

# *Analisis Hubungan Antara Tingkat Kemiskinan dan Penggunaan Internet di Indonesia: Dasar untuk Pembangunan Internet Gratis yang Tepat Sasaran*

Mohammad Ikhya Ulumuddin

*Program Studi : Sistem Informasi, Universitas Darwan Ali*

Email : [reqwyazka@gmail.com](mailto:reqwyazka@gmail.com)

---

**ABSTRACT**— This study investigates the correlation between poverty rates and internet usage across Indonesian provinces. Using data from Statistics Indonesia (BPS) in 2018—covering the proportion of internet users and poverty rates—this research applies K-Means Clustering to classify provinces based on the severity of poverty and the level of internet access. The goal is to identify regions that suffer from both high poverty and low internet usage, which are deemed high-priority targets for the implementation of free internet infrastructure. The results reveal several provinces, particularly in eastern Indonesia, that fall into this category. These findings are expected to serve as a data-driven reference for policymakers in designing equitable internet infrastructure development.

**Keywords**— Poverty, Internet Access, Clustering, K-Means, Digital Divide, Indonesia.

**ABSTRAK**— Penelitian ini menganalisis hubungan antara tingkat kemiskinan dan penggunaan internet di berbagai provinsi di Indonesia. Data yang digunakan berasal dari Badan Pusat Statistik (BPS) tahun 2018, mencakup proporsi pengguna internet dan persentase penduduk miskin. Metode yang digunakan adalah algoritma K-Means Clustering untuk mengelompokkan provinsi berdasarkan tingkat kemiskinan dan akses internet. Tujuan utama dari penelitian ini adalah mengidentifikasi wilayah-wilayah yang memiliki tingkat kemiskinan tinggi dan akses internet rendah, sehingga dapat dijadikan prioritas dalam pembangunan infrastruktur internet gratis. Hasil analisis menunjukkan bahwa beberapa provinsi di wilayah timur Indonesia masuk dalam kategori tersebut. Temuan ini diharapkan dapat menjadi dasar pertimbangan dalam perencanaan pembangunan internet gratis yang lebih tepat sasaran dan berbasis data.

**Kata kunci**— Kemiskinan, Akses Internet, Klastering, K-Means, Kesenjangan Digital, Indonesia.

---

## I. PENDAHULUAN

Indonesia merupakan negara kepulauan dengan tingkat pertumbuhan teknologi informasi dan komunikasi yang cukup tinggi dalam satu dekade terakhir. Akses internet telah menjadi kebutuhan pokok, baik untuk pendidikan, pekerjaan, maupun aktivitas sosial masyarakat. Namun, penyebaran akses internet di Indonesia masih belum merata. Beberapa wilayah terutama di bagian timur Indonesia masih mengalami kesulitan dalam mengakses jaringan internet yang layak.

Kondisi tersebut menjadi lebih kompleks ketika dikaitkan dengan tingkat kemiskinan di wilayah-wilayah tersebut. Akses yang terbatas terhadap internet pada masyarakat miskin menyebabkan keterisolasian digital (digital exclusion), yang pada akhirnya memperparah ketimpangan sosial dan ekonomi.

Menurut laporan Badan Pusat Statistik (BPS), pada tahun 2018 terdapat perbedaan yang signifikan dalam proporsi pengguna internet dan tingkat kemiskinan antarprovinsi di Indonesia. DKI Jakarta memiliki tingkat penggunaan internet mencapai lebih dari 80% dengan tingkat kemiskinan di bawah 5%, sementara Papua hanya mencatat sekitar 20% pengguna internet dengan tingkat kemiskinan lebih dari 25%.

Permasalahan mendasar yang diangkat dalam penelitian ini adalah bagaimana mengidentifikasi wilayah-wilayah di Indonesia yang menjadi prioritas dalam pembangunan internet gratis berdasarkan data empiris. Alih-alih melakukan pembangunan merata, pembangunan berbasis data dan kebutuhan riil akan lebih efisien dan berdampak.

