar global. Ketergantungan pada bauran energi nasional yang masih didominasi bahan bakar fosil menjadi hambatan kompetitif utama, mengingat investor teknologi global (Big Tech) kini mensyaratkan standar Net-Zero dalam pemilihan lokasi infrastruktur mereka.

##### 5. Penerapan algoritma *Machine Learning (K-Means)*

berhasil memvalidasi posisi Indonesia secara objektif dalam klaster *Emerging Market*, terpisah dari klaster *Mature Market* yang dihuni Singapura. Hal ini membuktikan secara komputasional bahwa Indonesia sedang dalam fase transisi kritis menuju negara ekonomi digital maju.

## 5.2 Saran

Berdasarkan temuan di atas, penulis mengajukan rekomendasi strategis bagi berbagai pemangku kepentingan:

### 1. Bagi Pemerintah (Regulator)

- a **Insentif Hijau:** Diperlukan kebijakan insentif fiskal (*Tax Allowance* atau *Tax Holiday*) khusus bagi pengembang pusat data yang mampu mengintegrasikan pembangkit listrik energi terbarukan (seperti PLTS Atap) secara mandiri.
- b **Zonasi Khusus:** Mempercepat pengembangan Kawasan Ekonomi Khusus (KEK) Digital yang menjamin ketersediaan lahan dan pasokan listrik hijau yang stabil untuk menarik investasi *Hyperscale*.

### 2. Bagi Pelaku Industri

- a **Efisiensi Energi:** Mengingat iklim tropis Indonesia, operator disarankan untuk beralih dari pendinginan konvensional ke teknologi *Liquid Cooling* atau *Immersion Cooling* untuk menurunkan nilai PUE (*Power Usage Effectiveness*).
- b **Ekspansi Kapasitas:** Fokus investasi sebaiknya tidak lagi hanya pada penambahan jumlah gedung, melainkan pada peningkatan densitas daya (MW per rack) untuk mengakomodasi beban kerja *Artificial Intelligence* (AI).

### 3. Bagi Pengembangan Penelitian Selanjutnya

- a **Integrasi Data Finansial:** Penelitian selanjutnya disarankan untuk menambahkan variabel ekonomi seperti harga sewa *colocation* per rak atau biaya listrik per kWh untuk memperkaya analisis kelayakan investasi.
- b **Prediksi Machine Learning:** Mengembangkan fitur *Forecasting* menggunakan algoritma *Machine Learning* untuk memprediksi kebutuhan kapasitas pusat data di Indonesia dalam 5–10 tahun ke depan berdasarkan tren pertumbuhan trafik data.

## **DAFTAR PUSTAKA**

Google, Temasek, & Bain. (2024). e-Economy SEA 2024: Profits at the precipice, harnessing the advantage.

Structure Research. (2023). Indonesia Data Centre Colocation Market Report. (Data pembanding industri untuk memvalidasi temuan Growth Rate).

APJII. (2025). Survei Profil Internet Indonesia 2025. Jakarta: Asosiasi Penyelenggara Jasa Internet Indonesia.

International Energy Agency (IEA). (2025). Electricity 2025: Analysis and forecast to 2027. Paris: IEA.

Uptime Institute. (2024). Annual Global Data Center Survey 2024. New York: Uptime Institute Intelligence.

Chidolue, O., Ohenhen, P., Umoh, A., et al. (2024). Green Data Centers: Sustainable Practices for Energy-Efficient IT Infrastructure. *Engineering Science & Technology Journal*, 5(1), 99-114.

DOI: 10.51594/estj.v5i1.730

Zhang, L., & Chen, S. (2024). Big Data Analytics for Carbon Neutrality in IT Infrastructure: A Review. *Sustainability*, 16(4), 1820.

DOI: <https://doi.org/10.3390/su16041820>

Ranjan, J., & Foropon, C. (2021). Big data analytics in building the competitive intelligence of organizations. *International Journal of Information Management*, 56, 102231.

DOI: <https://doi.org/10.1016/j.ijinfomgt.2020.102231>

United Nations. (2023). The Sustainable Development Goals Report 2023: Special Edition.

**McKinney, W. (2011).** *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.

**Plotly Technologies Inc. (2024).** *Collaborative data science*. Montreal, QC: Plotly Technologies Inc.

**Streamlit Inc. (2024).** *Streamlit Documentation*.