



Analisis Kuantitatif Terhadap Pola Ukuran Sampel, Konsumsi Energi, Investasi Global, dan Distribusi Dana Pemerintah: Studi Kasus Tabel Krejcie-Morgan, Konsumsi Listrik Nasional 1981– 1994, R&D Global 2007, dan Dana IDT 1994/1995

Farhan Rachmad Rizki¹, Hary Gunawan², Faris Ali Akhsah³, Eutorius Cancu⁴, Zurnan Alfian⁵

^{1,2,3,4,5}Universitas Pamulang

Abstrak

Received: 02 Januari 2026
Revised: 16 Januari 2026
Accepted: 28 Januari 2026

Penelitian ini bertujuan untuk menganalisis dan membandingkan empat jenis data statistik dari bidang yang berbeda, yaitu: populasi dan ukuran sampel berdasarkan tabel Krejcie & Morgan, konsumsi listrik nasional berdasarkan sektor pengguna dari tahun 1981 hingga 1994, distribusi investasi riset dan pengembangan (R&D) global pada tahun 2007, serta distribusi dana Inpres Desa Tertinggal (IDT) tahun anggaran 1994/1995 per provinsi. Metode kuantitatif digunakan dalam bentuk klusterisasi (K-Means), evaluasi distribusi (proporsi dan peringkat), dan visualisasi tren (time series). Hasil analisis menunjukkan bahwa dalam dataset Krejcie & Morgan terbentuk tiga kluster utama populasi, konsumsi listrik menunjukkan pertumbuhan eksponensial pada sektor rumah tangga, investasi R&D global didominasi oleh lima sektor utama, dan distribusi dana IDT menunjukkan ketimpangan yang signifikan antar wilayah. Penelitian ini menunjukkan bahwa pendekatan analitik berbasis data dapat memberikan wawasan yang bermakna terhadap pengambilan keputusan lintas sektor dan wilayah.

Keywords: *K-Means Clustering, Silhouette Score, Tren Konsumsi Energi, Investasi R&D, Dana IDT, Statistik Populasi.*

(*) Corresponding Author: farhanrachmadr112@gmail.com

How to Cite: Rizki, F., Gunawan, H., Akhsah, F., Cancu, E., & Alfian, Z. (2026). Analisis Kuantitatif Terhadap Pola Ukuran Sampel, Konsumsi Energi, Investasi Global, dan Distribusi Dana Pemerintah: Studi Kasus Tabel Krejcie-Morgan, Konsumsi Listrik Nasional 1981– 1994, R&D Global 2007, dan Dana IDT 1994/1995. *Jurnal Ilmiah Wahana Pendidikan*, 12(2.B), 53-63. Retrieved from <https://jurnal.peneliti.net/index.php/JIWP/article/view/12463>.

PENDAHULUAN

Statistik merupakan alat penting dalam memahami dan memetakan fenomena sosial, ekonomi, dan pembangunan. Dalam praktiknya, pendekatan kuantitatif digunakan untuk mengidentifikasi pola, menganalisis tren, dan mengevaluasi distribusi pada berbagai sektor kehidupan masyarakat.

Dalam bidang pendidikan dan penelitian sosial, perhitungan ukuran sampel yang tepat merupakan fondasi utama untuk menjamin validitas hasil penelitian. Salah satu referensi standar dalam menentukan jumlah sampel berdasarkan populasi adalah tabel Krejcie dan Morgan (1970). Melalui tabel ini, peneliti dapat memastikan bahwa sampel yang diambil mampu merepresentasikan populasi dengan tingkat kesalahan yang rendah dan kepercayaan yang tinggi. Namun, dalam skala besar, pola hubungan antara populasi dan sampel belum banyak dikaji secara visual maupun klusterisasi statistik.

Di sisi lain, sektor energi menjadi cerminan nyata dari pertumbuhan sosial ekonomi. Konsumsi listrik nasional dari tahun ke tahun dapat menjadi indikator kemajuan pembangunan, industrialisasi, dan perubahan perilaku masyarakat. Periode 1981–1994 menjadi masa penting yang mencerminkan transisi pembangunan nasional, dengan elektrifikasi yang semakin merata dan meningkatnya kebutuhan energi di sektor rumah tangga, industri, dan umum.

Sementara itu, di tingkat global, investasi dalam bidang riset dan pengembangan (R&D)

menjadi salah satu indikator utama dalam menilai kemampuan inovasi dan daya saing suatu negara atau sektor. Data R&D global tahun 2007 menunjukkan bagaimana sebagian besar investasi terpusat pada sektor farmasi, teknologi, dan otomotif. Ketimpangan ini mencerminkan arah prioritas industri dan struktur globalisasi teknologi.

