

Laporan Analisis Dashboard KUPAS : Analisis Faktor Pendidikan dan Infrastruktur terhadap Tingkat Kemiskinan di Kabupaten/Kota di Indonesia



Dosen Pengampu:

Yuliagnis Transver Wijaya, S.ST, M.Sc.

Disusun Oleh:

Qurany Nadhira Tsabita 222313323

PROGRAM STUDI D-IV KOMPUTASI STATISTIK

POLITEKNIK STATISTIKA STIS

TAHUN AJARAN 2024/2025

DAFTAR ISI

DAFTAR ISI.....	2
BAB 1	3
PENDAHULUAN	3
1.1 Latar Belakang	3
1.2 Rumusan Masalah	3
1.3 Tujuan Penulisan	4
BAB 2	5
METODOLOGI.....	5
2.1 Sumber Data	5
2.2 Metode Analisis.....	5
BAB 3	7
HASIL DAN PEMBAHASAN.....	7
3.1 Identifikasi Kebutuhan Data.....	7
3.2 Pengambilan Data.....	7
3.3 Pengintegritasan Data.....	8
3.4 Penelaahan Data	8
3.5 Pemvalidasian Data	9
3.6 Penentuan Objek Data	10
3.7 Bussiness Inteligence	11
3.7.1 Menu Beranda.....	11
3.7.2 Menu Manajemen Data.....	12
3.7.3 Menu Eksplorasi Data.....	14
3.7.4 Menu Analisis Data	18
BAB 4	26
PENUTUP.....	26
4.1 Kesimpulan.....	26
4.2 Saran.....	27
DAFTAR PUSTAKA	28

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kemiskinan merupakan permasalahan multidimensional yang tidak hanya berkaitan dengan aspek ekonomi, tetapi juga berhubungan erat dengan akses terhadap infrastruktur dasar serta kualitas sumber daya manusia. Di tingkat kabupaten/kota, ketimpangan akses terhadap layanan dasar seperti listrik, air bersih, dan sistem sanitasi yang layak sering kali menjadi indikator keterbelakangan pembangunan yang turut memengaruhi tingkat kemiskinan masyarakat.

Selain infrastruktur, faktor pendidikan memiliki peran penting dalam menentukan kapasitas individu dan rumah tangga untuk keluar dari lingkaran kemiskinan. Tingginya persentase penduduk dengan pendidikan rendah dan tingkat buta huruf yang signifikan menunjukkan adanya keterbatasan dalam akses informasi, peluang kerja, dan produktivitas, yang pada akhirnya berdampak terhadap kesejahteraan ekonomi.

Dengan mempertimbangkan hal tersebut, studi ini berfokus untuk menelaah apakah terdapat hubungan yang signifikan antara faktor pendidikan dan akses infrastruktur dengan tingkat kemiskinan di wilayah kabupaten/kota di Indonesia. Analisis dilakukan menggunakan pendekatan regresi linear berganda dengan lima variabel independen, yaitu: persentase rumah tangga tanpa listrik, persentase rumah tangga tanpa sistem drainase, persentase rumah tangga pengguna air bersih melalui pipa, persentase penduduk berusia ≥ 15 tahun dengan pendidikan rendah, dan persentase penduduk yang buta huruf. Melalui pendekatan ini, diharapkan dapat diperoleh gambaran yang lebih jelas mengenai kontribusi masing-masing faktor terhadap variasi tingkat kemiskinan di berbagai daerah, sehingga dapat menjadi dasar pertimbangan dalam perumusan kebijakan pembangunan yang lebih inklusif dan berbasis data.

1.2 Rumusan Masalah

- a. Bagaimana persebaran tingkat kemiskinan serta variabel-variabel terkait infrastruktur dan pendidikan di tingkat kabupaten/kota di Indonesia?
- b. Apakah terdapat pola pengelompokan wilayah berdasarkan karakteristik sosial ekonomi tersebut?
- c. Apakah variabel-variabel seperti akses listrik, sistem drainase, penggunaan air pipa, pendidikan rendah, dan buta huruf berpengaruh secara signifikan terhadap tingkat kemiskinan suatu daerah?
- d. Sejauh mana model regresi linear berganda dapat menjelaskan variasi tingkat kemiskinan berdasarkan faktor-faktor pendidikan dan infrastruktur?

1.3 Tujuan Penulisan

- a. Menyajikan visualisasi interaktif mengenai persebaran variabel kemiskinan, pendidikan, dan akses infrastruktur, termasuk identifikasi wilayah dengan nilai tertinggi dan terendah.
- b. Menganalisis pola pengelompokan wilayah berdasarkan karakteristik sosial ekonomi menggunakan metode clustering spasial.
- c. Melakukan pengujian asumsi klasik dan analisis statistik inferensial terhadap model regresi linear berganda.
- d. Menilai sejauh mana variabel-variabel terpilih mampu menjelaskan variasi tingkat kemiskinan, serta mengidentifikasi faktor-faktor yang memiliki pengaruh signifikan.

BAB 2

METODOLOGI

2.1 Sumber Data

Data yang digunakan dalam analisis ini bersumber dari hasil penelitian yang dilakukan oleh R. Kurniawan, B.I. Nasution, N. Agustina et al (2022) yang dipublikasikan dalam jurnal *Data in Brief* dengan judul "*Revisiting social vulnerability analysis in Indonesia: An optimized spatial fuzzy clustering approach*". Penelitian tersebut menyediakan dataset komprehensif yang mencakup 511 kabupaten/kota di Indonesia, dan memuat berbagai indikator sosial ekonomi yang relevan dalam kajian kerentanan sosial dan kemiskinan. Dataset tersebut dapat diakses secara terbuka melalui tautan https://raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/sovi_data.csv.

Selain data utama, analisis spasial dalam dashboard ini turut menggunakan matriks penimbang jarak antarwilayah yang berguna dalam mengidentifikasi pola pengelompokan geografis. Matriks ini juga tersedia secara daring melalui tautan <https://raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/distance.csv>.

Informasi metadata, termasuk penjelasan definisi masing-masing variabel dan metodologi pengumpulan data, dapat ditelusuri lebih lanjut pada publikasi asli hasil penelitian yang ada di laman <https://www.sciencedirect.com/science/article/pii/S2352340921010180>.

2.2 Metode Analisis

Metode analisis yang digunakan dalam pengembangan dashboard ini mengacu pada tahapan komputasi statistik yang terstruktur, dimulai dari identifikasi data hingga visualisasi serta interpretasi hasil. Proses analisis dilakukan secara sistematis untuk menggali informasi yang bermakna dari data sosial ekonomi tingkat kabupaten/kota, khususnya terkait dengan faktor pendidikan, infrastruktur, dan kemiskinan.

Tahapan awal dimulai dengan identifikasi dan pengenalan data, yang dilakukan terhadap dataset SOVI yang diperoleh dari sumber penelitian. Pada tahap ini dilakukan pemahaman terhadap struktur data, jenis variabel (numerik dan kategorik), serta skala pengukurannya. Data kemudian diekstrak dan diperiksa menggunakan perangkat lunak Microsoft Excel untuk memastikan kelengkapan dan konsistensinya. Berdasarkan hasil pemeriksaan, dataset tidak mengandung nilai hilang (missing values) sehingga dapat langsung digunakan dalam proses analisis lebih lanjut tanpa perlu pembersihan data tambahan.

Selanjutnya, proses pengolahan data dan pengembangan dashboard interaktif dilakukan menggunakan platform R Shiny. Dalam dashboard ini, dilakukan proses pengkategorian pada beberapa variabel numerik guna mempermudah eksplorasi visual, serta disediakan beragam bentuk visualisasi data, seperti barchart, histogram, dan boxplot untuk menampilkan sebaran variabel. Selain itu, dashboard juga dilengkapi dengan peta distribusi spasial untuk

menggambarkan persebaran nilai antar wilayah serta peta klaster untuk mengevaluasi kemungkinan pengelompokan berdasarkan kemiripan karakteristik sosial ekonomi.

Untuk memberikan pemahaman kuantitatif terhadap data, disajikan pula statistik deskriptif dari setiap variabel, yang mencakup nilai minimum, maksimum, rerata, dan standar deviasi. Analisis ini bertujuan untuk memberikan gambaran umum mengenai distribusi dan kecenderungan data. Dashboard ini juga menyediakan fitur untuk melakukan analisis statistik inferensial dan regresi linear berganda, yang dilengkapi dengan uji asumsi klasik seperti normalitas residual, multikolinearitas, dan heteroskedastisitas. Pendekatan ini digunakan untuk menilai apakah variabel-variabel yang merepresentasikan aspek pendidikan dan infrastruktur berpengaruh secara signifikan terhadap tingkat kemiskinan di kabupaten/kota.

BAB 3

HASIL DAN PEMBAHASAN

3.1 Identifikasi Kebutuhan Data

Tersedianya data sovi dari 511 kabupaten/kota di Indonesia memberikan peluang untuk melakukan analisis komprehensif terhadap berbagai faktor yang memengaruhi tingkat kemiskinan antarwilayah. Sesuai dengan tujuan utama analisis, fokus diarahkan pada identifikasi variabel-variabel yang mewakili aspek pendidikan dan akses infrastruktur, guna mengkaji sejauh mana keterkaitan faktor-faktor tersebut terhadap kemiskinan.

Dalam tahap awal analisis, dilakukan proses *business understanding* untuk merumuskan arah dan sasaran eksplorasi data. Proses ini mencakup identifikasi aset data yang tersedia, relevansi tiap variabel terhadap tujuan penelitian, serta bagaimana data akan dikelola dan dianalisis secara efektif dalam konteks pengembangan dashboard interaktif. Dari keseluruhan variabel dalam dataset, telah dipilih sejumlah variabel utama yang dinilai paling relevan dalam menjelaskan variasi kemiskinan. Variabel-variabel tersebut meliputi: persentase rumah tangga tanpa akses listrik, rumah tangga tanpa sistem drainase, rumah tangga pengguna air bersih melalui pipa, persentase penduduk berpendidikan rendah, dan persentase penduduk buta huruf. Variabel-variabel ini dipertimbangkan karena merepresentasikan dua dimensi penting dalam pembangunan sosial, yaitu infrastruktur dasar dan pendidikan.

