

**LAPORAN PROYEK DASHBOARD RSHINY  
UAS MATA KULIAH KOMPUTASI STATISTIK**



**Dosen Pengampu:  
Yuliagnis Transver Wijaya**

Tanggal: 23 Juli 2025

**Disusun Oleh:**  
Christiano Teddy Anta (222313031)  
Kelas: 2KS2

**PROGRAM STUDI D-IV KOMPUTASI STATISTIK  
POLITEKNIK STATISTIKA STIS  
JAKARTA TIMUR  
2024/2025**

**PAKTA INTEGRITAS UJIAN AKHIR SEMESTER**

**Mata Kuliah: Komputasi Statistik**

Prodi: DIV Komputasi Statistik

Saya yang bertanda tangan di bawah ini:

NIM : 222313031

Nama : Christiano Teddy Anta

Kelas : 2KS2

Dengan ini menyatakan bahwa dashboard dan laporan dashboard Komstat yang saya buat dan kumpulkan untuk Ujian Akhir Semester Mata Kuliah Komputasi Statistik adalah hasil karya saya sendiri. Saya menyatakan tidak melakukan tindakan kecurangan, plagiat, atau pelanggaran akademik lainnya.

Apabila di kemudian hari terbukti melakukan pelanggaran terhadap Pakta Integritas ini, saya bersedia menerima sanksi akademik sesuai ketentuan yang berlaku di Politeknik Statistika STIS.

Demikian Pakta Integritas ini saya buat dengan sebenar-benarnya dan tanpa paksaan dari pihak manapun.

Jakarta, 23 Juli 2025

ttd



(Christiano Teddy Anta)

NIM. 222313031

**Link Dashboard:** <https://christiano22.shinyapps.io/STATeddy/>

## 1. Identifikasi Kebutuhan Data

Tahap ini adalah fondasi dari setiap proyek analisis data, di mana tujuan bisnis dan kebutuhan data didefinisikan dengan jelas. Dashboard STATeddy secara implisit telah melalui tahap ini, yang tercermin dari desain dan fungsionalitasnya.

- Masalah yang Akan Dicari Solusinya:

Masalah utamanya adalah bagaimana cara menganalisis data Social Vulnerability Index (SOVI) di Indonesia secara komprehensif. Dashboard ini bertujuan untuk menyediakan platform terpadu untuk menjelajahi, menganalisis, dan memvisualisasikan data kerentanan sosial yang kompleks, sehingga dapat memberikan *insight* yang berguna untuk pengambilan kebijakan, khususnya dalam penanggulangan bencana.

- Tujuan Proyek:

Tujuan dari dashboard "STATeddy" adalah untuk menjadi alat analisis statistik yang interaktif dan lengkap, mulai dari manajemen data dasar hingga analisis inferensial dan pemodelan yang lebih canggih seperti regresi dan *clustering*.

- Data yang Dibutuhkan:

- sovi\_data.csv:

Dataset utama yang berisi 17 variabel indikator kerentanan sosial untuk 514 kabupaten/kota di Indonesia.

- distance.csv:

Matriks jarak berukuran 512x512 yang merepresentasikan dissimilaritas atau kemiripan karakteristik antar wilayah. Data ini krusial untuk analisis *clustering*.

- Data Koordinat Geografis:

Dashboard ini secara cerdas mengambil data koordinat dari repositori eksternal di GitHub (indonesia-locations-data) untuk memetakan DISTRICTCODE ke lokasi geografis riil di Indonesia, yang sangat penting untuk analisis spasial yang akurat.

- Data Kustom:

Dashboard menyediakan fungsionalitas bagi pengguna untuk mengunggah dataset mereka sendiri dalam format CSV, Excel, atau SPSS.

- Sumber Pengumpulan Data:

Sumber data utama bersifat sekunder dan terdokumentasi, yaitu dari repositori GitHub bmlmcmc/naspaclust dan coll-j/indonesia-locations-data. Pengguna juga bisa menjadi sumber data primer dengan mengunggah file mereka sendiri.