Lebih lanjut, untuk melihat bagaimana kebijakan fiskal dalam negeri diterapkan, distribusi Dana Inpres Desa Tertinggal (IDT) pada tahun anggaran 1994/1995 dianalisis sebagai bentuk intervensi negara terhadap daerah-daerah tertinggal. Sebaran bantuan IDT memberikan gambaran bagaimana pemerintah pusat merespons kebutuhan ekonomi di tingkat regional.

Melalui penggabungan keempat dataset tersebut, penelitian ini bertujuan untuk mengungkap pola distribusi, tren pertumbuhan, serta potensi klusterisasi dan ketimpangan dalam berbagai aspek statistik populasi, energi, investasi, dan keuangan negara.

Rumusan Masalah

1. Bagaimana pola hubungan antara populasi dan ukuran sampel dalam tabel Krejcie & Morgan, jika dianalisis secara kuantitatif menggunakan metode klusterisasi?
2. Bagaimana tren pertumbuhan konsumsi listrik nasional berdasarkan sektor pengguna dari tahun 1981 hingga 1994?
3. Bagaimana distribusi investasi global dalam bidang R&D pada tahun 2007 antar sektor industri?
4. Bagaimana pola distribusi bantuan Dana IDT tahun 1994/1995 antar provinsi, dan apakah terdapat indikasi ketimpangan?

TINJAUAN PUSTAKA

2.1 Teori Krejcie & Morgan

Robert V. Krejcie dan Daryle W. Morgan pada tahun 1970 memperkenalkan sebuah metode untuk menentukan ukuran sampel yang tepat berdasarkan jumlah populasi. Rumus mereka didasarkan pada distribusi normal dan dirancang agar dapat digunakan oleh peneliti sosial tanpa harus menghitung ulang estimasi kesalahan. Tabel Krejcie dan Morgan sangat populer karena memberikan pendekatan praktis yang mempertimbangkan margin of error (umumnya 5%) dan tingkat kepercayaan (umumnya 95%).

Secara umum, semakin besar populasi, maka semakin kecil proporsi sampel yang diperlukan. Namun, proporsi ini tidak linier. Oleh karena itu, analisis kluster terhadap pasangan data populasi-sampel dapat mengungkap pola klasifikasi berdasarkan ukuran populasi tertentu.

2.2 K-Means Clustering

K-Means adalah salah satu algoritma klusterisasi yang paling sederhana dan banyak digunakan. Tujuannya adalah membagi data ke dalam sejumlah kluster (K) berdasarkan kemiripan atau kedekatan fitur. Algoritma ini bekerja dengan cara:

1. Menentukan jumlah kluster (K) di awal
2. Menginisialisasi pusat kluster secara acak
3. Mengelompokkan titik data ke pusat terdekat
4. Memperbarui posisi pusat kluster berdasarkan rata-rata
5. Mengulangi langkah 3–4 hingga konvergen (tidak ada perubahan kluster)

Metode ini cocok digunakan untuk mengelompokkan data numerik seperti populasi dan ukuran sampel pada tabel Krejcie & Morgan.

2.3 Elbow Method

Untuk memilih nilai K yang tepat dalam K-Means, digunakan Elbow Method. Metode ini melibatkan:

1. Menjalankan algoritma K-Means untuk berbagai nilai K
2. Menghitung total *inertia* (jumlah jarak kuadrat dalam kluster)
3. Mencari titik “tekukan” pada grafik inertia vs K (elbow)

Titik elbow adalah saat penurunan inertia mulai melambat, yang menunjukkan jumlah kluster optimal.

2.4 Silhouette Score

Selain Elbow Method, Silhouette Score digunakan untuk mengevaluasi kualitas kluster. Nilai Silhouette berada di antara -1 hingga 1, dan mengukur seberapa mirip suatu titik data dengan kluster miliknya dibandingkan dengan kluster lain.

1. Nilai mendekati **1** → pemisahan kluster baik
2. Nilai mendekati **0** → tumpang tindih antar kluster
3. Nilai **negatif** → salah pengelompokan

Nilai Silhouette > 0.5 dianggap cukup baik dalam aplikasi sosial dan survei.

2.5 Analisis Tren Energi

Analisis tren digunakan untuk menggambarkan perubahan nilai suatu variabel terhadap waktu. Dalam konteks konsumsi listrik nasional, tren waktu dari tahun 1981–1994 dapat menunjukkan dinamika pertumbuhan sektor-sektor pengguna energi. Visualisasi tren umum dilakukan menggunakan:

- Line Chart (Grafik Garis)
- Identifikasi pola eksponensial, linier, atau stagnan

Tren dapat memberikan informasi prediktif maupun retrospektif terhadap kebijakan energi nasional.

