

VISUALISASI DAN STATISTIK DESKRIPTIF BESERTA ANALISIS DATASET STATUS GIZI BALITA TAHUN 2022 PER PROVINSI DI INDONESIA MENGGUNAKAN PEMROGRAMAN R

Faqih Rifaldy^{1*}, M N Alfi Syahrin², M Irsan Prayoga³, Fajar Syakbani⁴, Mhd Furqan⁵

^{1,2,3,4,5}Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara; Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara 20353

Keywords:

Status Gizi Balita;
Analisis Statistik Deskriptif;
Visualisasi Data;
Pemrograman R

Correspondent Email:

faqihrifaldy03@gmail.com

Abstrak. Dengan penekanan pada analisis statistik deskriptif dan visualisasi data menggunakan pemrograman R, penelitian ini mengkaji kesehatan gizi balita Indonesia pada tahun 2022. Perkembangan dan kualitas hidup generasi mendatang sangat dipengaruhi oleh masalah gizi balita, seperti stunting dan wasting, yang terus menjadi perhatian utama. Tujuan dari penelitian ini adalah untuk menggunakan statistik deskriptif dan visualisasi data untuk menyediakan data gizi balita dari 34 provinsi dengan cara yang lebih jelas. Metodologi penelitian ini mencakup pengumpulan informasi dari publikasi resmi Badan Pusat Statistik, seperti proporsi anak dalam kategori status gizi baik, buruk, rendah, dan gizi lebih. Perangkat lunak RStudio dengan paket ggplot2, dplyr, dan cluster digunakan untuk melakukan analisis. Histogram, boxplot, scatter plot, analisis korelasi, dan pengelompokan provinsi berdasarkan karakteristik data gizi merupakan beberapa metode yang digunakan. Temuan-temuan yang diperoleh menunjukkan adanya variasi regional yang mencolok dalam distribusi status gizi, dengan daerah-daerah seperti Papua dan Nusa Tenggara Timur memiliki tingkat gizi kurang dan gizi buruk tertinggi. Menurut analisis korelasi, gizi kurang dan gizi buruk berkorelasi negatif dengan kategori gizi baik. Menemukan pola tersembunyi dan pencilan yang sulit dikenali dalam tabel numerik dimungkinkan oleh visualisasi data. Studi ini menyoroti betapa pentingnya mendukung pengambilan keputusan berbasis bukti dengan statistik deskriptif dan visualisasi data. Temuan dari penelitian ini dapat digunakan sebagai panduan untuk membuat strategi berbasis data yang lebih efisien yang akan meningkatkan gizi balita.



JITET is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Abstract. With an emphasis on descriptive statistical analysis and data visualization using R programming, this study examines the nutritional health of Indonesia's under-fives in 2022. The development and quality of life of future generations is greatly influenced by under-five nutrition issues, such as stunting and wasting, which continue to be a major concern. The objective of this study is to use descriptive statistics and data visualization to provide under-five nutrition data from 34 provinces in a clearer way. The research methodology involved collecting information from official publications of the Central Bureau of Statistics, such as the proportion of children in the good, poor, low, and over-nutrition status categories. RStudio software with ggplot2, dplyr, and cluster packages was used to conduct the analysis. Histograms, boxplots, scatter plots, correlation analysis, and clustering of provinces based on nutrition data characteristics were among the methods used. The findings show that there are marked regional variations in the

distribution of nutritional status, with areas such as Papua and East Nusa Tenggara having the highest levels of undernutrition and malnutrition. According to the correlation analysis, undernutrition and malnutrition are negatively correlated with the good nutrition category. Finding hidden patterns and hard-to-spot outliers in numerical tables is made possible by data visualization. This study highlights the importance of supporting evidence-based decision-making with descriptive statistics and data visualization. The findings from this study can be used as a guide to create more efficient data-driven strategies that will improve under-five nutrition.