Proses identifikasi juga mencakup pengelompokan dan kategorisasi variabel numerik tertentu agar lebih mudah dianalisis dan divisualisasikan. Selain itu, dilakukan seleksi terhadap variabel yang memungkinkan pemetaan spasial, seperti nama wilayah administratif, untuk digunakan dalam visualisasi berbasis peta. Informasi-informasi ini selanjutnya disiapkan untuk mendukung pengembangan fitur visualisasi dalam dashboard, termasuk diagram batang, histogram, boxplot, peta distribusi variabel, serta peta kluster antarwilayah berdasarkan kesamaan karakteristik sosial ekonomi. Identifikasi kebutuhan data ini menjadi fondasi awal dalam menyusun struktur analisis dan desain antarmuka dashboard. Seluruh langkah ini bertujuan agar informasi yang ditampilkan tidak hanya akurat secara statistik, tetapi juga informatif dan bermakna dalam konteks pengambilan keputusan publik yang berbasis data.

3.2 Pengambilan Data

Pengambilan data dalam penelitian ini dilakukan melalui pemanfaatan sumber data terbuka dari proyek *naspaclust* oleh Kurniawan et al. (2022), yang menyediakan data sosial ekonomi wilayah kabupaten/kota di Indonesia. Dataset utama yang digunakan diunduh dalam format CSV melalui tautan https://raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/sovi_data.csv. Dataset ini berisi data dari 511 kabupaten/kota di Indonesia dan mencakup berbagai indikator sosial ekonomi

seperti persentase rumah tangga tanpa listrik, tanpa drainase, tanpa akses air bersih melalui pipa, tingkat kemiskinan, tingkat pendidikan rendah, serta variabel sosial demografis lainnya.

Untuk keperluan pemetaan dan analisis spasial, data CSV tersebut diintegrasikan dengan data batas administratif dalam format GeoJSON yang telah disederhanakan guna efisiensi pemrosesan. Integrasi ini memungkinkan visualisasi data dalam bentuk peta interaktif pada dashboard. File GeoJSON yang digunakan, yaitu `sovi_data_super_simplified.geojson`, diimpor menggunakan fungsi `st_read()` dari library `sf`, sedangkan dataset CSV diakses langsung melalui fungsi `read.csv()`.

Selain itu, untuk mendukung analisis spasial lanjutan seperti pemetaan klaster, digunakan juga matriks jarak antar wilayah kabupaten/kota dari tautan <https://raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/distance.csv>. Matriks ini penting dalam menghitung kedekatan spasial antar wilayah yang menjadi dasar dalam identifikasi pola pengelompokan wilayah.

Seluruh proses pengambilan, integrasi, dan pemrosesan data dilakukan secara langsung di lingkungan R melalui pengembangan aplikasi berbasis Shiny. Beberapa library yang digunakan dalam tahap ini meliputi `sf` untuk pengolahan data spasial, `leaflet` untuk visualisasi peta interaktif, dan `dplyr` untuk manipulasi data. Selain itu, pengembangan antarmuka dan fungsi interaktif dashboard dilakukan dengan memanfaatkan `shiny`, `shinyjs`, dan `bs4Dash` yang secara keseluruhan mendukung proses pengolahan data hingga penyajiannya dalam bentuk yang informatif dan interaktif.

3.3 Pengintegritasan Data

Setelah data berhasil diakses dari sumber daring, proses berikutnya adalah memastikan integritas data yang akan dianalisis, baik dari sisi struktur, kelengkapan, maupun keterkaitannya dengan data spasial. Data utama yang digunakan berasal dari https://raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/sovi_data.csv, yang merupakan versi terstruktur dari data indeks sosial dan kerentanan wilayah (SOVI) di Indonesia. Data ini kemudian diintegrasikan dengan file spasial berformat GeoJSON yang berisi batas administratif 511 kabupaten/kota untuk memungkinkan visualisasi peta dan analisis spasial. Proses integrasi dilakukan secara langsung melalui pemanggilan fungsi `read.csv()` dan `st_read()` pada R Shiny, tanpa perlu pengolahan tambahan karena data yang diakses tidak mengandung nilai kosong (NA) ataupun entri duplikat.

Pemeriksaan integritas data dilakukan dengan memverifikasi keterpaduan antaratribut dalam file CSV dengan koordinat spasial di GeoJSON. Hal ini dilakukan untuk menjamin bahwa setiap baris data memiliki padanan wilayah yang sesuai secara geografis. Selain itu, penggabungan dilakukan menggunakan variabel kunci yang konsisten, yaitu kode wilayah (ID/Kode Kab/Kota) yang telah distandarisasi. Proses ini berjalan mulus berkat dokumentasi data yang baik dan format yang bersih, sehingga memungkinkan pemrosesan lebih lanjut tanpa hambatan teknis berarti. Dengan struktur data yang telah terintegrasi secara spasial dan bebas dari anomali umum seperti nilai hilang atau ketidaksesuaian struktur, dataset dinyatakan siap untuk tahap eksplorasi, visualisasi, dan analisis lanjutan.

3.4 Penelaahan Data

Setelah data dinyatakan layak dan terintegrasi dengan baik, tahap selanjutnya adalah melakukan penelaahan terhadap isi dataset secara lebih mendalam. Sebelum masuk ke tahap interaktif di aplikasi,

dilakukan pula pengecekan awal di *RStudio* terkait sebaran data dan model yang akan diajukan. Pemeriksaan ini dilakukan untuk memahami karakteristik data, mengamati potensi pencilan atau outlier, serta mengevaluasi bentuk distribusi variabel secara umum. Langkah ini penting sebagai dasar dalam memilih jenis transformasi maupun pendekatan model regresi yang sesuai.

Penelaahan lebih lanjut mencakup identifikasi tipe data, eksplorasi relasi antar variabel, serta penyajian karakteristik data dalam bentuk grafik dan tabel interaktif. Tujuannya adalah memperoleh pemahaman yang utuh mengenai struktur dan pola dasar dalam data sebelum memasuki tahap analisis lanjutan.

Langkah pertama dalam proses ini adalah mengidentifikasi tipe data dari seluruh variabel. Karena pemrosesan dilakukan menggunakan perangkat lunak R, struktur data diperiksa melalui fungsi seperti `str()` dan `glimpse()` dari paket `dplyr`. Mayoritas variabel merupakan data numerik berskala rasio, seperti persentase rumah tangga tanpa listrik, tanpa drainase, dan variabel sosial ekonomi lainnya. Di sisi lain, atribut seperti nama wilayah dan kode kabupaten/kota termasuk dalam tipe data karakter atau faktor, tergantung kebutuhan penggunaannya dalam analisis visual dan tabular.

Untuk mendukung eksplorasi awal terhadap distribusi nilai variabel, dilakukan pengkategorian terhadap variabel numerik ke dalam kelas-kelas tertentu menggunakan metode *Interquartile Range (IQR)*. Pengkategorian ini diimplementasikan langsung pada server aplikasi Shiny, khususnya dalam blok kode `output$table_kategori <- DT::renderDataTable({ ... })`. Dalam blok tersebut, data hasil kategorisasi ditampilkan secara interaktif melalui *DataTable*, baik secara keseluruhan maupun per variabel sesuai input pengguna. Kategori disusun dengan menyisipkan kolom tambahan pada fitur di halaman manajemen data yang mengacu pada pengkategorian nilai sesuai dengan formula IQR, lalu ditata ulang agar tersaji dalam format yang mudah dibaca.

3.5 Pemvalidasian Data

Proses validasi data dalam penelitian ini dilakukan secara bertahap bersamaan dengan proses eksplorasi awal dan penyesuaian struktur data di lingkungan R. Mengingat data yang digunakan telah tersedia dalam format siap pakai dari sumber terbuka dan terpercaya, fokus validasi lebih diarahkan pada pengecekan konsistensi struktur data, pemeriksaan keberadaan duplikasi, serta evaluasi terhadap kewajaran sebaran nilai tiap variabel.

Langkah awal validasi dilakukan melalui fungsi `summary()`, `str()`, dan `skim()` untuk memastikan bahwa tidak terdapat variabel dengan nilai kosong (NA) atau tipe data yang tidak sesuai dengan semestinya. Berdasarkan hasil inspeksi, tidak ditemukan missing values pada variabel-variabel utama yang digunakan dalam analisis regresi maupun visualisasi spasial. Seluruh data numerik terdeteksi

dalam format yang konsisten dan berada dalam rentang nilai yang wajar sesuai konteks kabupaten/kota di Indonesia.

Selain itu, tidak ditemukan kasus duplikat berdasarkan identifikasi nama wilayah dan kode unik kabupaten/kota. Variabel kategorikal seperti provinsi, nama kabupaten/kota, dan kode wilayah juga telah dikonfirmasi bebas dari inkonsistensi penamaan atau pengkodean ganda.

Pengecekan tambahan terhadap keberadaan outlier dilakukan secara visual menggunakan histogram dan boxplot di RStudio. Dari hasil tersebut, ditemukan beberapa nilai ekstrem pada variabel tertentu seperti POPULATION atau PENDIDIKAN_RENDAH. Namun setelah dianalisis lebih lanjut, nilai-nilai tersebut merupakan representasi nyata dari ketimpangan antar wilayah (contoh: kota besar vs daerah tertinggal), sehingga tetap dipertahankan dalam analisis karena mencerminkan kondisi sosial ekonomi yang valid.

Dengan melalui tahapan validasi tersebut, data dinyatakan telah bersih, terstruktur, dan layak untuk dianalisis lebih lanjut. Tidak dilakukan proses imputasi karena tidak ditemukan nilai kosong atau tidak logis dalam dataset. Seluruh variabel telah memenuhi kelayakan teknis untuk digunakan dalam pembangunan model regresi serta visualisasi spasial berbasis wilayah.

3.6 Penentuan Objek Data

Penentuan objek data dalam analisis ini dilakukan secara terarah dengan mempertimbangkan tujuan utama eksplorasi, yaitu menyusun analisis komprehensif mengenai faktor-faktor yang memengaruhi tingkat kemiskinan di wilayah kabupaten/kota di Indonesia. Unit analisis yang digunakan mencakup seluruh kabupaten/kota yang tercatat dalam data sosial ekonomi tahun 2017, berjumlah 511 wilayah administratif. Seleksi variabel difokuskan pada dua aspek utama yang diasumsikan memiliki keterkaitan erat dengan kondisi kemiskinan, yaitu **akses terhadap infrastruktur dasar** dan **tingkat pendidikan masyarakat**.