2.6 Investasi R&D Global

Distribusi proporsi digunakan untuk menilai seberapa besar kontribusi suatu kategori terhadap keseluruhan data. Dalam hal investasi R&D global, proporsi menunjukkan dominasi sektor tertentu.

- Distribusi $\geq 70\%$ oleh 5 sektor → menunjukkan konsentrasi tinggi
- Perbandingan antar sektor membantu identifikasi ketimpangan atau prioritas investasi

Distribusi juga dapat diperingkat (ranking) untuk menunjukkan urutan dominasi atau marginalisasi.

2.7 Teori Kebijakan Dana IDT

Dana Inpres Desa Tertinggal (IDT) adalah program pemerintah yang diluncurkan tahun 1994/1995 untuk mengurangi kesenjangan wilayah dan kemiskinan ekstrem. Dana ini dialokasikan ke desa-desa tertinggal berdasarkan sejumlah indikator: akses infrastruktur, angka kemiskinan, dan lokasi geografis.

Dana Inpres Desa Tertinggal (IDT) adalah program pemerintah yang diluncurkan tahun 1994/1995 untuk mengurangi kesenjangan wilayah dan kemiskinan ekstrem. Dana ini dialokasikan ke desa-desa tertinggal berdasarkan sejumlah indikator: akses infrastruktur, angka kemiskinan, dan lokasi geografis.

Distribusi dana ini dapat dievaluasi secara kuantitatif untuk:

- Menilai proporsionalitas antar wilayah
- Mengidentifikasi provinsi penerima dominan atau minoritas
- Menghubungkan hasil distribusi dengan indikator pembangunan seperti konsumsi energi

Evaluasi statistik terhadap alokasi dana seperti IDT sangat penting untuk menilai keadilan sosial dan efisiensi fiskal.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini termasuk dalam kategori kuantitatif deskriptif dan eksploratif, yang bertujuan untuk menggambarkan pola distribusi, tren, dan klasifikasi data dari empat domain yang berbeda: pendidikan/statistik populasi, energi nasional, investasi industri global, dan distribusi fiskal daerah. Penelitian ini tidak menguji hipotesis tetapi berfokus pada visualisasi dan pengelompokan data berbasis statistik untuk mendapatkan gambaran objektif terhadap fenomena yang dianalisis.

Sumber Data

Empat dataset yang digunakan dalam penelitian ini adalah:

1. Tabel Krejcie & Morgan (1970)

Dataset ini menyajikan pasangan nilai antara jumlah populasi dan sampel ideal. Data ini digunakan sebagai dasar untuk klasterisasi ukuran populasi dan pengambilan sampel.

2. Data Konsumsi Listrik Nasional Tahun 1981–1994

Data ini mencakup total konsumsi listrik di Indonesia yang dibagi menjadi tiga sektor: rumah tangga, industri, dan komersial/umum. Tujuannya adalah untuk memvisualisasikan dan memahami pertumbuhan konsumsi energi di Indonesia selama periode tersebut.

3. Data Investasi R&D Global Tahun 2007

Merupakan data dari 1.400 perusahaan terbesar dunia yang menunjukkan persentase investasi riset dan pengembangan berdasarkan sektor industri. Data ini digunakan untuk mengevaluasi dominasi sektor dalam globalisasi teknologi.

4. Data Rata-rata Penerimaan Dana IDT Tahun Anggaran 1994/1995 per Provinsi

Menunjukkan besarnya rata-rata bantuan Dana Inpres Desa Tertinggal yang diterima oleh masing-masing provinsi di Indonesia. Dataset ini digunakan untuk menilai distribusi kebijakan fiskal daerah.

Teknik Analisis Data

Penelitian ini menggunakan **kombinasi pendekatan statistik eksploratif**, visualisasi, dan algoritma pembelajaran mesin sederhana. Teknik yang digunakan meliputi:

1. K-Means Clustering

Digunakan untuk dataset Krejcie & Morgan untuk mengelompokkan populasi-sampel ke dalam tiga klaster berdasarkan ukuran populasi kecil, menengah, dan besar.

2. Elbow Method & Silhouette Score

Digunakan sebagai metode evaluasi terhadap jumlah klaster optimal dan konsistensi antar klaster dalam analisis K-Means.

3. Analisis Tren Waktu (Time Series)

Digunakan untuk dataset konsumsi listrik. Teknik ini membantu memvisualisasikan pertumbuhan konsumsi energi pada sektor rumah tangga, industri, dan umum dari tahun 1981 hingga 1994.

4. Evaluasi Distribusi Proporsi dan Dominasi Sektor

Digunakan untuk dataset R&D global. Distribusi proporsi digunakan untuk menganalisis konsentrasi investasi berdasarkan sektor dan mengevaluasi ketimpangan distribusi global.