1. PENDAHULUAN

Di Indonesia, malnutrisi pada anak balita telah lama menjadi perhatian utama dalam pembangunan kesehatan masyarakat [1]. Malnutrisi membuat anak di bawah lima tahun lebih rentan terhadap masalah pertumbuhan, yang dapat berdampak jangka panjang pada produktivitas, daya tahan tubuh, dan perkembangan kognitif orang dewasa. Menurut Global Nutrition Report dan data Kementerian Kesehatan, prevalensi stunting dan wasting di Indonesia masih lebih tinggi daripada yang ditetapkan oleh WHO. Hal ini menunjukkan bahwa hambatan struktural, seperti akses yang tidak merata terhadap makanan padat gizi, sanitasi yang tidak memadai, dan kurangnya layanan kesehatan dasar di beberapa komunitas, terus menghambat upaya untuk meningkatkan status gizi [2]. Status gizi anak juga dipengaruhi oleh faktor sosioekonomi, termasuk urbanisasi, pendapatan rumah tangga, dan tingkat pendidikan orang tua [3].

Meskipun pemerintah secara rutin mengumpulkan informasi mengenai kondisi gizi balita melalui survei seperti Riskesdas dan Susenas, penggunaan informasi ini untuk analisis eksplorasi dan visualisasi spasial masih sangat sedikit. Data tersebut biasanya ditampilkan dalam bentuk tabel statis yang sulit dipahami oleh orang non-teknis atau pembuat kebijakan di tingkat daerah. Sebenarnya, representasi visual dari pola distribusi status gizi dapat memberikan lebih banyak informasi tentang wilayah yang paling berisiko serta bagaimana pola tersebut bervariasi dari tahun ke tahun. Selain itu, kita juga dapat menemukan hubungan antara kategori status gizi (baik, buruk, kurang, dan lebih) dengan menggunakan teknik berbasis analisis data kontemporer, serta mengelompokkan provinsi ke dalam kelompok tertentu sesuai dengan karakteristik gizi yang sebanding. Oleh karena itu, untuk mengubah

data yang tidak terstruktur menjadi pengetahuan yang dapat digunakan untuk mengambil keputusan, diperlukan pendekatan ilmiah.

Sayangnya, tanpa menggunakan visualisasi data yang komunikatif, data status gizi biasanya dipublikasikan dalam bentuk tabel numerik yang tekstual dan tidak fleksibel. Visualisasi data sebenarnya memberikan kontribusi yang signifikan terhadap kemudahan pemahaman informasi [4]. Bergantung pada tujuan analisis, berbagai visualisasi dapat digunakan: scatter plot menunjukkan hubungan antara dua variabel kuantitatif [5]; histogram menunjukkan distribusi frekuensi [6]; boxplot menunjukkan distribusi data, nilai rata-rata, dan pencilan [7]; pie chart berguna untuk menunjukkan proporsi [8]; dan barplot digunakan untuk membandingkan nilai antar kategori [9], seperti provinsi. Visualisasi semacam ini dapat menyoroti tren yang tersembunyi dari pandangan dalam tabel standar. Namun, untuk meringkas data dan memberikan gambaran umum tentang properti dataset, statistik deskriptif seperti rata-rata, median, deviasi standar, dan rentang nilai sangat penting [10].

Tujuan dari penelitian ini adalah untuk menginvestigasi dan mengevaluasi data status gizi balita di 34 provinsi di Indonesia pada tahun 2022. Tujuan utamanya adalah untuk menyampaikan data dengan menggunakan analisis statistik deskriptif dan visualisasi yang mendidik. Hal ini mencakup distribusi frekuensi, sebaran (rentang, standar deviasi), dan pengukuran sentral (mean) dari kategori status gizi. Untuk mengetahui hubungan antara prevalensi gizi buruk, kurang, baik, dan lebih, studi korelasi antar kategori status gizi juga dilakukan. Diharapkan bahwa temuan-temuan dari penelitian ini akan menjadi bahan pertimbangan dalam menyusun kebijakan

berbasis data, selain memberikan gambaran empiris tentang status gizi balita yang ada.