Beberapa penelitian sebelumnya telah membahas pentingnya internet dalam pengurangan kemiskinan dan peningkatan pembangunan manusia [3][4]. Namun, belum banyak penelitian yang mengintegrasikan data kemiskinan dan penggunaan internet secara bersamaan dalam satu model analisis berbasis klaster untuk kebutuhan kebijakan pembangunan infrastruktur digital.

Dengan menggunakan metode K-Means Clustering, penelitian ini mencoba mengelompokkan provinsi-provinsi berdasarkan dua parameter utama: tingkat kemiskinan dan tingkat penggunaan internet. Hasil klaster diharapkan memberikan gambaran wilayah mana saja yang menjadi target strategis untuk pembangunan internet gratis.

## II. METODOLOGI PENELITIAN

### A. Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari sumber resmi, yaitu Badan Pusat Statistik (BPS) Indonesia. Data yang digunakan mencakup dua indikator penting yang saling berkaitan dengan tujuan penelitian, yakni: (1) proporsi individu yang menggunakan internet menurut provinsi, dan (2) persentase penduduk miskin menurut provinsi.

- Proporsi Individu yang Menggunakan Internet Menurut Provinsi (%), 2018

Row No.	provinsi	Proporsi penggunaan Internet(Persen)
1	ACEH	30.690
2	SUMATERA ...	34.270
3	SUMATERA ...	36.490
4	RIAU	39.980
5	JAMBI	35.820

Gambar 1. Dataset individu yang menggunakan internet (Provinsi)

Data ini menyajikan informasi mengenai tingkat penetrasi penggunaan internet oleh individu di setiap provinsi di Indonesia. Indikator ini merepresentasikan sejauh mana masyarakat di suatu wilayah memiliki akses terhadap teknologi informasi, khususnya internet. Nilai persentase yang tercantum menunjukkan proporsi pengguna internet terhadap total jumlah penduduk di masing-masing provinsi. Data ini digunakan untuk menganalisis sebaran digitalisasi di seluruh wilayah Indonesia.

- Persentase Penduduk Miskin Menurut Provinsi, 2018

Row No.	provinsi	Persentase Penduduk Miskin
1	Aceh	15.680
2	Sumatera Utara	8.940
3	Sumatera Barat	6.550
4	Riau	7.210
5	Jambi	7.850
6	Sumatera Sel...	12.820

Gambar 2. Dataset penduduk miskin (Provinsi)

Data ini menunjukkan persentase penduduk yang berada di bawah garis kemiskinan di setiap provinsi. Variabel ini digunakan untuk mengidentifikasi wilayah-wilayah dengan kerentanan sosial ekonomi tinggi. Provinsi dengan tingkat kemiskinan tinggi menjadi prioritas utama dalam distribusi bantuan dan pembangunan, termasuk pembangunan infrastruktur digital seperti internet gratis.

Data diambil dari publikasi resmi BPS, lalu disatukan dalam satu set data dengan 34 entri (mewakili 34 provinsi). Dataset tersebut digunakan dalam proses analisis kluster menggunakan algoritma K-Means, dengan tujuan untuk mengelompokkan provinsi berdasarkan tingkat kemiskinan dan akses internet, guna

menentukan provinsi mana yang paling layak untuk dijadikan prioritas pembangunan internet gratis berbasis data.

### B. Variabel Penelitian

Penelitian ini menggunakan dua variabel utama dari data BPS tahun 2018, yaitu:

- X = Persentase Penduduk Miskin

Variabel ini merepresentasikan tingkat kemiskinan di masing-masing provinsi. Persentase ini menunjukkan seberapa besar proporsi penduduk yang berada di bawah garis kemiskinan. Provinsi dengan persentase penduduk miskin yang tinggi menandakan kondisi ekonomi yang lemah dan keterbatasan akses terhadap layanan dasar, termasuk akses terhadap teknologi. Sebaliknya, provinsi dengan tingkat kemiskinan yang rendah menunjukkan pembangunan yang relatif merata dan tingkat kesejahteraan yang lebih baik.

- Y = Persentase Pengguna Internet

Variabel ini menunjukkan tingkat pemanfaatan internet di masing-masing provinsi. Semakin tinggi persentasenya, semakin luas penggunaan internet di kalangan masyarakat. Provinsi dengan penggunaan internet yang rendah cenderung mengalami kesenjangan digital, baik akibat keterbatasan infrastruktur, biaya, maupun literasi teknologi. Sementara provinsi dengan tingkat penggunaan internet yang tinggi menunjukkan kemajuan dalam adopsi teknologi informasi.