Dalam konteks infrastruktur, variabel yang digunakan meliputi NOELECTRIC (persentase rumah tangga tanpa listrik), TAPWATER (akses terhadap air bersih melalui pipa), dan NOSEWER (tidak memiliki sistem pembuangan air limbah yang layak). Sementara dalam dimensi pendidikan, variabel yang dipertimbangkan adalah LOWEDU (penduduk dengan pendidikan rendah) dan ILLITERATE (penduduk buta huruf). Variabel POVERTY diposisikan sebagai variabel dependen (respon) yang mencerminkan tingkat kemiskinan daerah. Meskipun dataset memuat variabel-variabel lain seperti CHILDREN, FEMALE, FHEAD, DPRONE, dan NOTRAINING, analisis utama pada tahap awal ini berfokus hanya pada variabel-variabel yang relevan langsung terhadap tema infrastruktur dan pendidikan. Pemilihan ini bertujuan untuk menjaga fokus analisis tetap konsisten, serta memudahkan dalam penyusunan visualisasi dan interpretasi hasil.

Seluruh kabupaten/kota tetap dipertahankan sebagai observasi, tanpa dilakukan penyaringan berdasarkan provinsi atau wilayah tertentu, agar analisis dapat memberikan gambaran yang utuh terhadap ketimpangan sosial ekonomi di tingkat nasional. Variabel nmkab dan kddistrict tetap dipertahankan sebagai *identifier* untuk keperluan pelacakan wilayah, pengelompokan spasial, dan

penyusunan visualisasi interaktif berbasis peta, namun tidak dilibatkan langsung dalam model kuantitatif.

3.7 Bussiness Intelligence

Sebagai bentuk implementasi dari proses analisis data sosial ekonomi yang telah dilakukan, sistem Business Intelligence dikembangkan dalam bentuk dashboard interaktif KUPAS (*Kemiskinan, Infrastruktur, dan Pendidikan di Kabupaten/Kota Indonesia*). Dashboard terdiri dari beberapa tab menu utama yang saling terhubung, masing-masing menyajikan informasi berbeda namun saling melengkapi. Berikut ini adalah penjelasan tiap subbagian:

3.7.1 Menu Beranda

Menu **Beranda** merupakan tampilan awal dari dashboard **KUPAS (Kemiskinan, Pendidikan, dan Akses Infrastruktur)** yang dirancang sebagai *landing page* informatif bagi pengguna. Halaman ini menampilkan pengantar umum mengenai tujuan analisis serta ruang lingkup variabel yang digunakan, sehingga pengguna memiliki pemahaman kontekstual sebelum mengeksplorasi data lebih lanjut. Secara garis besar, tampilan beranda dibagi menjadi tiga bagian utama:

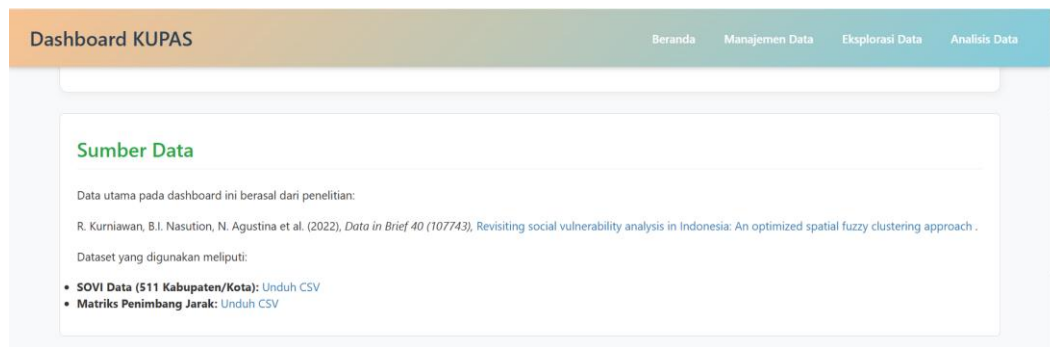
1. Deskripsi Umum Dashboard

Di bagian atas halaman, terdapat penjelasan mengenai latar belakang dan fokus utama dashboard. Analisis yang disajikan berfokus pada faktor-faktor yang memengaruhi tingkat kemiskinan di wilayah Kabupaten/Kota di Indonesia, dengan menitikberatkan pada dua aspek utama, yaitu akses Infrastruktur, pendidikan.



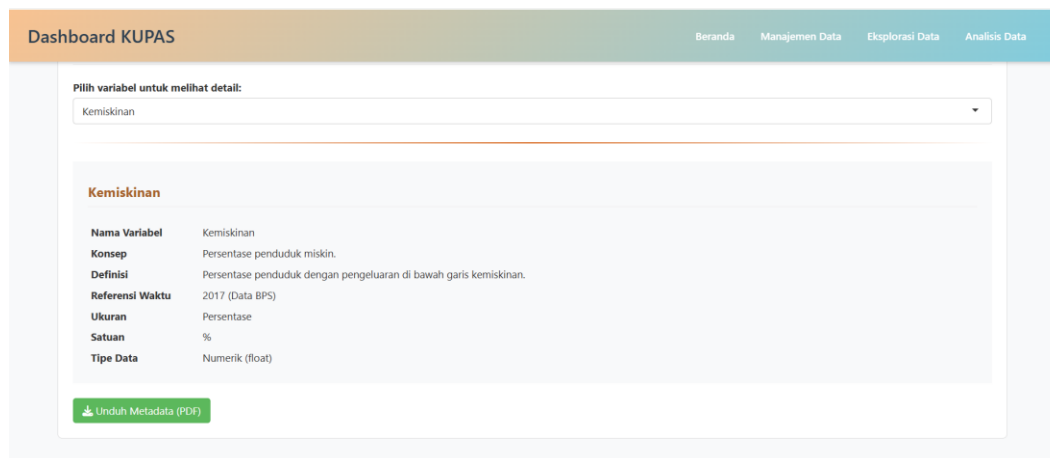
2. Sumber Data

Bagian ini mencantumkan sumber dataset yang digunakan dalam dashboard. Dataset utama diambil dari publikasi ilmiah oleh Kurniawan dkk. (2022) yang memuat data sosial dari 511 kabupaten/kota di Indonesia. Tautan unduhan untuk dataset utama dan matriks penimbang jarak juga disediakan, sehingga pengguna bisa mengakses data mentah secara langsung.



3. Eksplorasi Metadata Variabel

Fitur interaktif ini memungkinkan pengguna untuk memilih dan menelusuri informasi deskriptif dari setiap variabel dalam dataset. Pengguna dapat memilih nama variabel melalui dropdown, lalu sistem akan menampilkan penjelasan terkait deskripsi, skala pengukuran, serta interpretasi. Tersedia pula opsi unduhan metadata dalam format PDF.



3.7.2 Menu Manajemen Data

Menu Manajemen Data berfungsi sebagai pusat pengelolaan dataset sebelum dilakukan analisis lanjutan. Terdapat dua subfitur utama dalam menu ini, yaitu Kategorisasi Variabel dan Tabel Data, yang masing-masing memiliki peran penting dalam proses eksplorasi dan persiapan data. Secara garis besar, menu Manajemen Data terdiri dari dua sub-halaman yang dirancang untuk membantu pengguna memahami dan menyiapkan data sebelum masuk ke tahap analisis yang lebih mendalam.

1. Kategorisasi Variabel

Halaman ini memungkinkan pengguna untuk mengubah variabel numerik kontinyu menjadi variabel kategori ordinal (*Rendah / Menengah / Tinggi*) menggunakan metode pembagian kuartil. Transformasi ini bertujuan untuk menyederhanakan interpretasi serta memfasilitasi analisis deskriptif dan komparatif antar wilayah. Fitur-fitur yang tersedia dalam subpage ini meliputi:

- Dropdown pilihan variabel: Memungkinkan pengguna memilih variabel yang ingin dikategorisasi, atau menampilkan seluruh variabel sekaligus.
- Tabel interaktif: Menampilkan hasil kategorisasi dalam bentuk tabel dinamis.
- Tombol unduh (CSV): Memungkinkan pengguna menyimpan hasil kategorisasi untuk keperluan analisis lanjutan di luar dashboard.

Dashboard KUPAS Beranda Manajemen Data Eksplorasi Data Analisis Data

Gunakan fitur ini untuk mengubah variabel kontinu menjadi kategori Rendah / Menengah / Tinggi berdasarkan kuartil.

Pilih Variabel untuk Ditampilkan:

[Download Data \(CSV\)](#)

Show entries Search:

	Kode	Nama Wilayah	Kemiskinan	Kategori Kemiskinan	Rumah Tangga Tanpa Listrik	Kategori Rumah Tangga Tanpa Listrik	Rumah Tangga Guna Air Pipa	Kategori Rumah Tangga Guna Air Pipa	Rumah Tangga Tanpa Drainase	Kategori Rumah Tangga Tanpa Drainase	Pendidikan Rendah	Kategori Pendidikan Rendah
1	5305	TIMOR TENGAH UTARA	23.52	Tinggi	26.71624211	Tinggi	2.898932897	Rendah	11.72406519	Menengah	39.7729225	Tinggi
2	5308	LEMBATA	26.48	Tinggi	19.88028683	Tinggi	17.22775887	Menengah	9.027102577	Menengah	41.15813577	Tinggi
3	5307	ALOR	21.67	Tinggi	17.60029926	Tinggi	7.33657533	Menengah	15.35584027	Menengah	37.35738588	Menengah
4	5302	SUMBA	31.03	Tinggi	18.32966288	Tinggi	21.86082406	Menengah	34.13895006	Tinggi	40.32162117	Tinggi

2. Tabel Data

Halaman ini menyajikan keseluruhan dataset yang telah digabungkan, termasuk informasi wilayah (kode dan nama kabupaten/kota) serta penamaan variabel yang telah disesuaikan agar lebih intuitif. Tabel ini sangat berguna sebagai referensi utama bagi pengguna sebelum melakukan proses analisis atau visualisasi data. Fitur-fitur yang tersedia dalam subpage ini antara lain:

- **Tabel data mentah (raw):** Menyajikan seluruh observasi dari dataset SOVI dalam format interaktif.
- **Tombol unduh (CSV):** Mengizinkan pengguna mengunduh dataset lengkap dalam format CSV untuk kebutuhan eksplorasi atau dokumentasi lebih lanjut.