- Pemahaman Jenis Data:

Dashboard ini dirancang untuk menangani berbagai jenis data seperti yang dijelaskan dalam paparan:

- Numerik (Kontinyu/Rasio): Seperti POPULATION, POVERTY, FAMILYSIZE.
- Kategorik (Nominal/Ordinal): Seperti DISTRICTCODE, dan variabel baru yang dihasilkan dari transformasi (misalnya, kategori pendapatan 'Rendah', 'Sedang', 'Tinggi').
- Pemahaman Proses Pengolahan: Kebutuhan data telah disesuaikan dengan seluruh alur analisis yang tersedia di dashboard, mulai dari pembersihan di menu "Manajemen Data" hingga visualisasi, uji statistik, dan pemodelan di menu-menu berikutnya.

## 2. Mengambil Data

Dashboard "STATeddy" menggunakan metode pengambilan data yang modern dan otomatis, sesuai dengan praktik analisis data saat ini.

- Metode dan Tools:
  - Dokumentasi (Otomatis):

Metode utama yang digunakan adalah pengambilan data dari sumber online (dokumentasi digital). Dashboard secara otomatis mengunduh file CSV dari URL GitHub menggunakan fungsi `read.csv()` di R saat aplikasi pertama kali dijalankan. Ini memastikan data yang digunakan selalu versi terbaru dari sumbernya.
  - Upload File Pengguna:

Dashboard ini juga memfasilitasi pengguna untuk "mengambil" data dari file lokal mereka. Ini sesuai dengan konsep dokumentasi di mana pengguna menyediakan data yang sudah ada. Tools yang digunakan adalah `fileInput` dari R Shiny yang mendukung format `.csv`, `.xlsx`, dan `.sav` (SPSS).

## 3. Mengintegrasikan Data

Mengintegrasikan data adalah proses fundamental untuk memastikan bahwa data yang akan dianalisis bersifat utuh, akurat, dan dapat dipercaya. Dashboard STATeddy tidak hanya memuat data, tetapi juga secara aktif melakukan beberapa langkah integrasi untuk meningkatkan kualitas dan keandalan data.

- Akurasi Data

Akurasi data adalah jaminan bahwa data merepresentasikan fakta atau kejadian di dunia nyata dengan benar. Dashboard STATeddy secara proaktif meningkatkan akurasi data, terutama pada aspek spasial (geografis):

  - Integrasi Data Koordinat:

Dataset `sovi_data.csv` yang menjadi data utama tidak memiliki informasi `latitude` dan `longitude`. Menyadari kekurangan ini, dashboard dirancang untuk

secara otomatis mengintegrasikan data dari sumber eksternal, yaitu repositori GitHub indonesia-locations-data.

- Fungsi `generate_real_indonesia_coordinates`:

Fungsi ini secara spesifik dibuat untuk mengambil kode unik wilayah (DISTRICTCODE) dari data SOVI, lalu mencocokkannya dengan database koordinat untuk mendapatkan latitude dan longitude yang riil. Proses ini secara signifikan meningkatkan akurasi lokasi geografis yang ditampilkan pada menu "Analisis Peta SOVI" dan "Clustering". Tanpa langkah integrasi ini, analisis spasial tidak akan valid.

- Logika Fallback:

Jika DISTRICTCODE tidak ditemukan di database koordinat, dashboard tidak berhenti, melainkan memiliki logika *fallback* untuk menghasilkan koordinat acak di dalam batas provinsi yang benar. Ini memastikan bahwa setiap titik data tetap dapat divisualisasikan secara spasial, meskipun dengan tingkat akurasi yang lebih rendah untuk kasus tertentu.