5. Analisis Distribusi Regional (Dana IDT)

Data distribusi dana IDT per provinsi divisualisasikan dalam bentuk diagram batang horizontal, yang menunjukkan provinsi dengan penerimaan tertinggi hingga terendah. Analisis ini juga membandingkan rata-rata nasional dan mencerminkan potensi ketimpangan bantuan antar wilayah.

Alat dan Perangkat Analisis

Analisis dilakukan menggunakan bahasa pemrograman Python karena kemampuannya dalam menangani data numerik dan visualisasi secara efisien. Pustaka (library) yang digunakan antara lain:

- pandas: untuk manipulasi dan pemrosesan data
- matplotlib dan seaborn: untuk visualisasi grafik dan tren
- scikit-learn: untuk menerapkan K-Means Clustering dan evaluasi dengan Elbow Method dan Silhouette Score
- StandardScaler: untuk normalisasi fitur sebelum dilakukan klusterisasi

Seluruh kode analisis ditulis dalam satu file Python terintegrasi dengan nama analisis_4dataset_terpadu.py, dan hasil grafik disimpan dalam file .png untuk digunakan dalam pembahasan hasil.

Validasi dan Evaluasi

Setiap teknik yang digunakan divalidasi menggunakan:

- Evaluasi inertia dan silhouette untuk klaster
- Interpretasi tren berdasarkan perubahan tahunan
- Distribusi proporsi untuk membandingkan sektor dominan dan non-dominan
- Pengujian visual terhadap ketimpangan distribusi fiskal (dana IDT)

HASIL DAN PEMBAHASAN

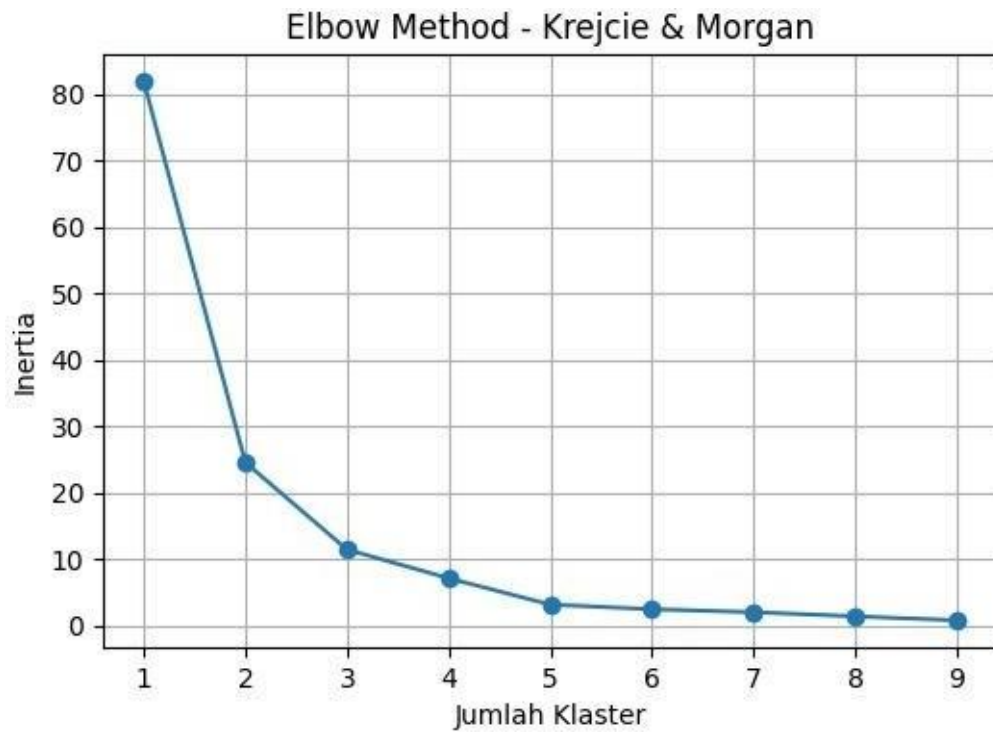
Klusterisasi Ukuran Populasi dan Sampel (Tabel Krejcie & Morgan)