Dashboard KUPAS Beranda Manajemen Data Eksplorasi Data Analisis Data

Menampilkan dataset SOVI yang telah digabung dengan nama wilayah dan nama kolom yang disesuaikan.

[Download Data \(CSV\)](#)

Show entries Search:

	Kode Wilayah	Nama Wilayah	Kemiskinan (%)	Rumah Tangga Tanpa Listrik (%)	Rumah Tangga Tanpa Drainase (%)	Rumah Tangga Guna Air Pipa (%)	Pendidikan Rendah (%)	Buta Huruf (%)
1	5305	TIMOR TENGAH UTARA	23.52	26.71624211	11.72406519	2.898932897	39.7729225	10.31522401
2	5308	LEMBATA	26.48	19.88028683	9.027102577	17.22775887	41.15813577	5.145441514
3	5307	ALOR	21.67	17.60029926	15.35584027	7.33657533	37.35738588	7.174933258
4	5302	SUMBA TIMUR	31.03	18.32966288	34.13895006	21.86082406	40.32162117	11.62298855
5	5301	SUMBA BARAT	29.28	42.41426726	50.90212092	0	28.61113514	19.12269363
6	5304	TIMOR TENGAH SELATAN	29.44	43.07146532	5.956797646	10.86738701	42.54090122	16.73075219
7	5303	KUPANG	22.91	22.67748025	8.52749764	3.20184808	36.99995054	11.05458361
8	3509	JEMBER	11	0.0732369	30.87049322	4.205278896	47.79775459	12.31755959

3.7.3 Menu Eksplorasi Data

Menu ini dirancang sebagai bagian dari *Business Intelligence* yang memungkinkan pengguna untuk mengeksplorasi data sosial ekonomi kabupaten/kota di Indonesia melalui pendekatan interaktif dan visual. Tujuannya adalah membantu dalam menemukan pola, hubungan, atau penyimpangan yang signifikan sebelum proses analisis lanjutan dilakukan. Eksplorasi dilakukan melalui tiga sub-halaman utama, yaitu:

1. Statistik Deskriptif

Halaman ini menyajikan ringkasan statistik dari variabel numerik yang tersedia dalam data. Statistik deskriptif menjadi fondasi awal untuk memahami distribusi, kecenderungan pusat, dan penyebaran data secara umum. Fitur-fitur yang tersedia dalam subpage ini antara lain:

- **Pemilihan Variabel**
Pengguna dapat memilih satu variabel numerik dari daftar variabel yang tersedia untuk dianalisis secara deskriptif.
- **Ringkasan Statistik Otomatis**
Setelah variabel dipilih, sistem akan menampilkan statistik seperti: nilai minimum, maksimum, median, kuartil, rata-rata, dan standar deviasi.
- **Interpretasi**
Sistem akan menghasilkan interpretasi otomatis dari ringkasan statistik, guna memudahkan pemahaman makna angka-angka tersebut dalam konteks sosial ekonomi.
- **Tombol Unduh**
Tersedia opsi untuk mengunduh hasil ringkasan dan interpretasi dalam format PDF untuk keperluan dokumentasi atau pelaporan.

The screenshot shows the 'Eksplorasi Data' section of the 'Dashboard KUPAS'. At the top, there's a navigation bar with 'Beranda', 'Manajemen Data', 'Eksplorasi Data', and 'Analisis Data'. Below the navigation bar, a dropdown menu labeled 'Pilih Variabel:' has 'Kemiskinan' selected. The main content area is divided into three sections: 'Ringkasan Statistik', 'Interpretasi', and a 'Unduh Ringkasan PDF' button at the bottom.

Ringkasan Statistik

Ringkasan statistik untuk 'Kemiskinan':

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.760	7.305	11.100	12.943	15.980	43.630

Interpretasi

Data ini merepresentasikan kondisi sosial di seluruh kabupaten/kota di Indonesia terkait variabel **Kemiskinan**. Nilai yang ditampilkan adalah persentase, sehingga angka mencerminkan proporsi rumah tangga atau penduduk yang terdampak oleh kondisi tersebut. Secara nasional, rata-rata nilai variabel ini adalah **12.94%**, yang menunjukkan bahwa secara umum, sekitar 12.94% dari populasi di setiap kabupaten/kota mengalami kondisi ini. Median (nilai tengah) tercatat sebesar **11.1%**, memberikan gambaran tipikal di tingkat kabupaten/kota. Nilai minimum ditemukan sebesar **1.76%**, dan maksimum mencapai **43.63%**.

Karena rata-rata lebih tinggi dari median, hal ini mengindikasikan adanya beberapa daerah dengan persentase yang sangat tinggi, sehingga menarik rata-rata ke atas. Distribusi seperti ini disebut *condong ke kanan* atau *positif skew*.

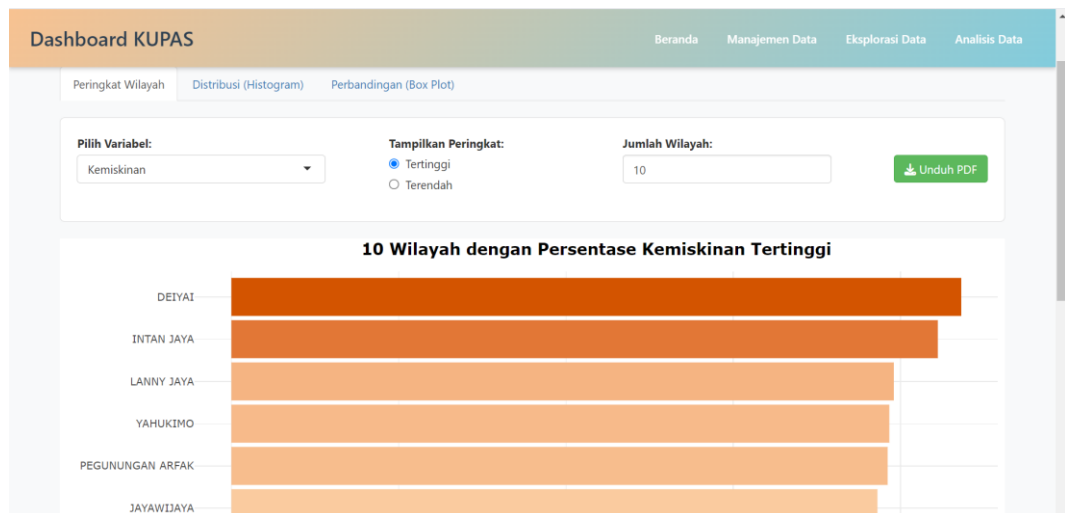
[Unduh Ringkasan PDF](#)

2. Visualisasi Data

Halaman ini menyediakan berbagai grafik eksploratif untuk memahami distribusi, perbandingan antar wilayah, serta hubungan antar kelompok. Visualisasi membantu pengguna mengidentifikasi outlier, pola distribusi, dan segmentasi wilayah. Fitur-fitur yang tersedia dalam subpage ini antara lain:

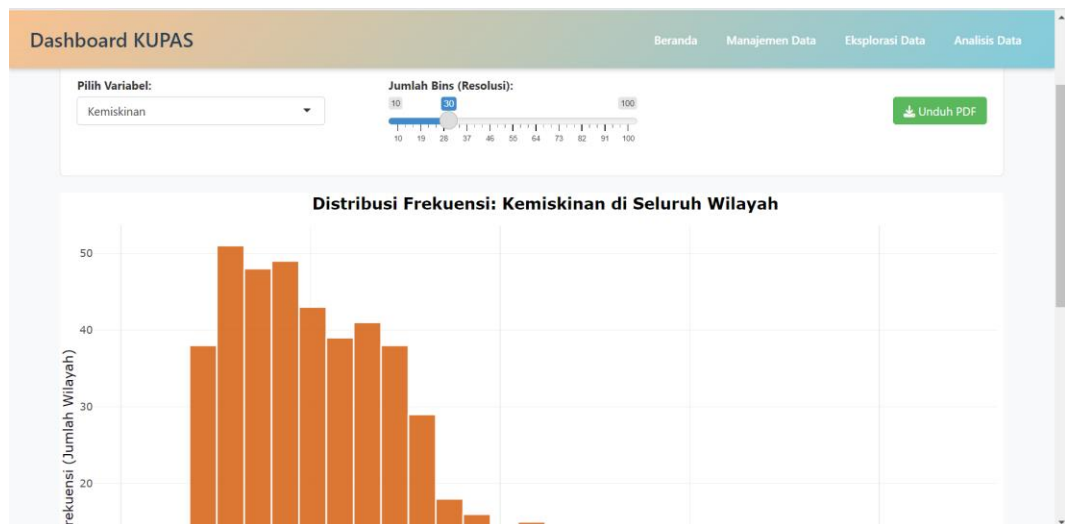
- Peringkat Wilayah (Bar Chart)

Berisi fitur seperti, Dropdown Variabel untuk memilih indikator sosial ekonomi. Radio Button Sortir (Tertinggi/Terendah) untuk mengatur arah peringkat. Input Jumlah Wilayah untuk menyesuaikan jumlah wilayah yang ditampilkan. Plot Interaktif untuk memvisualisasikan ranking antar wilayah. Unduh PDF hasil chart.