- Kelengkapan Data

Kelengkapan data mengacu pada sejauh mana kumpulan data memiliki semua nilai yang diperlukan. Dashboard STATeddy membantu pengguna dalam menilai kelengkapan data sejak awal.

- Identifikasi Awal:

Pada menu "Manajemen Data", terdapat kotak "Data Summary" yang secara eksplisit menampilkan jumlah total *missing values* yang ada di dalam dataset. Informasi ini krusial karena data yang tidak lengkap dapat menghasilkan analisis yang bias atau bahkan gagal.

- Penanganan Kegagalan Akses:

Dashboard ini dirancang dengan ketahanan. Jika koneksi ke URL data utama di GitHub gagal, aplikasi tidak akan macet. Sebaliknya, fungsi `load_data` akan menangkap *error* ini dan secara otomatis membuat dataset sampel. Ini menjamin "kelengkapan" dari sisi fungsionalitas, di mana pengguna tetap dapat mencoba dan memahami fitur-fitur dashboard meskipun data primer tidak tersedia.

- Keamanan Data

Aspek keamanan bertujuan melindungi data dari akses atau perubahan yang tidak sah. Dalam konteks aplikasi web seperti STATeddy:

- Akses Terkontrol:

Data diambil dari URL GitHub yang spesifik dan *read-only*. Pengguna dashboard tidak dapat mengubah data asli di server sumber.

- Sesi Lokal:

Saat pengguna mengunggah data kustom mereka, file tersebut diproses dalam sesi R yang berjalan di server. Data ini tidak disimpan secara permanen di server setelah sesi berakhir, sehingga melindungi kerahasiaan data pengguna.

- Tata Kelola Data

Tata kelola data memastikan bahwa data dikelola dengan baik dan memenuhi kebutuhan analisis.

STATeddy menerapkan prinsip ini melalui fitur dokumentasi yang superior.

- Menu Metadata:

Dashboard ini memiliki menu "Metadata" yang sangat komprehensif. Menu ini berfungsi sebagai kamus data terpusat yang menjelaskan setiap detail dari dataset SOVI dan Matriks Jarak. Informasi mencakup deskripsi untuk 17 variabel SOVI, unitnya, sumber datanya (BPS, Susenas), hingga perannya dalam komponen PCA. Adanya dokumentasi yang jelas seperti ini adalah pilar utama tata kelola data yang baik, karena memastikan semua pengguna memiliki pemahaman yang sama tentang data yang mereka analisis.

- Validitas Data

Validitas memastikan bahwa data yang ada di dalam sistem sudah benar dan sesuai format. Dashboard

STATeddy melakukan validasi secara implisit maupun eksplisit.

- Pratinjau Data:

Fitur pratinjau tabel interaktif di menu "Manajemen Data" memungkinkan pengguna untuk secara visual memvalidasi data mereka setelah dimuat, memastikan tidak ada kesalahan pemisahan kolom atau format yang salah.

- Validasi Fungsional:

Dashboard secara cerdas memvalidasi input pengguna. Misalnya, menu "Regresi Linear" hanya akan menampilkan variabel numerik sebagai pilihan untuk variabel dependen dan independen. Ini mencegah pengguna melakukan analisis yang tidak valid secara statistik (misalnya, meregresikan variabel teks), sehingga menjaga validitas seluruh proses analisis.

#### 4. Menelaah data

Menelaah data adalah langkah eksplorasi untuk memahami karakteristik, pola, dan potensi masalah dalam dataset. Dashboard STATeddy menyediakan serangkaian alat interaktif yang kuat untuk mendukung tahap ini secara menyeluruh.

- Identifikasi Tipe Data & Skala Pengukuran

Dashboard secara efektif mengidentifikasi dan memanfaatkan berbagai tipe data dan skala pengukuran (nominal, rasio) untuk analisis yang tepat.

- Identifikasi Otomatis: Saat data dimuat, lingkungan R secara otomatis mengklasifikasikan setiap kolom sebagai numerik, karakter, atau faktor.