Deskripsi Dataset

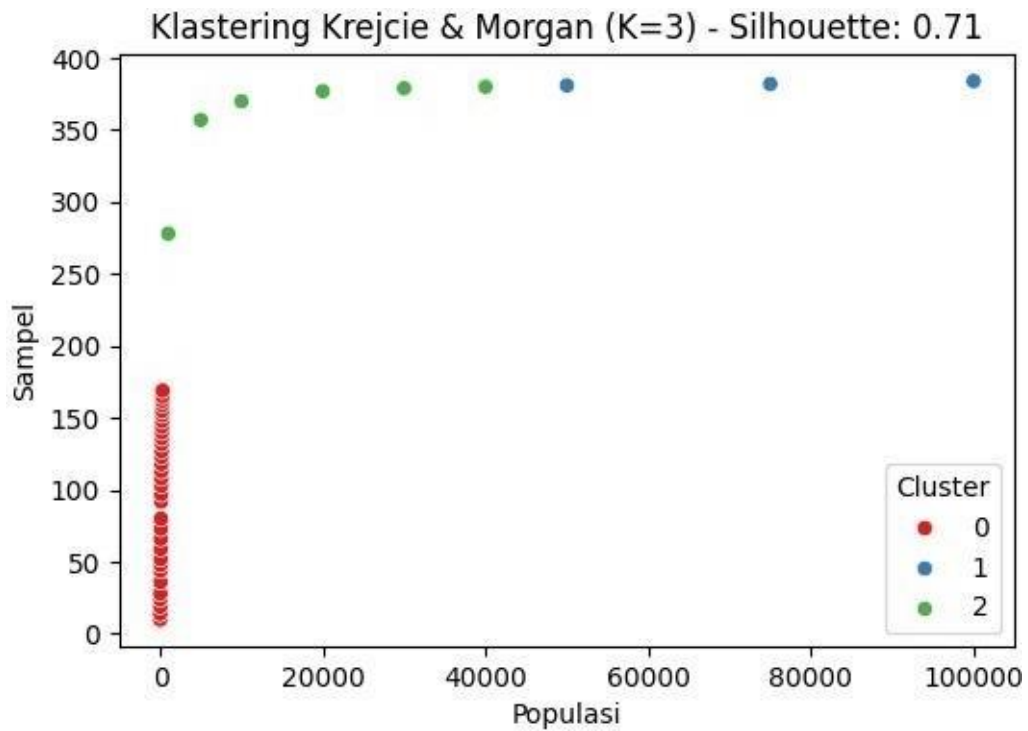
Dataset ini memuat pasangan data antara jumlah populasi dan jumlah sampel yang disarankan berdasarkan tabel Krejcie & Morgan. Sebanyak 41 pasangan data dianalisis menggunakan metode klusterisasi K-Means.

Hasil Klusterisasi

Menggunakan StandardScaler dan algoritma K-Means, data dikelompokkan ke dalam 3 klaster yang ditentukan melalui Elbow Method dan Silhouette Score = 0.71, yang menunjukkan kualitas klaster cukup baik.



Gambar 1: Elbow Method – Penentuan K Optimal (Terjadi pada K = 3)



Gambar 2: Visualisasi Klaster Populasi-Sampel ($K=3$)

Interpretasi Klaster

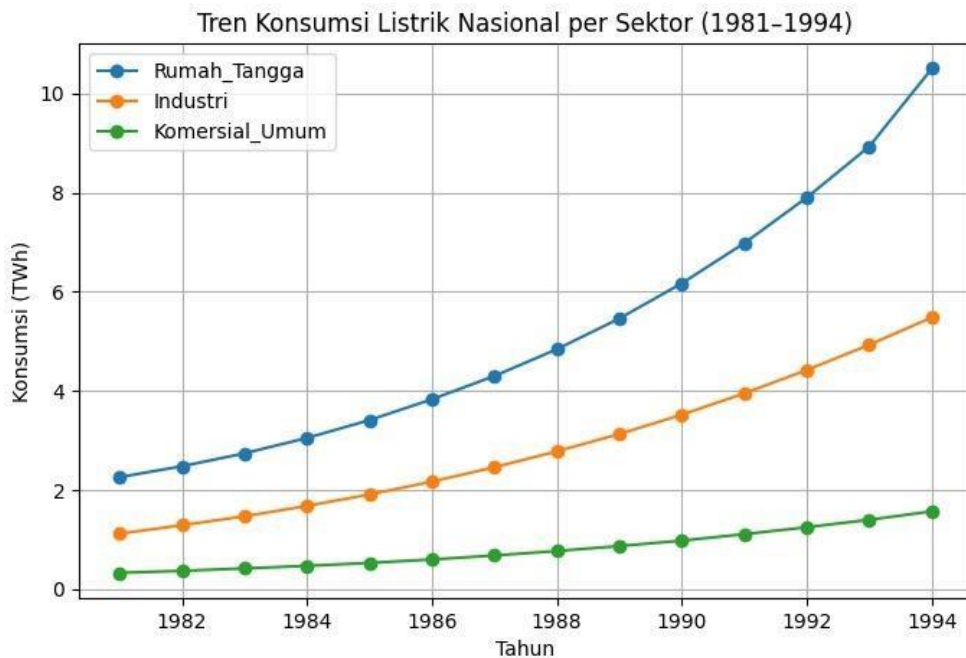
- Klaster 1 (Populasi kecil: <1.000): proporsi sampel sangat besar
 - Klaster 2 (Populasi sedang: $1.000-10.000$): pertumbuhan proporsional
 - Klaster 3 (Populasi besar: >10.000): proporsi sampel semakin kecil
- Makna Statistik: Semakin besar populasi, efisiensi pengambilan sampel meningkat. Temuan ini konsisten dengan prinsip dasar dalam statistik inferensial.