Program open source R, yang terkenal dengan analisis statistik dan visualisasi data, digunakan dalam metodologi eksplorasi data kuantitatif studi ini. “Status Gizi Balita Menurut Provinsi Tahun 2022,” sebuah publikasi resmi dari Badan Pusat Statistik (BPS), digunakan sebagai dataset. Data tersebut menunjukkan persentase dari setiap kategori status gizi-baik, kurang, buruk, dan lebih-untuk 34 provinsi di Indonesia. Statistik deskriptif pertama-tama diperiksa untuk menentukan fitur keseluruhan dari data. Diagram batang, diagram kotak, dan peta panas yang menunjukkan korelasi kategori kemudian digunakan untuk memvisualisasikan distribusi kategori. Semua proses analisis dilakukan di RStudio, memanfaatkan paket-paket seperti ggplot2, dplyr, cluster, dan factoextra. Pendekatan ini tidak hanya memperkuat validitas analisis, tetapi juga menghasilkan visualisasi yang dapat dipahami dengan mudah dan disajikan secara profesional dalam publikasi ilmiah.

2. TINJAUAN PUSTAKA

2.1 Status Gizi Balita di Indonesia

Status gizi pada balita merupakan salah satu tolok ukur utama dalam menggambarkan taraf kesehatan dan kesejahteraan anak di suatu negara. Organisasi Kesehatan Dunia (WHO) mengelompokkan status gizi berdasarkan indikator seperti berat badan, tinggi badan, dan usia anak. Di Indonesia, perhatian terhadap status gizi balita cukup besar karena prevalensi masalah gizi seperti stunting, wasting, dan underweight masih tergolong tinggi dan dapat berpengaruh pada kualitas generasi mendatang. Informasi terkait status gizi balita diperoleh melalui berbagai sumber resmi, seperti Survei Status Gizi Balita Indonesia (SSGBI) dan sistem e-PPGBM. Data yang dihimpun dari kegiatan tersebut menjadi acuan utama dalam menyusun strategi penanganan gizi, menentukan arah kebijakan pemerintah, serta menilai keberhasilan program-program kesehatan yang telah dilaksanakan [11].

2.2 Statistik Deskriptif dalam Analisis Data Kesehatan

Statistik deskriptif merupakan pendekatan awal dalam analisis data yang bertujuan untuk menyajikan, merangkum, dan menggambarkan ciri-ciri utama dari suatu kumpulan data. Dalam konteks status gizi balita, metode ini berguna untuk menghitung nilai-nilai seperti rata-rata, median, modus, standar deviasi, serta distribusi frekuensi pada tiap provinsi. Penggunaan statistik deskriptif membantu dalam memahami pola umum, mengenali data yang menyimpang (outlier), dan melihat bagaimana data tersebar sebelum dilakukan analisis yang lebih mendalam [12].

2.3 Visualisasi Data

Visualisasi data adalah teknik untuk menyajikan informasi dalam bentuk visual seperti grafik, diagram, maupun peta agar lebih mudah dimengerti. Dalam penelitian ini, visualisasi berperan untuk memperlihatkan kondisi status gizi balita di tiap provinsi di Indonesia. Berbagai jenis visual seperti diagram batang, diagram lingkaran, dan heatmap sangat efektif dalam menyampaikan informasi kepada pengambil kebijakan maupun masyarakat luas. Selain itu, penggunaan peta tematik dapat memperjelas perbedaan geografis dalam distribusi status gizi antar wilayah [13].

2.4 Pemrograman R dalam Analisa Data

R adalah bahasa pemrograman open-source yang banyak digunakan dalam bidang analisis data dan statistika. Bahasa ini dilengkapi dengan berbagai paket, seperti ggplot2 untuk membuat visualisasi data, dplyr untuk memudahkan manipulasi data, serta tidyverse yang menyatukan berbagai alat untuk pengolahan data secara sistematis dan efisien. R memiliki kelebihan dalam hal kemampuan mereproduksi analisis, fleksibilitas penggunaan, dan dokumentasi yang lengkap, sehingga sangat cocok digunakan untuk menganalisis data status gizi secara komprehensif [14].