- Penerapan Cerdas: Dashboard menggunakan klasifikasi ini untuk mengontrol antarmuka pengguna.
  - Rasio/Numerik:
 

Variabel seperti POPULATION, POVERTY, atau FEMALE diidentifikasi sebagai numerik. Oleh karena itu, variabel-variabel ini tersedia untuk perhitungan statistik (mean, median, standar deviasi) di menu "Statistik Deskriptif", sebagai sumbu X/Y di *scatter plot*, dan sebagai variabel dependen/independen dalam "Regresi Linear".
  - Nominal/Kategorik:
 

Variabel seperti DISTRICTCODE atau variabel baru yang dibuat dari transformasi (misalnya, kategori kuantil POVERTY\_cat) diidentifikasi sebagai data kategorik. Variabel ini secara otomatis tersedia sebagai opsi "Group By" di menu "Statistik Deskriptif" atau sebagai "Faktor" dalam analisis ANOVA, yang merupakan penggunaan yang tepat untuk data berskala nominal.
- Pemeriksaan Missing Value(s), Konsistensi Data, dan Outlier(s)
 

Dashboard menyediakan fitur-fitur visual dan ringkasan untuk memeriksa masalah umum pada data.

  - Missing Values:
 

Seperti yang disebutkan sebelumnya, jumlah data yang hilang (missing values) langsung ditampilkan di ringkasan data, menjadikannya langkah pertama dalam menelaah kualitas data.
  - Konsistensi Data:
 

Menu "Visualisasi" sangat berguna untuk memeriksa konsistensi. Pengguna dapat membuat *scatter plot* antara dua variabel yang seharusnya berkorelasi (misalnya, CHILDREN dan ELDERLY dengan POPULATION). Titik data yang menyimpang jauh dari pola umum dapat mengindikasikan adanya masalah konsistensi atau kesalahan input.
  - Outliers: STATeddy menawarkan beberapa cara untuk mendeteksi *outlier*.
    1. Box Plot: Alat utama untuk identifikasi *outlier* ada di menu "Visualisasi". *Box plot* secara grafis menampilkan titik-titik data yang berada di luar "pagar" (1.5 kali *Interquartile Range*), yang merupakan definisi umum dari *outlier*.
    2. Statistik Deskriptif: Perbedaan yang besar antara nilai *mean* dan *median*, serta nilai *min* atau *max* yang sangat ekstrem, bisa menjadi indikasi awal adanya *outlier*.
    3. Analisis Regresi: Untuk pengguna tingkat lanjut, plot diagnostik di menu "Regresi Linear", seperti plot Cook's Distance, dirancang khusus untuk mengidentifikasi *outlier* yang memiliki pengaruh besar (influential) terhadap model regresi.

## 5. Memvalidasi Data

Proses validasi data adalah tahap kritis untuk memastikan keandalan hasil analisis. Dashboard STATeddy mengintegrasikan beberapa mekanisme yang memungkinkan pengguna untuk melakukan validasi data secara efektif dan efisien, sesuai dengan lima pilar validasi data yang Anda sebutkan.

- Ketepatan Data (Accuracy)

Ketepatan data memastikan bahwa nilai yang tercatat sesuai dengan fakta. Meskipun dashboard tidak dapat secara otomatis memverifikasi data dengan sumber eksternal, ia menyediakan alat bantu bagi pengguna untuk melakukan pemeriksaan akurasi:

- Analisis Deskriptif:

Menu "Statistik Deskriptif" memungkinkan pengguna untuk dengan cepat melihat ringkasan statistik seperti nilai minimum, maksimum, dan rata-rata. Jika sebuah variabel persentase seperti POVERTY menunjukkan nilai maksimum di atas 100, pengguna dapat langsung mengidentifikasi adanya masalah akurasi data.