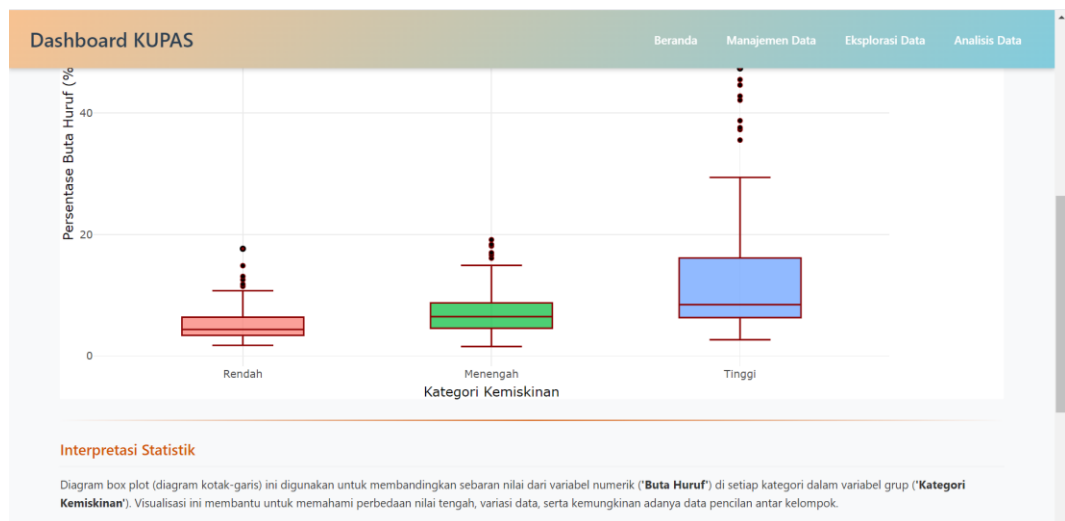
- Distribusi (Histogram)

Digunakan untuk mengevaluasi bentuk distribusi dari suatu variabel numerik secara keseluruhan. Fitur ini mempermudah dalam mendeteksi sebaran data seperti simetri, skewness, dan potensi pencilan. Berisi fitur-fitur seperti : Menu pilihan Variabel, untuk memilih variabel numerik yang akan dianalisis. Slider Jumlah Bin, untuk menyesuaikan jumlah interval (kelas) pada histogram. Plot Histogram Interaktif, memudahkan dalam mengamati pola distribusi secara visual. Interpretasi, disediakan untuk membantu memahami makna dari pola distribusi yang ditampilkan. Tombol Unduh PDF, untuk menyimpan hasil visualisasi dan interpretasi.



- **Perbandingan (Box Plot)**

Fitur ini digunakan untuk membandingkan distribusi suatu variabel numerik antar kelompok kategorik. Box plot berguna dalam mengevaluasi perbedaan tendensi pusat, sebaran, dan pencilan antar kategori. Fitur yang tersedia: Dropdown Variabel Numerik (Y), yaitu variabel utama yang ingin dibandingkan. Dropdown Variabel Kategorik (X), untuk menentukan kelompok pembanding (misalnya status wilayah, zona, dsb). Visualisasi Box Plot Interaktif, menampilkan median, kuartil, dan outlier untuk masing-masing kelompok. Interpretasi, dihasilkan untuk menjelaskan perbedaan antar kategori. Tombol Unduh PDF, untuk menyimpan hasil plot yang terbentuk.



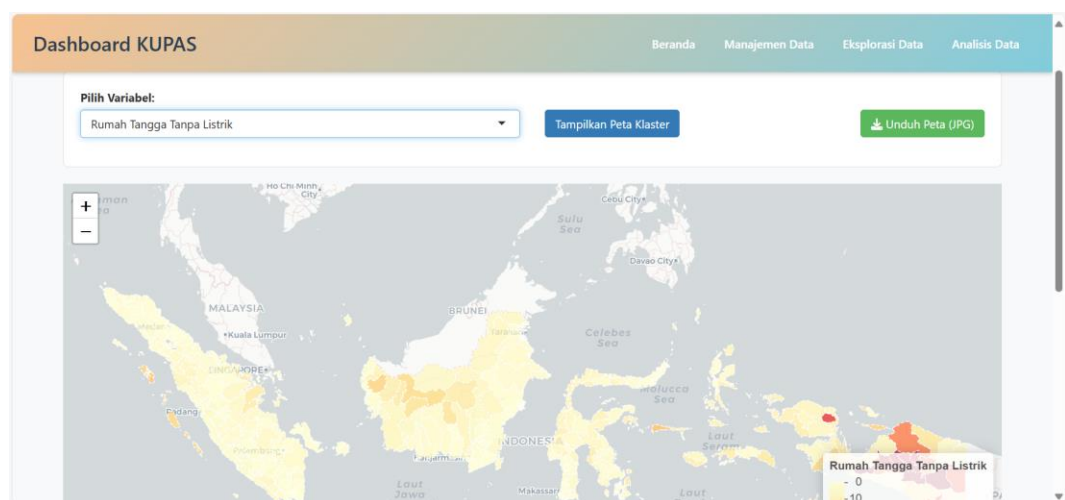
3. Pemetaan

Halaman ini menyajikan pemetaan spasial indikator sosial ekonomi untuk memberikan pemahaman yang lebih mendalam mengenai distribusi geografis antar wilayah. Dengan pendekatan visual berbasis peta, pengguna dapat dengan mudah mengidentifikasi pola spasial

seperti konsentrasi kemiskinan, kesenjangan infrastruktur, maupun distribusi pendidikan rendah secara regional. Fitur yang ada di halaman ini antara lain:

- Peta Persebaran Variabel

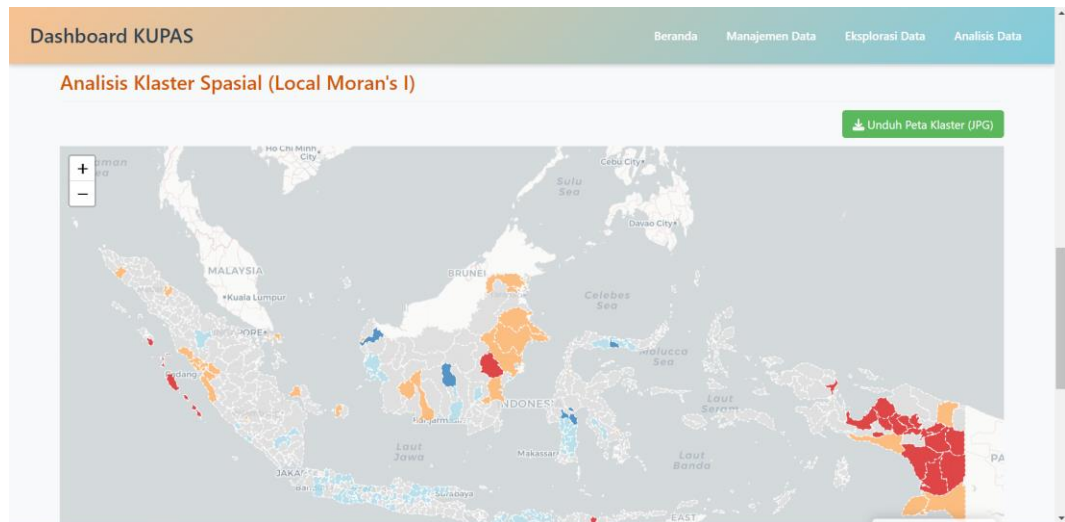
Subbagian ini memungkinkan pengguna untuk memilih satu variabel dari daftar indikator sosial ekonomi, kemudian memvisualisasikannya dalam bentuk peta choropleth berdasarkan kuartil. Wilayah-wilayah pada peta diberi warna yang merepresentasikan tingkat nilai dari variabel tersebut (misalnya rendah, menengah, tinggi), sehingga memudahkan analisis spasial secara deskriptif. Fitur-fitur yang tersedia meliputi: menu pilihan Variabel, memungkinkan pengguna memilih variabel yang ingin divisualisasikan dari daftar indikator. Peta Interaktif, pengguna dapat mengarahkan kursor ke suatu wilayah untuk melihat informasi detail seperti nama kabupaten/kota dan nilai variabel terkait. Tombol Unduh Peta (JPG), menyediakan opsi untuk menyimpan peta yang telah ditampilkan ke dalam format gambar (.jpg) untuk kebutuhan dokumentasi atau laporan. Panel Interpretasi Otomatis, menampilkan penjelasan teks yang secara dinamis menyesuaikan variabel yang dipilih, membantu memahami konteks dan makna dari pola distribusi yang muncul.



- Peta Klaster LISA

Selain peta persebaran, pengguna juga dapat menampilkan analisis klaster spasial berbasis Local Indicators of Spatial Association (LISA). Analisis ini memberikan insight lebih lanjut tentang wilayah yang secara signifikan membentuk klaster tinggi-tinggi, rendah-rendah, atau outlier (tinggi-rendah/rendah-tinggi). Fitur yang tersedia: Tombol “Tampilkan Peta Klaster”, saat diklik, sistem akan menjalankan analisis LISA untuk variabel yang sedang dipilih. Peta Klaster Interaktif, menunjukkan hasil identifikasi klaster dengan pewarnaan yang berbeda untuk tiap tipe klaster (misalnya High-High, Low-Low, High-Low, Low-High). Panel Interpretasi, penjelasan

deskriptif mengenai hasil klasterisasi yang muncul, untuk membantu pengguna memahami wilayah-wilayah yang memiliki hubungan spasial signifikan. Tombol Unduh Peta (JPG), menyediakan opsi untuk menyimpan peta yang telah ditampilkan ke dalam format gambar (.jpg).