Setiap provinsi kemudian direpresentasikan sebagai satu titik data berdimensi dua (X, Y):

- X = tingkat kemiskinan,
- Y = tingkat penggunaan internet.

Analisis dilakukan dengan K-Means Clustering untuk mengelompokkan provinsi ke dalam beberapa kluster berdasarkan pola hubungan kedua variabel tersebut. Hasil kluster ini akan membantu dalam mengidentifikasi wilayah yang layak menjadi prioritas pembangunan internet gratis.

Misalnya:

- Provinsi dengan kemiskinan tinggi dan penggunaan internet rendah tergolong wilayah yang sangat membutuhkan intervensi, seperti program internet gratis.
- Provinsi dengan kemiskinan rendah namun penggunaan internet rendah, bisa jadi terkendala oleh infrastruktur atau wilayah geografis tertentu.
- Provinsi dengan kemiskinan tinggi namun penggunaan internet tinggi menandakan adanya potensi pemanfaatan teknologi yang cukup baik di tengah keterbatasan ekonomi.
- Sementara itu, provinsi dengan kemiskinan dan penggunaan internet yang sama-sama rendah kemungkinan perlu pendekatan ganda: peningkatan kesejahteraan dan penyediaan akses digital.

Dengan pendekatan ini, hasil penelitian diharapkan dapat memberikan rekomendasi yang berbasis data bagi

perencanaan kebijakan publik yang lebih adil dan tepat sasaran.

### C. Metode Clustering

Penelitian ini menggunakan algoritma K-Means Clustering sebagai metode utama untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan dua variabel, yaitu persentase penduduk miskin dan persentase pengguna internet. K-Means merupakan salah satu algoritma unsupervised learning yang paling populer dalam data mining dan analisis multivariat. Algoritma ini bertujuan untuk membagi sekumpulan data ke dalam K kluster yang berbeda, di mana setiap kluster memiliki karakteristik yang serupa berdasarkan nilai atribut tertentu.

K-Means bekerja dengan cara mencari titik pusat (centroid) untuk setiap kluster, kemudian menghitung jarak setiap titik data ke masing-masing centroid. Setiap data akan dimasukkan ke dalam kluster dengan centroid terdekat, dan proses ini akan diulang hingga posisi centroid tidak berubah secara signifikan atau hingga jumlah iterasi maksimum tercapai. Tujuan akhir dari K-Means adalah meminimalkan total jarak kuadrat (sum of squared distances) antara setiap titik data dan pusat klusternya, sehingga diperoleh kelompok data yang homogen di dalam kluster dan heterogen antar kluster.

Dalam penelitian ini, setiap provinsi direpresentasikan sebagai satu titik data dua dimensi, dengan sumbu X mewakili persentase penduduk miskin dan sumbu Y mewakili persentase pengguna internet. Proses pengelompokan dilakukan untuk mengidentifikasi wilayah-wilayah yang memiliki pola karakteristik sosial dan digital yang serupa, sehingga dapat diberikan perlakuan kebijakan yang berbeda sesuai dengan kelompoknya.

Untuk menentukan jumlah kluster yang optimal, digunakan metode Elbow, yaitu teknik yang memplot nilai Within-Cluster Sum of Squares (WCSS) terhadap berbagai nilai K. Titik siku (elbow) pada grafik menunjukkan jumlah kluster optimal, karena setelah titik tersebut, penurunan WCSS tidak lagi signifikan meskipun jumlah kluster ditambah. Berdasarkan hasil metode Elbow pada data penelitian ini, diperoleh bahwa K = 3 merupakan jumlah kluster yang paling ideal, karena memberikan keseimbangan antara detail segmentasi dan kejelasan kelompok.

Dengan jumlah kluster tersebut, provinsi-provinsi di Indonesia dapat dikelompokkan menjadi tiga kluster besar yang mencerminkan kombinasi tingkat kemiskinan dan tingkat penggunaan internet. Hasil dari kluster ini kemudian dianalisis lebih lanjut untuk menentukan wilayah prioritas pembangunan infrastruktur internet gratis, khususnya pada kluster yang memiliki karakteristik kemiskinan tinggi dan penggunaan internet rendah.