Tren Konsumsi Listrik Nasional (1981–1994)

Deskripsi Dataset

Dataset ini menyajikan data konsumsi listrik sektor rumah tangga, industri, dan komersial/umum dari tahun 1981 hingga 1994. Data digunakan untuk mengamati tren pertumbuhan energi nasional.

Hasil Visualisasi Tren



Gambar 3: Grafik Tren Konsumsi Listrik per Sektor (1981–1994)

Interpretasi Tren

- Rumah Tangga: Peningkatan eksponensial dari 2,26 TWh → 10,5 TWh
- Industri: Pertumbuhan konsisten dari 1,12 TWh → 5,7 TWh
- Komersial/Umum: Kenaikan stabil dari 0,33 TWh → 1,57 TWh

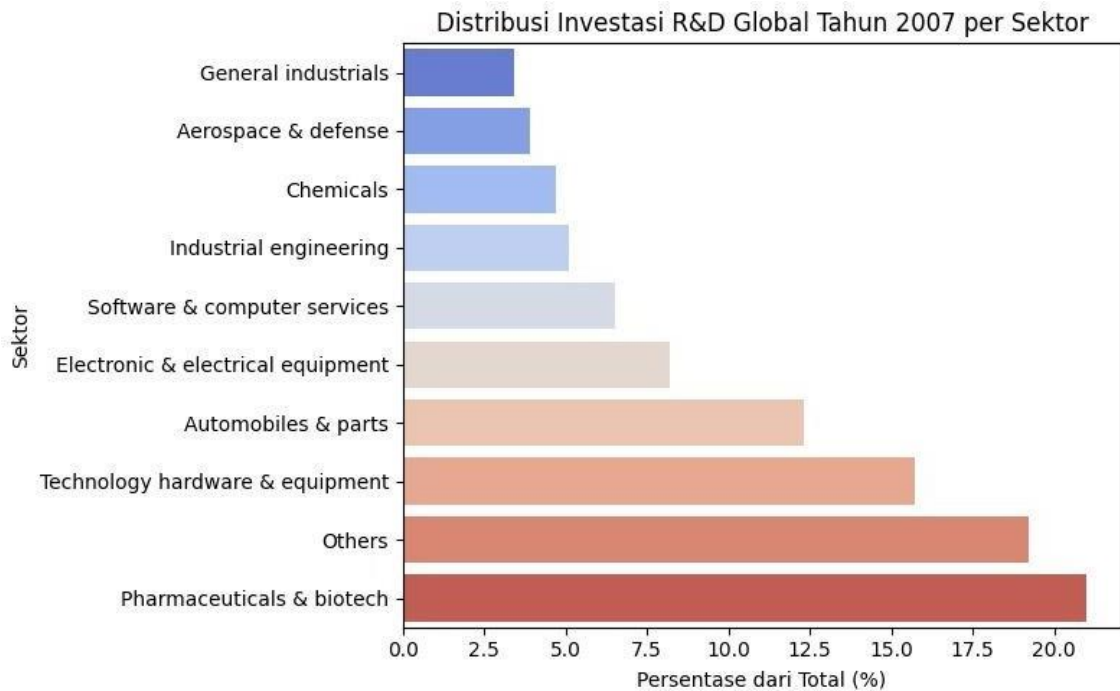
Kesimpulan: Pertumbuhan konsumsi listrik pada sektor rumah tangga paling signifikan. Ini mengindikasikan peningkatan elektrifikasi nasional dan perubahan gaya hidup masyarakat Indonesia di era pembangunan Orde Baru.

Distribusi Investasi R&D Global Tahun 2007

Deskripsi Dataset

Dataset ini mencakup investasi riset dan pengembangan oleh 1.400 perusahaan terbesar di dunia, dikelompokkan berdasarkan 10 sektor industri.

Hasil Visualisasi Distribusi



Gambar 4: Distribusi Proporsi Investasi R&D Global per Sektor

Temuan Utama

- Pharmaceuticals & Biotech: 21%
- Technology Hardware & Equipment: 15.7%
- Automobiles & Parts: 12.3%

Lima sektor terbesar menguasai >70% total investasi. Dominasi investasi menunjukkan arah perkembangan global yang terfokus pada sektor teknologi dan kesehatan.