3.7.4 Menu Analisis Data

Halaman Analisis Data dirancang untuk membantu pengguna dalam menguji hipotesis, memverifikasi asumsi dasar, dan melakukan analisis lanjutan terhadap data sosial ekonomi yang tersedia. Modul ini menyatukan berbagai metode statistik seperti uji asumsi klasik, uji beda rata-rata, analisis ragam (ANOVA), dan regresi linear, semuanya disajikan secara interaktif dan mudah diakses melalui antarmuka berbasis web. Secara garis besar, menu analisis data dibagi menjadi tiga bagian utama:

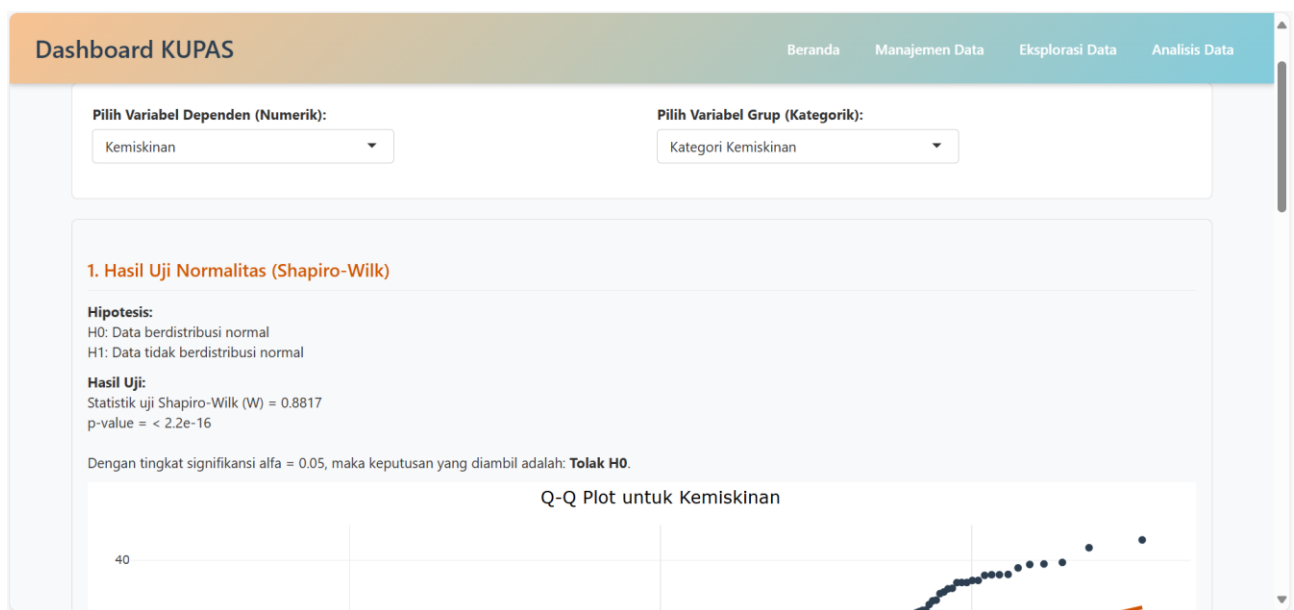
1. Uji Asumsi

Halaman ini dirancang untuk melakukan pengujian Normalitas Sebaran Data dan Homogenitas Varian antar kelompok secara interaktif. Fitur-fitur pada Subhalaman Uji Asumsi, yaitu:

- Menu pilihan Variabel Dependen (Numerik), pengguna memilih variabel numerik yang akan diuji distribusinya dan variansinya antar kelompok.
- Menu pilihan Variabel Grup (Kategorik), digunakan untuk mengelompokkan data saat menguji homogenitas varian (misalnya berdasarkan wilayah, jenis kelamin, status sosial, dsb.).
- Uji Normalitas (Shapiro-Wilk), sistem akan menampilkan hasil statistik uji Shapiro-Wilk secara otomatis, lengkap dengan QQ-plot interaktif sebagai alat bantu visual untuk memeriksa sebaran data. Komponen yang tersedia: Nilai statistik uji dan p-

value, Plot QQ interaktif untuk melihat pola sebaran data terhadap distribusi normal teoritis, Interpretasi otomatis berdasarkan hasil uji dan visualisasi.

- Uji Homogenitas Varian (Bartlett). Pengujian ini digunakan untuk mengetahui apakah varians antar kelompok sama atau tidak. Hasil disajikan dalam bentuk statistik uji dan p-value, serta didukung dengan boxplot interaktif untuk visualisasi persebaran data tiap grup. Komponen yang tersedia: Ringkasan hasil uji Bartlett, Boxplot interaktif untuk melihat persebaran dan outlier antar kelompok, Interpretasi otomatis berdasarkan hasil uji dan distribusi visual.
- Unduh Laporan PDF, pengguna dapat mengunduh hasil uji asumsi beserta grafik dan interpretasinya ke dalam format PDF melalui tombol "Unduh Laporan Uji Asumsi".



2. Statistik Inferensia

Halaman Analisis Inferensia menyediakan berbagai metode uji hipotesis statistik yang bertujuan untuk menggeneralisasi temuan dari sampel ke populasi. Fitur-fitur pada halaman ini memungkinkan pengguna untuk melakukan pengujian perbedaan rata-rata, proporsi, varian, dan analisis varians (ANOVA) secara interaktif. Setiap metode dilengkapi dengan panduan penggunaan, input variabel melalui dropdown, serta hasil analisis yang ditampilkan secara otomatis. Fitur-fitur tersebut dan penjelasannya adalah sebagai berikut:

- Uji Beda Rata-rata (Uji-t)
Sub-menu ini memungkinkan pengguna untuk memilih antara: Uji-t Satu Sampel: untuk membandingkan rata-rata suatu variabel numerik terhadap nilai acuan tertentu. Uji-t Dua Sampel Independen: untuk menguji perbedaan rata-rata antara dua

kelompok berdasarkan variabel kategorik (misalnya kategori 'Tinggi' dan 'Rendah'). Hasil pengujian ditampilkan dalam bentuk tabel statistik, nilai p, dan interpretasi ringkas mengenai signifikansi hasil.

Dashboard KUPAS

Pilih Variabel: Kemiskinan Nilai Hipotesis Rata-rata (μ_0): 12,94

Hasil Uji-t

Hipotesis:

- $H_0: \mu = 12,94$
- $H_1: \mu \neq 12,94$

Statistik Uji (t): 0.0083

p-value: 0.9934

Keputusan ($\alpha = 0.05$): **Gagal Tolak H_0**

Kesimpulan Statistik:
Tidak terdapat cukup bukti untuk menyatakan adanya perbedaan rata-rata yang signifikan.

Interpretasi:
Berdasarkan hasil uji t dengan tingkat signifikansi 0.05 (95% tingkat kepercayaan), **tidak terdapat cukup bukti** untuk menyatakan bahwa rata-rata populasi berbeda dari nilai

- Uji Proporsi, digunakan untuk menguji apakah proporsi kejadian tertentu (disebut sebagai ‘sukses’) secara signifikan berbeda dari nilai proporsi yang dihipotesiskan. Pengguna dapat memilih variabel numerik, menetapkan ambang batas untuk mendefinisikan sukses, serta memasukkan nilai hipotesis sebagai pembanding. Hasil analisis ditampilkan dalam bentuk estimasi proporsi, nilai statistik uji, dan simpulan inferensial.

Dashboard KUPAS

Hasil Uji Proporsi Satu Sampel

Data: Terdapat 205 wilayah 'sukses' dari total 511 wilayah.

Hipotesis:

- $H_0: p = 0.5$
- $H_1: p \neq 0.5$

Statistik Uji (X-squared): 19.5695

p-value: 9.701e-06

Keputusan ($\alpha = 0.05$): **Tolak H_0**

Kesimpulan:
Proporsi sampel berbeda secara signifikan dari nilai hipotesis.

Interpretasi:
Berdasarkan hasil uji proporsi dengan tingkat signifikansi 0.05 (95% tingkat kepercayaan), terdapat cukup bukti untuk menyatakan bahwa proporsi populasi **berbeda** dari nilai hipotesis (0.5).

[Unduh Laporan Uji \(PDF\)](#)

- Uji Ragam (Uji F untuk Dua Ragam), fitur ini digunakan untuk menguji kesamaan ragam (varians) antara dua kelompok data berdasarkan variabel kategorik. Uji ini penting untuk memenuhi asumsi homogenitas varian sebelum melakukan uji-t dua

sampel. Pengguna memilih satu variabel numerik dan satu variabel grup, kemudian sistem akan menampilkan hasil uji F termasuk nilai statistik dan p-value.

Dashboard KUPAS

Variable Numerik (Y): Variable Grup (X):

Hasil Uji F untuk Dua Ragam

Hipotesis:
 H_0 : Rasio ragam = 1 (varians kedua kelompok sama)
 H_1 : Rasio ragam \neq 1 (varians kedua kelompok berbeda)

Statistik Uji (F): 0.0322

p-value: $< 2.2e-16$

Keputusan ($\alpha = 0.05$): Tolak H_0

Kesimpulan :
 Terdapat perbedaan ragam (varians) yang signifikan secara statistik antar kelompok.

- ANOVA (Analysis of Variance), dalam fitur ini tersedia dua jenis uji ANOVA: ANOVA Satu Arah: Menguji perbedaan rata-rata antara tiga atau lebih kelompok berdasarkan satu variabel kategorik. ANOVA Dua Arah: Menguji pengaruh dua faktor kategorik terhadap satu variabel numerik, termasuk interaksi antara kedua faktor. Pengguna memilih variabel numerik sebagai variabel dependen, serta satu atau dua variabel kategorik sebagai faktor. Hasil pengujian ditampilkan dalam tabel ANOVA yang mencakup nilai F, derajat bebas, p-value, dan interpretasi singkat.

Dashboard KUPAS

ANOVA Satu Arah: Pilih satu variabel dependen (numerik) dan satu variabel grup (faktor) yang memiliki 3 kategori ('Rendah', 'Menengah', 'Tinggi').

ANOVA Dua Arah: Pilih satu variabel dependen (numerik) dan dua variabel grup (faktor) yang berbeda untuk melihat pengaruh masing-masing faktor dan interaksinya.

ANOVA Satu Arah

Pilih Variabel Dependen (Numerik): Pilih Variabel Grup (Kategorik):

Hipotesis Uji ANOVA

- H_0 : Rata-rata semua kelompok adalah sama
- H_1 : Setidaknya ada satu rata-rata kelompok yang berbeda

Tabel Ringkasan ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
KEHISKINAN_cat	2	24284	12142	738.5	$< 2e-16$ ***
Residuals	508	8352	16		

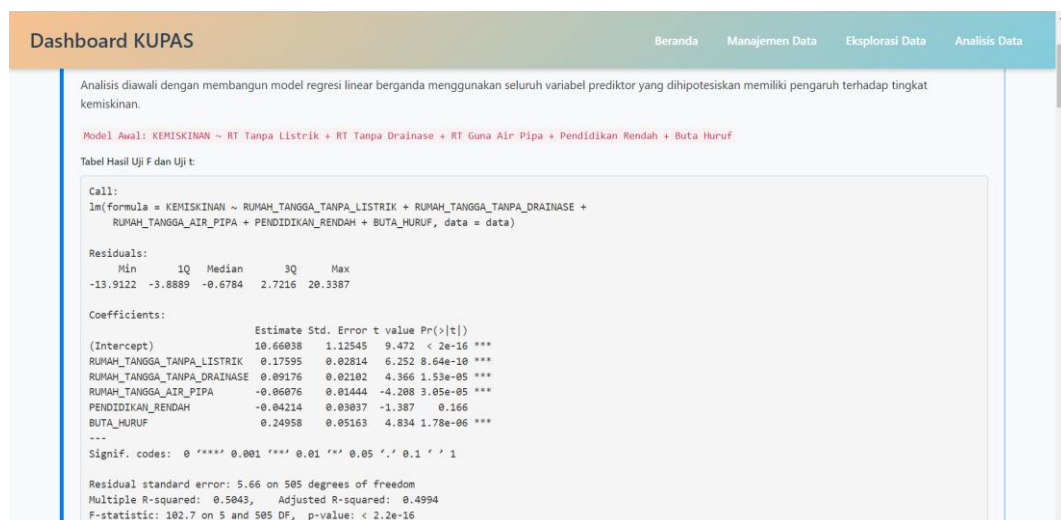
- Unduh Laporan, pengguna dapat mengunduh hasil analisis inferensia dalam format PDF dengan menekan tombol “Unduh Laporan Uji”. Laporan akan memuat hasil uji

yang telah dilakukan beserta interpretasinya, berguna untuk dokumentasi maupun pelaporan formal.