### D. Penerapan Rumus K-Means

#### 1. Rumus Dasar K-Means

Algoritma K-Means adalah metode unsupervised learning yang membagi data menjadi k kluster berdasarkan kemiripan fitur. Langkah utama dalam algoritma ini:

- Inisialisasi: Pilih k centroid secara acak.
- Pengelompokan (Assignment): Hitung jarak setiap titik data ke semua centroid dan kelompokkan ke centroid terdekat.

$$d(x_i, \mu_j) = \sqrt{\sum_{l=1}^n (x_{il} - \mu_{jl})^2}$$

Di mana:

- $x_i$  = data ke-i
  - $\mu_j$  = centroid ke-j
  - n = jumlah fitur
  - Fungsi d = jarak Euclidean
- Pembaruan (Update)
 
$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

Di mana  $C_j$  adalah himpunan anggota kluster ke-j.
  - Ulangi langkah 2–3 sampai posisi centroid tidak berubah (konvergen) atau mencapai iterasi maksimum.

#### 2. Penerapan dalam Penelitian Ini

Fitur yang digunakan:

- $x_1$ : Proporsi Penggunaan Internet (dalam persen)
- $x_2$ : Persentase Penduduk Miskin

Jumlah kluster (k): Ditentukan secara manual menjadi 3, dengan asumsi bahwa Indonesia dapat dikategorikan ke dalam tiga kelompok:

- Tinggi internet, rendah kemiskinan
- Sedang
- Rendah internet, tinggi kemiskinan

Normalisasi Data: Sebelum dihitung, semua data dinormalisasi agar skala kemiskinan dan internet setara, sehingga tidak ada dominasi fitur. Contoh Penghitungan Jarak (misalnya untuk provinsi X dengan internet 20%, kemiskinan 25% ke centroid  $\mu_j = (35, 15)$ ):

$$\begin{aligned} d &= \sqrt{(20 - 35)^2 + (25 - 15)^2} \\ &= \sqrt{225 + 100} \\ &= \sqrt{325} \approx 18.03 \end{aligned}$$

Setiap provinsi dibandingkan ke semua centroid, lalu ditempatkan pada kluster dengan jarak terdekat.

Setelah proses iteratif dijalankan:

- Cluster 0: Rata-rata internet sedang, kemiskinan sedang. Misalnya: Kalimantan Selatan, Sulawesi Tenggara, Jambi

- Cluster 1:  
Internet tinggi (>50%), kemiskinan rendah (<10%). Misalnya: DKI Jakarta, Bali, Kepulauan Riau. Centroid: (52, 6)
- Cluster 2:  
Internet rendah (<25%), kemiskinan tinggi (>20%). Papua, Nusa Tenggara Timur. Centroid: (21, 24). Wilayah prioritas utama pembangunan internet gratis.

#### E. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sistematis yang bertujuan untuk menghasilkan pengelompokan (klasterisasi) provinsi-provinsi di Indonesia berdasarkan tingkat kemiskinan dan penggunaan internet. Setiap tahapan dirancang untuk mendukung validitas hasil dan ketepatan rekomendasi kebijakan. Berikut penjabaran tiap tahap secara lengkap:

- Studi Literatur (Literature Review)  
Tahap awal penelitian dimulai dengan melakukan kajian literatur terhadap berbagai sumber akademik dan laporan resmi. Tujuan dari tahap ini adalah untuk memahami konsep dasar mengenai kemiskinan digital, pemerataan akses internet, serta metode clustering, khususnya algoritma K-Means. Literatur yang dikaji meliputi jurnal ilmiah, buku, serta laporan dari lembaga resmi seperti Badan Pusat Statistik (BPS).
- Pengumpulan dan Integrasi Data  
Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik (BPS) tahun 2018, yaitu:
  - Data persentase penduduk miskin menurut provinsi
  - Data proporsi pengguna internet menurut provinsiKedua jenis data ini kemudian diintegrasikan dalam satu set data (data frame) yang memuat 34 provinsi sebagai observasi dan dua variabel utama sebagai fitur analisis.
- Pra-Pemrosesan dan Normalisasi Data  
Sebelum dilakukan proses analisis lebih lanjut, data melalui tahap pra-pemrosesan yang meliputi pengecekan data hilang, pembersihan (cleaning), dan pengecekan konsistensi antar nilai. Karena kedua variabel memiliki skala yang berbeda, maka dilakukan normalisasi agar data dapat dibandingkan secara setara. Normalisasi ini penting agar algoritma K-Means tidak bias terhadap variabel yang memiliki rentang nilai lebih besar.
- Penentuan Jumlah Kluster Optimal (Metode Elbow)  
Dalam proses clustering, pemilihan jumlah kluster (nilai K) yang tepat sangat krusial. Oleh karena itu, dilakukan analisis dengan metode Elbow, yaitu teknik yang memplot nilai Within-Cluster Sum of Squares (WCSS) terhadap berbagai nilai K. Titik

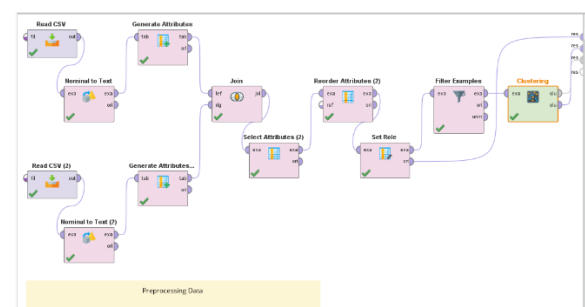
"siku" pada grafik menunjukkan jumlah kluster optimal. Berdasarkan hasil grafik Elbow, nilai K = 3 ditetapkan sebagai jumlah kluster terbaik dalam penelitian ini.

- Pelaksanaan K-Means Clustering di RapidMiner  
Setelah nilai K ditentukan, proses clustering dilakukan menggunakan perangkat lunak RapidMiner, yaitu sebuah platform analisis data yang mendukung implementasi algoritma machine learning secara visual. Data yang telah dinormalisasi dimasukkan ke dalam operator K-Means dan dikonfigurasi sesuai parameter yang telah ditentukan. Proses ini menghasilkan output berupa pembagian provinsi ke dalam tiga kluster berbeda.
- Visualisasi Hasil Klasterisasi  
Untuk memperjelas hasil pengelompokan, dilakukan visualisasi dalam bentuk scatter plot dua dimensi dengan sumbu X sebagai persentase penduduk miskin dan sumbu Y sebagai persentase pengguna internet. Masing-masing provinsi ditampilkan sebagai titik data dan diberi warna sesuai kluster yang terbentuk. Visualisasi ini membantu dalam memahami pola distribusi provinsi berdasarkan kedua variabel tersebut.
- Interpretasi Hasil dan Rekomendasi Kebijakan  
Tahap akhir adalah melakukan interpretasi terhadap hasil kluster yang telah diperoleh. Setiap kluster dianalisis karakteristiknya, terutama fokus pada kluster dengan tingkat kemiskinan tinggi dan penggunaan internet rendah, karena wilayah-wilayah dalam kluster ini menjadi kandidat prioritas dalam pembangunan akses internet gratis yang tepat sasaran. Rekomendasi kebijakan disusun berdasarkan hasil analisis ini agar dapat memberikan dampak yang optimal terhadap peningkatan kesejahteraan masyarakat melalui teknologi digital.

### III. DESAIN, HASIL DAN PEMBAHASAN

#### A. Desain Klasifikasi

Proses desain clustering dilakukan dengan memanfaatkan software RapidMiner menggunakan algoritma K-Means Clustering. Gambar di bawah menunjukkan tahapan preprocessing dan proses clustering secara keseluruhan.