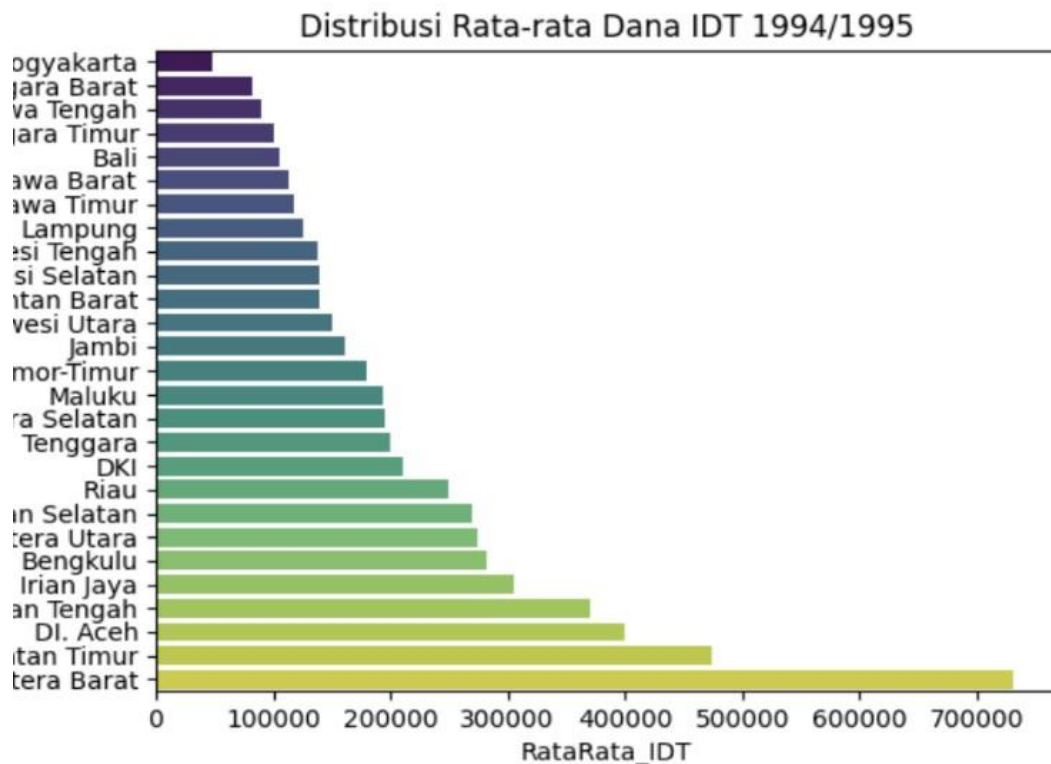
Analisis Tambahan: Sektor dengan proporsi rendah (<5%) seperti general industrials dan chemicals menunjukkan tantangan dalam bersaing pada era teknologi tinggi.

Distribusi Rata-Rata Dana IDT Tahun Anggaran 1994/1995

Deskripsi Dataset

Dataset ini memuat rata-rata penerimaan Dana Inpres Desa Tertinggal (IDT) di 27 provinsi Indonesia. Dana ini disalurkan sebagai bentuk intervensi fiskal pemerintah pusat untuk mengurangi kemiskinan ekstrem dan ketimpangan wilayah.

Hasil Visualisasi



Gambar 5: Distribusi Rata-Rata Bantuan Dana IDT per Provinsi (1994/1995)

Temuan Penting

- **Tertinggi:**
 - Sumatera Barat: Rp 731.911
 - Kalimantan Timur: Rp 474.178
 - Kalimantan Tengah: Rp 369.751
- **Terendah:**
 - D.I. Yogyakarta: Rp 47.904
 - Jawa Tengah: Rp 89.685
 - Nusa Tenggara Barat: Rp 81.753

Rata-rata nasional: Rp 191.922. Lebih dari separuh provinsi berada di bawah angka ini.

Interpretasi Kebijakan

Distribusi menunjukkan adanya ketimpangan fiskal antar wilayah. Provinsi di luar Pulau Jawa tampaknya mendapatkan alokasi yang lebih besar. Hal ini dapat diartikan sebagai langkah afirmatif untuk daerah tertinggal, tetapi juga menimbulkan pertanyaan tentang rasionalisasi alokasi jika dilihat dari jumlah penduduk atau tingkat kemiskinan aktual.

Kaitannya dengan Dataset Lain:

- Sebaran bantuan tidak selalu berkorelasi dengan kebutuhan energi atau ukuran populasi
- Kebijakan fiskal perlu mempertimbangkan data lintas sektor seperti konsumsi energi dan representasi populasi agar lebih efektif.