2.5 Analisis Dataset Kesehatan Publik

Analisis data dalam konteks kesehatan masyarakat bertujuan untuk mengidentifikasi pola, keterkaitan, serta kegagalan dalam data populasi. Metode seperti analisis korelasi, distribusi, dan pengelompokan (clustering) sering dimanfaatkan untuk memahami penyebaran status gizi dan faktor-faktor yang

memengaruhinya. Dengan menggunakan pendekatan yang berlandaskan data, pengambilan keputusan oleh pemerintah maupun lembaga kesehatan dapat dilakukan secara lebih tepat dan didasarkan pada bukti yang kuat (evidence-based policy) [15].

3. METODE PENELITIAN

3.1. Jenis Penelitian

Berdasarkan data yang tersedia di setiap provinsi di Indonesia, studi deskriptif ini mencoba untuk mengkaraktirasi secara metodis kesehatan gizi balita pada tahun 2022. Tanpa menggunakan eksperimen atau pengujian hipotesis, teknik deskriptif digunakan untuk memberikan gambaran yang realistis dan akurat tentang kondisi aktual yang ada [16].

3.2. Teknik Pengumpulan Data

Data yang digunakan merupakan data sekunder yang diperoleh dari Survei Status Gizi Indonesia (SSGI) tahun 2022, yang dirilis oleh Kementerian Kesehatan Republik Indonesia. Dataset ini mencakup informasi terkait indikator status gizi balita, antara lain:

- Stunting (pendek berdasarkan tinggi badan menurut usia)
- Wasting (kurus berdasarkan berat badan menurut tinggi badan)
- Underweight (berat badan rendah menurut usia)
- Overweight.

Data disusun dan dipublikasikan dalam bentuk agregat provinsi dan tersedia untuk diunduh secara publik

3.3. Pengolahan Data

Data diperoleh dengan cara mengakses dan mengunduh dataset resmi dari sumber terbuka pemerintah. Data tersebut disimpan dalam format digital (CSV/Excel), kemudian diolah dan dianalisis menggunakan perangkat lunak pengolahan data. Karena bersifat sekunder, tidak dilakukan proses pengumpulan data primer seperti wawancara atau observasi langsung di lapangan [17].

3.4. Analisis Data

Data dianalisis menggunakan statistik deskriptif, yang memperoleh :

- Penghitungan rata-rata, median, minimum, dan maksimum nilai pada tiap indikator.
- Pengelompokan data berdasarkan interval dan kategori gizi.

3.5. Visualisasi Data

Visualisasi data digunakan untuk menyajikan informasi status gizi balita secara lebih komunikatif dan mudah dipahami. Visualisasi membantu mengidentifikasi pola, perbandingan antar provinsi, serta anomali data [18]. Beberapa jenis visualisasi yang digunakan dalam penelitian ini meliputi:

- Scatter Plot
- Boxplot
- Histogram
- Bar Plot
- Pie Chart

3.6. Perangkat lunak dan Alat Bantu

Analisis dilakukan menggunakan bahasa pemrograman R dengan bantuan beberapa paket statistik dan visualisasi:

- dplyr dan tidyr untuk manipulasi data
- summarytools untuk menghasilkan tabel statistik deskriptif
- ggplot2 untuk membuat grafik visual
- tmap dan sf untuk membuat peta distribusi spasial status gizi per provinsi
- readr untuk membaca file data

Pemilihan alat bantu visualisasi ini bertujuan untuk mendukung data-driven storytelling, memastikan bahwa hasil analisis dapat disampaikan dengan cara yang informatif dan menarik secara visual.

4. HASIL DAN PEMBAHASAN

4.1 Persiapan Dataset

Untuk memahami dan menerapkan konsep dasar analisis data, khususnya visualisasi dan statistik deskriptif, digunakan sebuah dataset berjudul “*Status Gizi Balita 2022*” yang berisi informasi persentase status gizi pada balita di 34 provinsi di Indonesia. Dataset ini dipilih sebagai contoh karena memiliki struktur yang sederhana namun cukup representatif dalam menggambarkan variabel numerik dan kategorikal. Analisis dilakukan menggunakan pemrograman R dengan tujuan untuk memperkenalkan teknik-teknik eksplorasi data, termasuk pembuatan grafik seperti bar chart,

boxplot, histogram, serta perhitungan ukuran statistik deskriptif seperti rata-rata, median, modus, simpangan baku, dan rentang nilai. Fokus utama analisis ini bukan pada isi atau interpretasi dari data gizi itu sendiri, melainkan pada penerapan teknis metode eksplorasi data dalam konteks pembelajaran atau demonstrasi pemrograman statistik menggunakan R.