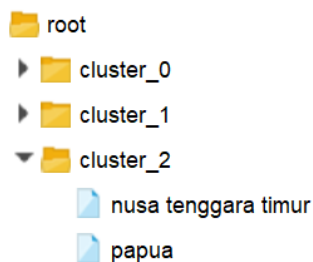


Gambar 3. Desain Klasifikasi Rapidminer

Langkah-langkah pada desain ini dijelaskan sebagai berikut:

1. Read CSV: Dua dataset dari BPS yaitu Proporsi Penggunaan Internet 2018 dan Persentase Penduduk Miskin 2018 dibaca dari file CSV terpisah.
2. Nominal to Text: Operasi ini mengubah atribut nominal (seperti nama provinsi) menjadi format teks agar bisa digabungkan.
3. Generate Attributes: Membuat atribut tambahan atau format kolom agar bisa digunakan untuk join.
4. Join: Kedua dataset digabung berdasarkan atribut provinsi.
5. Select Attributes: Memilih hanya kolom yang diperlukan (proporsi internet dan persentase miskin).
6. Reorder Attributes: Mengatur ulang urutan kolom untuk memudahkan analisis.
7. Set Role: Menentukan atribut yang menjadi label atau identifier (dalam hal ini nama provinsi).
8. Filter Examples: Menghapus entri kosong atau provinsi yang datanya tidak lengkap.
9. Clustering: Data yang telah bersih dan relevan diproses menggunakan algoritma K-Means untuk mengelompokkan provinsi berdasarkan kemiripan nilai dua indikator utama: penggunaan internet dan tingkat kemiskinan.

## B. Visualisasi Data



Gambar 4. Hasil Clustering

Folder ini adalah hasil dari proses K-Means Clustering yang mengelompokkan provinsi-provinsi di Indonesia ke dalam beberapa kluster berdasarkan dua variabel:

1. Proporsi penggunaan internet (%)
2. Persentase penduduk miskin (%)

Setiap folder cluster\_0, cluster\_1, dan cluster\_2 mewakili satu kluster. Di dalam masing-masing folder, terdapat nama-nama provinsi yang masuk ke dalam kluster tersebut.

Isi Folder :

1. cluster\_0/

Berisi provinsi-provinsi yang berada dalam kluster 0, yaitu provinsi dengan akses internet sedang dan kemiskinan relatif rendah.

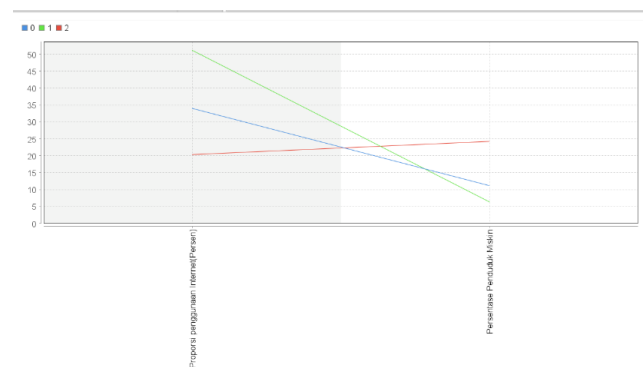
2. cluster\_1/

Berisi provinsi dengan akses internet tinggi dan kemiskinan rendah. Umumnya wilayah maju.

3. cluster\_2/

Berisi provinsi dengan akses internet rendah dan kemiskinan tinggi. Contohnya:

nusa tenggara timur, papua. Kluster ini adalah prioritas utama dalam pembangunan internet gratis berbasis pemerataan digital.



Gambar 5. Visualisasi Hasil Clustering (Kemiskinan vs Internet)

Kluster dapat dijelaskan sebagai berikut :

TABEL I Penjelasan Hasil Clustering (Kemiskinan vs Internet)

Kluster	Warna	Karakteristik Utama
0	Biru	Penggunaan internet sedang, kemiskinan menengah
1	Hijau	Penggunaan internet tinggi, kemiskinan rendah (wilayah maju)
2	Merah	Penggunaan internet rendah, kemiskinan tinggi (wilayah rentan)

- Biru (Cluster 0):
  - Rata-rata proporsi penggunaan internet: ~33%
  - Rata-rata penduduk miskin: ~12%
  - Ciri-ciri: Provinsi dengan penggunaan internet sedang dan kemiskinan rendah-sedang
- Hijau (Cluster 1):
  - Rata-rata proporsi penggunaan internet: ~52%
  - Rata-rata penduduk miskin: ~6%
  - Ciri-ciri: Provinsi dengan akses internet tinggi dan kemiskinan rendah
  - Ini termasuk provinsi maju seperti DKI Jakarta, Bali, dll.
- Merah (Cluster 2):
  - Rata-rata proporsi penggunaan internet: ~21%