- Visualisasi Distribusi:

Dengan menggunakan Histogram atau Box Plot di menu "Visualisasi", pengguna dapat memeriksa distribusi data. Nilai yang berada sangat jauh dari pusat distribusi (pencilan ekstrim) dapat menandakan kesalahan input data yang perlu diperiksa ulang kebenarannya, meskipun secara teknis masih dalam rentang yang mungkin.

- Validasi Spasial:

Pada "Analisis Peta SOVI", data koordinat yang telah diintegrasikan divisualisasikan pada peta leaflet. Jika pengguna melihat sebuah titik data untuk kabupaten di Jawa muncul di wilayah Papua, ini adalah validasi visual yang kuat bahwa terdapat ketidaktepatan pada data koordinat.

- Konsistensi Data

Konsistensi data menjamin bahwa data tidak memiliki kontradiksi logis di seluruh dataset. STATeddy mendukung pemeriksaan konsistensi melalui:

- Konsistensi Fungsional:

Dashboard memastikan konsistensi selama sesi analisis. Ketika pengguna melakukan transformasi pada sebuah variabel, variabel baru tersebut akan tersedia secara konsisten di semua menu analisis lainnya. Logika `updateSelectInput` yang dinamis memastikan bahwa hanya variabel yang relevan dan konsisten dengan tipe data yang dibutuhkan yang akan muncul, mencegah error akibat inkonsistensi input.

- Pemeriksaan Hubungan Logis:



Pengguna dapat menggunakan "Correlation Plot" di menu "Visualisasi" untuk memeriksa konsistensi hubungan antar variabel. Misalnya, variabel LOWEDU (pendidikan rendah) dan POVERTY (kemiskinan) secara logis seharusnya memiliki korelasi positif. Jika plot menunjukkan korelasi negatif yang kuat, ini adalah sinyal adanya inkonsistensi data yang perlu diinvestigasi lebih lanjut.

- Kelengkapan Data

Pemeriksaan ini bertujuan untuk memastikan semua data yang dibutuhkan sudah lengkap dan tidak ada yang hilang.

- Deteksi Eksplisit:

Dashboard STATeddy secara langsung menangani ini di menu "Manajemen Data". Kotak "Data Summary" dengan jelas menyatakan jumlah *missing values* dalam dataset. Hal ini memberikan gambaran langsung kepada pengguna mengenai tingkat kelengkapan data mereka.

- Penanganan Otomatis saat Analisis:

Banyak fungsi analisis di R yang digunakan oleh dashboard (seperti `lm()`, `t.test()`, `cor()`) secara inheren melakukan validasi kelengkapan dengan mengabaikan baris yang memiliki data tidak lengkap (`complete.cases`) pada variabel yang sedang dianalisis. Ini memastikan bahwa hasil perhitungan statistik tidak terdistorsi oleh data yang hilang.

- Keunikan Data

Keunikan data memastikan tidak ada duplikasi observasi atau entitas.

- Inspeksi Visual:

Fitur utama untuk validasi keunikan adalah tabel interaktif (`DT::dataTableOutput`) di menu "Manajemen Data". Pengguna dapat mengurutkan data berdasarkan kolom kunci (misalnya `DISTRICTCODE`). Dengan melakukan pengurutan, duplikasi pada data akan muncul sebagai baris yang berurutan dan dapat dengan mudah diidentifikasi secara visual.

- Format Data

Validasi format memastikan setiap data memiliki tipe yang benar (misalnya, angka sebagai numerik, tanggal sebagai date).

- Penanganan Saat Impor:

Saat pengguna mengunggah file (CSV, Excel, SPSS), fungsi di R seperti `read.csv()` atau `haven::read_sav()` secara otomatis mencoba mengurai dan menetapkan format data yang paling sesuai untuk setiap kolom.

- Konversi Cerdas:

Kode di `app.R` bahkan memiliki logika untuk mencoba mengonversi kolom karakter yang berisi angka menjadi format numerik, yang merupakan langkah validasi format proaktif.