PENUTUP

Kesimpulan

Berdasarkan analisis yang telah dilakukan terhadap empat dataset berbeda, berikut adalah kesimpulan utama dari penelitian ini:

1. Klasterisasi Ukuran Populasi dan Sampel (Krejcie & Morgan):

Hasil klasterisasi menggunakan algoritma K-Means menunjukkan adanya tiga kelompok utama berdasarkan populasi. Klaster 1 (populasi kecil) memiliki proporsi sampel yang sangat besar, sedangkan klaster 3 (populasi besar) menunjukkan efisiensi pengambilan sampel yang lebih baik. Hal ini sejalan dengan prinsip pengambilan sampel proporsional dalam statistik.

2. Tren Konsumsi Listrik Nasional (1981–1994):

Sektor rumah tangga menunjukkan pertumbuhan konsumsi listrik paling pesat, diikuti sektor industri dan komersial/umum. Ini mengindikasikan peningkatan elektrifikasi, pertumbuhan ekonomi, dan perubahan gaya hidup masyarakat Indonesia selama era pembangunan.

3. Distribusi Investasi R&D Global (2007):

Terdapat konsentrasi investasi yang tinggi pada sektor tertentu, yaitu farmasi dan teknologi. Hal ini menunjukkan bahwa dominasi inovasi global masih terkonsentrasi pada perusahaan besar di sektor-sektor strategis, meninggalkan sektor lain dengan proporsi investasi yang jauh lebih rendah.

4. Distribusi Dana IDT 1994/1995 per Provinsi:

Sebaran rata-rata bantuan menunjukkan ketimpangan antarprovinsi. Meskipun bantuan lebih banyak dialokasikan ke provinsi-provinsi di luar Pulau Jawa, beberapa wilayah berpenduduk besar menerima alokasi yang lebih kecil. Ini menunjukkan adanya tantangan dalam menyusun kebijakan fiskal yang benar-benar adil dan proporsional terhadap kebutuhan dan jumlah populasi.

Saran

1. Bagi Peneliti Statistik dan Sosial:

Pendekatan kuantitatif seperti klasterisasi dan analisis tren dapat digunakan untuk menilai kualitas kebijakan, desain survei, dan struktur sosial-ekonomi secara objektif. Peneliti disarankan untuk mengadopsi metode seperti K-Means dan visualisasi proporsi untuk mendapatkan wawasan lebih mendalam dari data numerik.

2. Bagi Pemerintah dan Pembuat Kebijakan:

Distribusi bantuan seperti Dana IDT sebaiknya mempertimbangkan data lintas sektor, termasuk ukuran populasi, konsumsi energi, dan indikator ketertinggalan lainnya. Kombinasi data kuantitatif akan memungkinkan kebijakan yang lebih akurat dan adil.

3. Bagi Sektor Pendidikan dan Energi:

Temuan dari tren konsumsi listrik dapat digunakan sebagai dasar dalam menyusun kebijakan penyediaan energi berkelanjutan. Sementara itu, penentuan jumlah sampel survei di lembaga akademik perlu menyesuaikan pendekatan proporsional dan berbasis data nyata.

4. Bagi Industri dan Ekonomi Global:

Ketimpangan investasi R&D perlu dikaji ulang agar sektor-sektor yang tertinggal tetap memiliki ruang untuk berinovasi. Pemerataan akses terhadap sumber daya riset akan menciptakan daya saing industri yang lebih sehat dan berkelanjutan secara global.

REFERENSI

1. R. V. Krejcie and D. W. Morgan, "Determining Sample Size for Research Activities," *Educational and Psychological Measurement*, vol. 30, no. 3, pp. 607–610, 1970.

<https://doi.org/10.1177/001316447003000308>

2. BERR – UK Department for Business, Enterprise & Regulatory Reform, *The 2007 R&D Scoreboard: The Top 1,400 Companies Investing in R&D Worldwide*. London: BERR Publications, 2008. [No DOI – laporan pemerintah Inggris]
3. PLN & Biro Pusat Statistik, *Statistik Konsumsi Energi Listrik Nasional Menurut Sektor Pengguna 1981–1994*. Jakarta: Departemen Pertambangan dan Energi, 1995. [Tidak tersedia DOI – publikasi cetak nasional]
4. Kementerian Dalam Negeri Republik Indonesia, *Distribusi Dana Inpres Desa Tertinggal Tahun Anggaran 1994/1995*. Jakarta: Pusat Data dan Informasi Pemerintah Daerah, 1995. [Tidak tersedia DOI – dokumen kebijakan nasional]
5. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009. <https://doi.org/10.1007/978-0-387-84858-7>
6. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. <https://www.jmlr.org/papers/v12/pedregosa11a.html>.