- Rata-rata penduduk miskin: ~24%
- Ciri-ciri: Provinsi dengan akses internet rendah dan kemiskinan tinggi
- Isinya seperti Papua dan Nusa Tenggara Timur (NTT)

C. Table Hasil Kluster

TABEL II Table hasil kluster

Provinsi	% Kemiskinan	% Internet	Kluster
Papua	27.430	19.590	2
NTT	21.030	21.100	2
Dki Jakarta	3.550	65.890	1
Yogyakarta	11.810	55.450	1
Jawa Barat	7.250	45.330	1
Sumatera Utara	8.940	34.270	0
Sumatera Barat	6.550	36.490	0
-----	-----	-----	-----

D. Interpretasi

Provinsi dalam Kluster 2 seperti Papua, Papua Barat, dan NTT menjadi kandidat prioritas pembangunan internet gratis karena dua alasan utama:

- Masyarakatnya miskin dan belum mendapat akses digital yang memadai.
- Potensi pengembangan ekonomi digital tinggi jika difasilitasi dengan internet.

Sebaliknya, provinsi dalam Kluster 0 tidak menjadi prioritas utama karena akses sudah tinggi dan tingkat kemiskinan rendah.

Lebih lengkap :

- Cluster 1 (Hijau): Target bukan prioritas untuk pembangunan internet gratis. Mereka sudah maju.
- Cluster 0 (Biru): Menengah, bisa dipertimbangkan sebagai target sekunder.
- Cluster 2 (Merah): Prioritas utama pembangunan internet gratis, karena kondisi ganda: miskin + internet rendah.

Hasil klustering menunjukkan bahwa terdapat satu kluster (Cluster 2) yang berisi provinsi dengan tingkat kemiskinan yang tinggi serta akses internet yang rendah, yaitu Papua dan Nusa Tenggara Timur. Kluster ini menjadi sasaran utama dalam upaya pembangunan internet gratis berbasis data. Sedangkan Cluster 1 merupakan wilayah dengan kondisi sosial digital paling baik, dan Cluster 0 berada di kategori menengah.

E. Validasi & Keterbatasan

Model K-Means Clustering merupakan algoritma yang efektif dan efisien dalam mengelompokkan data berbasis kedekatan jarak antar titik (dalam hal ini, provinsi) berdasarkan atribut numerik. Pada proyek ini, algoritma K-Means digunakan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan proporsi penggunaan internet (%) dan persentase penduduk miskin (%). Dengan hanya menggunakan dua dimensi, visualisasi kluster menjadi lebih intuitif dan mudah

dianalisis, baik melalui grafik sebar maupun plot centroid antar kelompok.

Namun, meskipun bekerja optimal pada data berdimensi rendah, model K-Means memiliki beberapa keterbatasan:

1. Sensitivitas terhadap Outlier

Algoritma ini cukup sensitif terhadap data yang memiliki nilai ekstrem (outlier), karena perhitungan pusat kluster (centroid) sangat bergantung pada posisi data. Dalam konteks ini, DKI Jakarta dapat dikategorikan sebagai outlier karena memiliki proporsi penggunaan internet yang sangat tinggi dibanding provinsi lainnya, serta tingkat kemiskinan yang sangat rendah. Kedua hal tersebut bisa menyebabkan pergeseran centroid kluster.

2. Mitigasi dengan Normalisasi

Untuk mengurangi efek negatif dari keberadaan outlier, data pada proyek ini telah melalui proses normalisasi, yaitu penyetaraan skala antar fitur (fitur internet dan kemiskinan) agar memiliki bobot yang setara dalam perhitungan jarak. Hal ini membantu mengurangi dominasi variabel tertentu serta membuat distribusi data lebih seragam. Oleh karena itu, meskipun terdapat outlier seperti DKI Jakarta, dampaknya terhadap pembentukan kluster secara keseluruhan menjadi minimal dan tidak mendistorsi hasil klustering secara signifikan.