- Pencegahan Kesalahan Tipe Data:

Seperti yang disebutkan sebelumnya, antarmuka dashboard mencegah pengguna menggunakan variabel dengan format yang salah untuk analisis tertentu. Anda tidak bisa memilih variabel teks untuk dihitung rata-ratanya. Ini adalah bentuk penegakan format data yang sangat efektif.

## 6. Menentukan Objek Data

Menentukan objek data adalah proses strategis di mana analis memilih data yang paling relevan (baik kolom maupun baris) untuk menjawab pertanyaan penelitian. Dashboard STATeddy secara fundamental dirancang untuk memberdayakan pengguna dalam proses ini.

- Tujuan Utama dan Relevansi Penelitian

Dashboard ini memungkinkan pengguna untuk secara langsung menghubungkan tujuan penelitian mereka dengan data yang digunakan.

- Antarmuka Berbasis Tujuan:

Struktur menu dashboard (misalnya, "Uji Rata-rata", "ANOVA", "Regresi Linear") secara alami mengarahkan pengguna untuk memilih alat yang sesuai dengan tujuan mereka.

- Pemilihan Variabel Dinamis:

Di setiap menu analisis, pengguna diberikan kontrol penuh melalui widget `selectInput` untuk memilih variabel yang relevan dengan hipotesis mereka. Jika tujuannya adalah memprediksi POVERTY (kemiskinan), pengguna akan memilihnya sebagai "Variabel Dependen" di menu "Regresi Linear" dan memilih variabel-variabel lain yang secara teoretis relevan sebagai "Variabel Independen". Ini adalah inti dari penentuan objek data berdasarkan relevansi.

- Kualitas Data

Keputusan dalam menentukan objek data sangat dipengaruhi oleh kualitas data yang telah dievaluasi pada tahap sebelumnya.

- Seleksi Berdasarkan Kualitas:

Setelah pengguna menelaah (Tahap 4) dan memvalidasi (Tahap 5) data, mereka dapat membuat keputusan yang terinformasi. Jika ditemukan bahwa variabel FAMILYSIZE memiliki banyak *missing values* atau *outlier* yang tidak dapat dijelaskan, pengguna dapat memutuskan untuk tidak menyertakannya dalam model regresi mereka. Proses seleksi aktif ini, yang didasarkan pada penilaian kualitas, adalah bagian penting dari penentuan objek data.

- Ketersediaan Data

Ketersediaan data adalah prasyarat dasar, dan dashboard ini membuatnya sangat transparan.

- Refleksi Ketersediaan di UI:

Pilihan yang tersedia di semua `selectInput` (pemilih variabel) di seluruh aplikasi adalah cerminan langsung dari data yang tersedia di dataset yang sedang aktif. Jika pengguna mengunggah dataset kustom, semua pilihan ini

akan diperbarui secara dinamis. Ini memastikan bahwa pengguna hanya dapat memilih dari objek data yang benar-benar tersedia.

- Cakupan Wilayah dan Rentang Waktu

Parameter ini mendefinisikan batas-batas dari objek data.

- Cakupan Wilayah:

Untuk data default, cakupan wilayah didefinisikan dengan jelas, yaitu 514 kabupaten/kota di seluruh Indonesia. Informasi ini dipertegas di menu "Metadata", sehingga pengguna memahami batasan geografis dari analisis mereka.

- Rentang Waktu:

Data SOVI yang digunakan adalah data *cross-section* dari tahun 2017. Sama seperti cakupan wilayah, rentang waktu ini dikomunikasikan dengan jelas. Meskipun pengguna tidak bisa mengubah rentang waktu data default, kesadaran akan parameter ini sangat penting untuk interpretasi hasil yang benar. Pengguna yang mengunggah data *time-series* mereka sendiri secara implisit menentukan rentang waktu dari objek data yang mereka gunakan.