Provinsi	Gizi Baik (%)	Gizi Kurang (%)	Gizi Buruk (%)	Gizi Lebih (%)
1. Aceh	84,7	22,3	5,5	7,5
2. Sumatera Utara	88,9	21,1	5,1	6,9
3. Sumatera Barat	89,3	20,0	4,0	6,7
4. Riau	88,0	21,1	4,8	6,3
5. Jambi	88,5	21,5	5,2	6,8
6. Sumatera Selatan	88,0	23,0	5,4	6,3
7. Bengkulu	84,4	22,9	6,2	6,3
8. Lampung	83,8	23,5	6,8	6,3
9. Kepulauan Bangka Belitung	89,3	19,9	4,3	6,3
10. Kepulauan Riau	87,4	21,4	5,0	6,2
11. DKI Jakarta	75,5	17,2	3,0	4,3
12. Jawa Barat	73,3	20,6	4,1	5,0
13. Jawa Tengah	69,2	22,1	4,8	5,1
14. DI Yogyakarta	75,1	20,8	3,3	5,7
15. Jawa Timur	68,4	21,5	4,8	5,3
16. Banten	66,1	21,6	5,5	6,8
17. Bali	70,9	20,4	5,8	5,2
18. Nusa Tenggara Barat	65,7	22,3	5,8	6,4
19. Nusa Tenggara Timur	62,1	24,3	7,8	5,8
20. Kalimantan Barat	63,5	22,1	6,4	6,0
21. Kalimantan Tengah	66,4	21,5	6,0	7,1
22. Kalimantan Selatan	66,6	21,3	5,8	6,3
23. Kalimantan Timur	67,2	20,8	5,5	6,5
24. Kalimantan Utara	68,9	21,1	5,9	6,1
25. Sulawesi Utara	68,3	20,2	5,0	6,0
26. Sulawesi Tengah	64,1	23,5	6,4	6,0
27. Sulawesi Selatan	64,8	22,7	6,0	6,3
28. Sulawesi Tenggara	63,3	22,0	6,3	6,4
29. Gorontalo	66,2	21,6	6,1	6,1
30. Sulawesi Barat	64,3	22,4	6,3	6,8
31. Maluku	68,8	21,2	6,0	6,0
32. Maluku Utara	67,0	20,8	6,1	6,0
33. Papua Barat	63,0	23,5	6,7	7,8
34. Papua	63,2	23,9	7,4	8,3

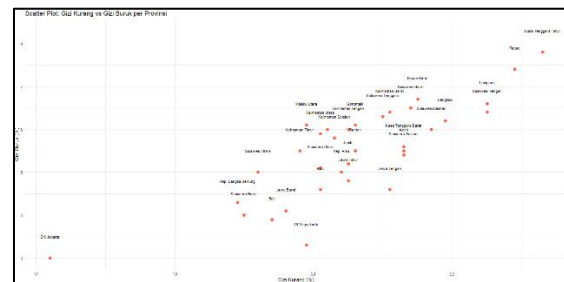
Gambar 1 Dataset Status Gizi Balita 2022

4.2 Visualisasi

Visualisasi data adalah inti dan merupakan sarana penting untuk melakukan analisis data, dan ketika makna sudah ditemukan dan dimengerti, makna yang didapatkan tersebut akan di komunikasikan kepada orang lain [19].