3. Keterbatasan Jumlah Kluster (K)

Penentuan jumlah kluster (k) juga menjadi tantangan tersendiri. Pada studi ini digunakan pendekatan metode elbow untuk menetapkan jumlah kluster optimal, yaitu k = 3. Namun, pemilihan k ini tetap bersifat subjektif dan bisa berubah tergantung konteks dan tujuan analisis.

4. Keterbatasan Konteks Sosial dan Geografis

Algoritma tidak mempertimbangkan faktor-faktor sosial, geografis, atau kebijakan lokal yang mungkin berdampak besar pada pembangunan infrastruktur internet atau tingkat kemiskinan. Misalnya, Papua memiliki tantangan geografis dan aksesibilitas yang tidak tertangkap langsung dalam data numerik.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa metode K-Means Clustering dapat digunakan untuk mengidentifikasi wilayah prioritas pembangunan internet gratis berbasis data. Dengan menggunakan data kemiskinan dan penggunaan internet dari BPS 2018, provinsi-provinsi di Indonesia berhasil dikelompokkan ke dalam tiga kluster.

Provinsi dalam Kluster 2 yang mencakup Papua, dan Nusa Tenggara Timur sebaiknya dijadikan target

utama pembangunan infrastruktur internet gratis untuk mengurangi kesenjangan digital dan sosial. Pendekatan berbasis data ini mendukung efisiensi dalam pengambilan kebijakan publik.

Penelitian ini dapat dikembangkan lebih lanjut dengan menambahkan variabel lain seperti indeks pembangunan manusia, kualitas pendidikan, dan infrastruktur jaringan untuk membentuk model yang lebih kompleks dan holistik

#### V. REFERENSI

- [1] Badan Pusat Statistik, "Statistik Kesejahteraan Rakyat 2018," [www.bps.go.id](http://www.bps.go.id), 2018.
- [2] Badan Pusat Statistik, "Statistik Telekomunikasi Indonesia 2018," [www.bps.go.id](http://www.bps.go.id), 2018.
- [3] H. Purwanto, "Peran Teknologi Informasi dalam Peningkatan Ekonomi Wilayah Tertinggal," *Jurnal Teknologi dan Pembangunan*, vol. 5, no. 2, 2019.
- [4] S. Hidayat dan R. Anjani, "Pemanfaatan Internet Gratis untuk Pendidikan di Daerah 3T," *Seminar Nasional Teknologi Informasi*, 2020.
- [5] A. Nugroho, "Clustering Daerah Berdasarkan Tingkat Kemiskinan Menggunakan K-Means," *Jurnal Informatika*, vol. 12, 2021.
- [6] RapidMiner Documentation. [Online]. Available: <https://docs.rapidminer.com>
- [7] M. Prasetyo, "Pembangunan Infrastruktur Digital dan Pemerataan Teknologi," *Kominfo Journal*, vol. 8, 2020.
- [8] B. Siswanto, "Digital Divide dan Akses Internet di Indonesia Timur," *Jurnal Pembangunan Daerah*, vol. 7, no. 1, 2019.
- [9] R. Kurniawan, "Analisis K-Medoids dalam Pemodelan Data Sosial," *Jurnal Sains Data*, 2022.
- [10] L. Syahril, "Model Penilaian Prioritas Wilayah Pembangunan Teknologi," *Tesis Magister*, Universitas Indonesia, 2021.
- [11] Hudaya Latuconsinaa, Khusainia , & Sri Jaya Lesmanaa, "Pendidikan dan Penggunaan Internet Menurunkan Kemiskinan di Banten" *Jurnal Ekonomi dan Pembangunan Indonesia* Vol. 24 No. 2 Juli 2024
- [12] Mita Amelia, Ahmad Faqih, & Ade Rizki Rinaldi. "PENERAPAN METODE K-MEANS CLUSTERING DALAM PEMETAAN KEMISKINAN KABUPATEN/KOTA DI INDONESIA UNTUK PERENCANAAN KEBIJAKAN YANG TEPAT" *Jurnal Informatika Dan Teknik Elektro Terapan*, 13. <https://doi.org/10.23960/jitet.v13i2.6231>