3. Analisis Regresi

Halaman Analisis Regresi ini menyajikan tahapan analisis regresi linear berganda secara sistematis, mulai dari pemodelan awal hingga penyempurnaan model dengan transformasi data. Tujuan utama dari analisis ini adalah untuk mengevaluasi pengaruh berbagai faktor sosial, khususnya akses terhadap infrastruktur dasar dan kondisi pendidikan, terhadap tingkat kemiskinan di wilayah kabupaten/kota di Indonesia.

Analisis diawali dengan membentuk model regresi linear berganda yang mencakup seluruh variabel prediktor yang dianggap berpengaruh terhadap tingkat kemiskinan. Variabel-variabel tersebut meliputi akses terhadap infrastruktur dasar seperti listrik, drainase, dan air bersih melalui pipa, serta kondisi pendidikan yang direpresentasikan oleh proporsi rumah tangga berpendidikan rendah dan tingkat buta huruf. Model awal ini digunakan untuk mengidentifikasi pengaruh simultan dari kelima variabel tersebut terhadap tingkat kemiskinan. Langkah ini bertujuan sebagai fondasi awal untuk mengevaluasi model secara statistik serta menjadi acuan untuk proses penyederhanaan model di tahap selanjutnya.



Berdasarkan hasil estimasi model awal, dilakukan evaluasi terhadap signifikansi masing-masing variabel. Pada tahap ini ditemukan bahwa variabel Pendidikan Rendah memiliki nilai p-value di atas 0,05, yang menandakan bahwa pengaruhnya terhadap tingkat kemiskinan tidak signifikan secara statistik pada taraf signifikansi 5%. Oleh karena itu, variabel tersebut dikeluarkan dari model guna menyederhanakan persamaan regresi dan meningkatkan interpretabilitas model. Perbaikan ini menghasilkan model baru dengan empat prediktor, yaitu: rumah tangga tanpa listrik, tanpa drainase, menggunakan air pipa, dan rumah tangga

dengan kepala keluarga buta huruf. Model yang lebih ringkas ini diuji kembali performanya dan dibandingkan dengan model sebelumnya, baik dari sisi R^2 maupun nilai signifikansi tiap koefisien.

Dashboard KUPAS

BerandaManajemen DataEksplorasi DataAnalisis Data

Berdasarkan hasil analisis Model Awal, variabel **Pendidikan Rendah (PENDIDIKAN_REDAH)** menunjukkan nilai p-value yang melebihi batas signifikansi ($p > 0.05$), sehingga tidak terbukti memiliki pengaruh yang signifikan terhadap tingkat kemiskinan. Jika dibandingkan dengan variabel lain dalam model yang terbukti signifikan, kontribusi variabel ini relatif rendah. Oleh karena itu, untuk memperoleh model yang lebih baik dan interpretatif, variabel tersebut dikeluarkan dari model pada tahap perbaikan.

Model Perbaikan: $KEMISKINAN \sim RT \text{ Tanpa Listrik} + RT \text{ Tanpa Drainase} + RT \text{ Guna Air Pipa} + Buta Huruf$

Hasil Uji F dan Uji t pada Model Perbaikan

```
Call:
lm(formula = KEMISKINAN ~ RUMAH_TANGGA_TANPA_LISTRIK + RUMAH_TANGGA_TANPA_DRAINASE + RUMAH_TANGGA_AIR_PIPA + BUTA_HURUF, data = data)
```

Residuals:					
Min	1Q	Median	3Q	Max	
-14.0641	-3.8670	-0.7695	2.7707	20.6163	

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.33998	0.69140	15.530	<.2e-16	***
RUMAH_TANGGA_TANPA_LISTRIK	0.18305	0.02770	6.608	9.87e-11	***
RUMAH_TANGGA_TANPA_DRAINASE	0.08477	0.02042	4.151	3.89e-05	***
RUMAH_TANGGA_AIR_PIPA	-0.05505	0.01385	-3.974	8.18e-05	***
BUTA_HURUF	0.24069	0.05128	4.694	3.45e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sebelum model perbaikan dapat digunakan sebagai dasar interpretasi dan pengambilan keputusan, perlu dilakukan pengujian asumsi-asumsi klasik regresi linear. Pengujian mencakup: Normalitas residual menggunakan uji Shapiro-Wilk, Homoskedastisitas menggunakan uji Breusch-Pagan. Hasil uji menunjukkan bahwa residual dari model belum memenuhi asumsi normalitas ($p\text{-value} < 0,05$). Ini mengindikasikan bahwa distribusi galat tidak normal, yang dapat mempengaruhi validitas inferensi statistik dari model. Selain itu, terdapat indikasi ringan heteroskedastisitas. Maka, dilakukan langkah perbaikan lanjutan berupa transformasi logaritmik.

Dashboard KUPAS

BerandaManajemen DataEksplorasi DataAnalisis Data

Pengujian asumsi klasik dilakukan terhadap Model Perbaikan guna mengevaluasi kesesuaian model dengan prasyarat analisis regresi linear.

Uji Multikolinearitas (VIF)

RUMAH_TANGGA_TANPA_LISTRIK	RUMAH_TANGGA_TANPA_DRAINASE	RUMAH_TANGGA_AIR_PIPA	BUTA_HURUF
2.755892	1.530395	1.151890	2.661194

→ Jika semua nilai VIF < 10, maka tidak ada indikasi multikolinearitas yang serius.

Uji Normalitas (Shapiro-Wilk)

```
Shapiro-Wilk normality test
data: residuals(model)
W = 0.94919, p-value = 3.047e-12
```

→ H_0 : error berdistribusi normal. H_1 : error tidak berdistribusi normal.
p-value = 0. Karena p-value < 0.05, tolak H_0 → error tidak berdistribusi normal.

Uji Homoskedastisitas (Breusch-Pagan)

```
studentized Breusch-Pagan test
data: model
BP = 25.272, df = 4, p-value = 4.436e-05
```

Untuk mengatasi pelanggaran asumsi klasik, dilakukan transformasi logaritmik terhadap variabel dependen (KEMISKINAN), menghasilkan model regresi log-lin (log-linear). Model ini tidak hanya membantu dalam memperbaiki distribusi residual dan kestabilan variansi, tetapi juga memberikan interpretasi yang lebih relevan secara ekonometrik. Koefisien pada

model log-lin dapat diinterpretasikan sebagai persentase perubahan pada tingkat kemiskinan akibat perubahan satu unit pada variabel prediktor. Hasil uji asumsi klasik pada model final menunjukkan perbaikan signifikan, di mana distribusi residual menjadi lebih mendekati normal, serta variansi residual cenderung homogen.

Dashboard KUPAS

BerandaManajemen DataEksplorasi DataAnalisis Data

Transformasi logaritma natural diterapkan pada variabel dependen KEMISKINAN untuk mengatasi pelanggaran asumsi klasik, khususnya ketidakterpenuhan normalitas residual dan indikasi heteroskedastisitas. Digunakan model log-lin (log-linear) menghasilkan distribusi residual menjadi lebih mendekati normal dan variansi model menjadi lebih stabil. Selain itu, model log-lin memungkinkan interpretasi koefisien sebagai perubahan persentase pada variabel dependen akibat perubahan satu unit pada variabel prediktor, yang relevan dalam analisis sosial-ekonomi.

Model Final: $\log(\text{KEMISKINAN}) \sim \text{RT Tanpa Listrik} + \text{RT Tanpa Drainase} + \text{RT Guna Air Pipa} + \text{Buta Huruf}$

4.1 Hasil Model Final

Tabel Koefisien Model Log-Linear:

Interpretasi koefisien (β) pada model ini adalah dalam bentuk semi-elastisitas: jika variabel X naik 1 unit, maka Y akan berubah sebesar ($\beta \times 100$)%.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.168314992	0.051201277	42.348846	1.835686e-168
RUMAH_TANGGA_TANPA_LISTRIK	0.006059077	0.002358285	2.569273	1.047660e-02
RUMAH_TANGGA_TANPA_DRAINASE	0.008983330	0.001738620	5.120917	4.329219e-07
RUMAH_TANGGA_AIR_PIPA	-0.006138148	0.001179400	-5.204467	2.832860e-07
BUTA_HURUF	0.016306004	0.004365450	3.735240	2.088687e-04

Interpretasi Koefisien:

$b_0 = 2.1683$, artinya saat seluruh variabel prediktor bernilai nol, maka logaritma tingkat kemiskinan diperkirakan sebesar 2.1683. Ini adalah baseline model.
 $b_1 = 0.0061$, artinya setiap kenaikan 1% pada proporsi Rumah Tangga Tanpa Listrik akan meningkatkan tingkat kemiskinan sekitar 0.61%, dengan asumsi variabel lain tetap konstan.
 $b_2 = 0.0089$, artinya setiap kenaikan 1% pada proporsi Rumah Tangga Tanpa Drainase akan meningkatkan tingkat kemiskinan sekitar 0.89%, dengan asumsi variabel lain tetap konstan.
 $b_3 = -0.0061$, artinya setiap kenaikan 1% pada proporsi Rumah Tangga Guna Air Pipa akan menurunkan tingkat kemiskinan sekitar 0.61%, dengan asumsi variabel lain tetap konstan.
 $b_4 = 0.0163$, artinya setiap kenaikan 1% pada proporsi Buta Huruf akan meningkatkan tingkat kemiskinan sekitar 1.63%, dengan asumsi variabel lain tetap konstan.

Dalam halaman ini juga diberikan kesimpulan terkait model final yang dihasilkan telah memenuhi kriteria statistik serta asumsi klasik yang diperlukan untuk regresi linear berganda, serta fitur unduh laporan sesuai analisis regresi yang ditampilkan pada halaman ini.

Dashboard KUPAS

BerandaManajemen DataEksplorasi DataAnalisis Data

Berdasarkan hasil analisis regresi yang dilakukan, model log-linear (dengan variabel logaritma Y) tanpa variabel pendidikan rendah dipilih sebagai model terbaik. Alasan pemilihan ini didasarkan pada kriteria statistik dan diagnostik model berikut:

- Seluruh variabel prediktor dalam model final signifikan secara statistik berdasarkan uji t ($p\text{-value} < 0.05$), menunjukkan bahwa variabel-variabel tersebut memiliki kontribusi yang nyata terhadap tingkat kemiskinan.
- Uji F menunjukkan bahwa model signifikan secara keseluruhan, artinya model secara umum menjelaskan variasi dalam data dengan baik.
- Tidak terdapat indikasi multikolinearitas yang serius karena seluruh nilai Variance Inflation Factor (VIF) berada di bawah ambang batas 10.
- Transformasi logaritma pada variabel dependen berhasil memperbaiki indikasi heteroskedastisitas, seperti ditunjukkan oleh $p\text{-value}$ dari uji Breusch-Pagan yang meningkat.
- Meskipun asumsi normalitas belum sepenuhnya terpenuhi ($p\text{-value}$ Shapiro-Wilk = 0.0245), ukuran sampel yang besar memungkinkan penerapan **Teorema Limit Pusat**. Oleh karena itu, inferensi berdasarkan model ini tetap dapat dipertanggungjawabkan secara statistik.

Kesimpulan

Dashboard ini merupakan alat analisis komprehensif yang dirancang untuk mengevaluasi pengaruh akses terhadap infrastruktur dasar (listrik, air bersih melalui pipa, dan sistem drainase) serta kondisi pendidikan (tingkat pendidikan rendah dan buta huruf) terhadap tingkat kemiskinan di wilayah kabupaten/kota di Indonesia.

Melalui pendekatan regresi linear berganda yang disempurnakan dengan transformasi logaritma (log-linear), analisis ini menemukan bahwa faktor-faktor seperti keterbatasan akses listrik, drainase, dan air bersih, serta tingginya angka buta huruf, memiliki hubungan yang signifikan terhadap tingkat kemiskinan.

Model log-linear yang digunakan tidak hanya memenuhi sebagian besar asumsi klasik regresi, tetapi juga menunjukkan kapasitas yang kuat dalam menjelaskan variasi kemiskinan antar wilayah. Temuan ini menegaskan pentingnya intervensi kebijakan yang berfokus pada peningkatan akses infrastruktur dan perbaikan kualitas pendidikan sebagai strategi pengurangan kemiskinan di wilayah Kab/Kota di Indonesia.

karena itu, interensi berdasarkan model ini tetap dapat dipertanggungjawabkan secara statistik.

Kesimpulan

Dashboard ini merupakan alat analisis komprehensif yang dirancang untuk mengevaluasi pengaruh akses terhadap infrastruktur dasar (listrik, air bersih melalui pipa, dan sistem drainase) serta kondisi pendidikan (tingkat pendidikan rendah dan buta huruf) terhadap tingkat kemiskinan di wilayah kabupaten/kota di Indonesia.

Melalui pendekatan regresi linear berganda yang disempurnakan dengan transformasi logaritma (log-linear), analisis ini menemukan bahwa faktor-faktor seperti keterbatasan akses listrik, drainase, dan air bersih, serta tingginya angka buta huruf, memiliki hubungan yang signifikan terhadap tingkat kemiskinan.

Model log-linear yang digunakan tidak hanya memenuhi sebagian besar asumsi klasik regresi, tetapi juga menunjukkan kapasitas yang kuat dalam menjelaskan variasi kemiskinan antar wilayah. Temuan ini menegaskan pentingnya intervensi kebijakan yang berfokus pada peningkatan akses infrastruktur dan perbaikan kualitas pendidikan sebagai strategi pengurangan kemiskinan di wilayah Kab/Kota di Indonesia.

[Unduh Laporan Analisis Regresi](#)

BAB 4

PENUTUP

4.1 Kesimpulan

Dashboard interaktif ini dikembangkan sebagai alat bantu visual dan analitis dalam memahami berbagai faktor sosial ekonomi yang memengaruhi tingkat kemiskinan di 511 kabupaten/kota di Indonesia berdasarkan data tahun 2017. Dengan memanfaatkan teknologi *Shiny* dalam bahasa pemrograman R, dashboard ini dirancang untuk menyajikan informasi statistik secara komprehensif, interaktif, dan mudah diakses, baik oleh peneliti, pengambil kebijakan, maupun publik umum. Struktur dashboard terbagi ke dalam beberapa menu utama, yaitu *Beranda*, *Eksplorasi Data*, *Eksplorasi Spasial*, *Analisis Data*, dan *Unduh Laporan*. Masing-masing menu memiliki subhalaman yang menyajikan fitur-fitur berbeda, mulai dari ringkasan statistik deskriptif, visualisasi peta tematik, uji asumsi klasik, hingga analisis regresi linier berganda.

Berdasarkan analisis regresi yang telah dilakukan, ditemukan bahwa beberapa variabel sosial ekonomi memiliki pengaruh signifikan terhadap tingkat kemiskinan. Variabel-variabel tersebut antara lain: persentase rumah tangga tanpa akses listrik, rumah tangga tanpa drainase, rumah tangga tanpa air bersih dari pipa, serta tingkat pendidikan kepala keluarga. Keempat variabel ini menunjukkan hubungan yang signifikan dengan tingkat kemiskinan, baik secara statistik maupun secara interpretatif berdasarkan konteks sosial. Misalnya, akses terhadap infrastruktur dasar seperti listrik dan air bersih dapat secara langsung memengaruhi kualitas hidup dan produktivitas masyarakat, sementara pendidikan kepala keluarga menjadi indikator penting dalam mobilitas ekonomi dan akses terhadap pekerjaan yang layak.

Sementara itu, meskipun pada analisis awal variabel *Low Education* (persentase kepala keluarga berpendidikan rendah) menunjukkan signifikansi secara statistik ketika diuji secara terpisah terhadap tingkat kemiskinan, variabel ini tidak dimasukkan dalam model akhir. Hal ini disebabkan karena kontribusinya menjadi tidak signifikan ketika dikombinasikan dengan variabel lain yang lebih dominan. Kemungkinan terdapat efek tumpang tindih atau kolinearitas dengan variabel pendidikan formal dan faktor struktural lainnya. Oleh karena itu, meskipun *Low Education* tetap relevan secara substantif, model akhir lebih mengedepankan variabel-variabel dengan kekuatan prediktif yang lebih kuat dan tidak saling mengaburkan.

Secara keseluruhan, dashboard ini berhasil menunjukkan bahwa pendekatan analisis data eksploratif dan inferensial dapat digunakan untuk mengidentifikasi faktor-faktor kunci dalam perumusan kebijakan pengentasan kemiskinan. Visualisasi spasial dalam bentuk peta kluster juga memberikan wawasan tambahan mengenai persebaran wilayah-wilayah yang memerlukan intervensi kebijakan

prioritas. Dengan demikian, pengembangan sistem informasi berbasis data ini diharapkan dapat menjadi contoh penerapan *data-driven decision making* dalam konteks pembangunan sosial di tingkat daerah.

4.2 Saran

Berdasarkan hasil analisis dan pengembangan dashboard ini, terdapat beberapa saran yang dapat disampaikan untuk pengembangan penelitian selanjutnya maupun pemanfaatan hasil analisis dalam praktik kebijakan:

1. Pentingnya Variabel Kontekstual Tambahan

Meskipun model regresi telah menangkap beberapa variabel signifikan yang berpengaruh terhadap kemiskinan, masih terdapat kemungkinan adanya faktor penting lain yang belum tercakup, seperti tingkat ketimpangan pendapatan, akses layanan kesehatan, atau keterlibatan dalam sektor informal. Penelitian lanjutan disarankan untuk menggali variabel-variabel tersebut agar model menjadi lebih komprehensif dan representatif terhadap kondisi sosial ekonomi daerah.

2. Analisis Lanjutan dengan Pendekatan Multilevel atau Spasial

Mengingat data yang digunakan memiliki struktur wilayah yang beragam dan tersebar secara geografis, pendekatan lanjutan seperti *Multilevel Regression* atau *Spatial Econometrics* dapat digunakan untuk menangkap efek spasial dan heterogenitas antar wilayah.

3. Peningkatan Kualitas dan Cakupan Data

Data tahun 2017 memberikan gambaran sebaran data yang kuat, namun untuk keperluan pemantauan kebijakan jangka panjang, perlu dilakukan integrasi data lintas waktu (*time series*) agar dinamika kemiskinan dan faktor penyebabnya bisa dianalisis secara lebih menyeluruh. Selain itu, kualitas data input dari pemerintah daerah juga perlu ditingkatkan agar tidak terjadi bias atau missing values yang dapat mengganggu hasil analisis.

DAFTAR PUSTAKA

Dataset: https://raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/sovi_data.csv

Distance matrix: <https://raw.githubusercontent.com/bmlmcmc/naspaclust/main/data/distance.csv>

Kurniawan, R., Wijayanto, A. W., & Yuniarto, B. (2021). *Modul Praktikum Komputasi Statistik*. Politeknik Statistika STIS.

Kurniawan, R., Nasution, B. I., Agustina, N., et al. (2022). Revisiting social vulnerability analysis in Indonesia: An optimized spatial fuzzy clustering approach. *Data in Brief*, 40, 107743. <https://doi.org/10.1016/j.dib.2021.107743>