4.2.1 Visualisasi Scatter Plot

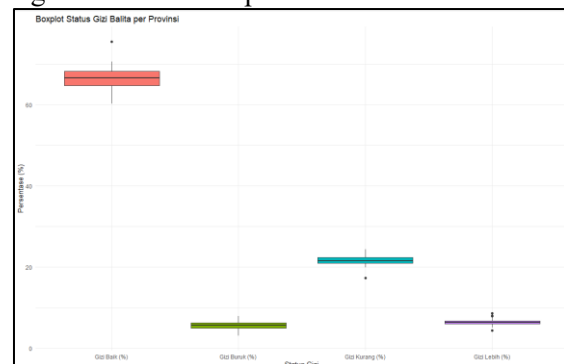
Scatter plot dalam gambar selanjutnya menunjukkan hubungan antara persentase gizi kurang dan gizi buruk di berbagai provinsi di Indonesia. Setiap titik merah pada grafik mewakili sebuah provinsi, dengan posisi horizontal (sumbu x) menunjukkan persentase gizi kurang dan posisi vertikal (sumbu y) menunjukkan persentase gizi buruk. Provinsi dengan angka gizi kurang yang tinggi cenderung memiliki tingkat gizi buruk yang lebih tinggi pula, menunjukkan kemungkinan korelasi antara kedua variabel tersebut. Dengan label sumbu x sebagai "Gizi Kurang (%)" dan sumbu y sebagai "Gizi Buruk (%)". Nama provinsi ditampilkan di dekat titik data untuk mempermudah identifikasi. Beberapa provinsi seperti Nusa Tenggara Timur, Papua, dan Lampung tampak memiliki persentase gizi buruk dan gizi kurang yang tinggi, sedangkan DKI Jakarta memiliki angka yang lebih rendah untuk kedua indikator tersebut.



Gambar 1 Visualisasi scatter plot

4.2.2 Visualisasi Boxplot

Boxplot pada gambar selanjutnya menggambarkan distribusi empat kategori status gizi balita per provinsi, yaitu Gizi Baik (%), Gizi Kurang (%), Gizi Buruk (%), dan Gizi Lebih (%). Sumbu x menampilkan kategori status gizi, sedangkan sumbu y menunjukkan persentase (%). Setiap boxplot memvisualisasikan nilai median, kuartil pertama dan ketiga, serta outlier yang ada pada setiap kategori. Misalnya, boxplot Gizi Baik (%) menunjukkan rentang persentase yang relatif lebar dengan nilai median yang cukup tinggi, yang mengindikasikan bahwa mayoritas balita berada dalam kategori gizi yang baik, meskipun terdapat beberapa nilai ekstrim di atas rentang normal. Di sisi lain, boxplot Gizi Kurang (%) dan Gizi Buruk (%) memperlihatkan nilai median yang lebih rendah, dengan Gizi Buruk (%) menunjukkan distribusi persentase yang sangat rendah, namun disertai beberapa outlier yang mengindikasikan situasi khusus di beberapa provinsi. Untuk Gizi Lebih (%), walaupun nilai median tetap rendah, adanya titik-titik outlier di atas rentang normal menyiratkan variasi yang signifikan di antara provinsi.



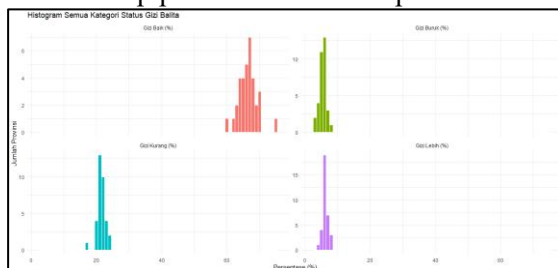
Gambar 2 Visualisasi box plot

4.2.3 Visualisasi Histogram

Histogram pada gambar tersebut menunjukkan distribusi data dari empat

kategori status gizi balita di masing-masing provinsi, yaitu Gizi Baik (%), Gizi Kurang (%), Gizi Buruk (%), dan Gizi Lebih (%). Setiap panel histogram mewakili satu kategori gizi dan menampilkan jumlah provinsi (sumbu y) yang memiliki rentang persentase tertentu (sumbu x).

Pada panel Gizi Baik (%), terlihat bahwa sebagian besar provinsi memiliki persentase gizi baik yang berada di kisaran tengah, dengan puncak distribusi mendekati 50%, menunjukkan bahwa mayoritas daerah menyajikan kondisi gizi yang dianggap baik. Sedangkan pada panel Gizi Buruk (%) dan Gizi Lebih (%), jumlah provinsi yang tercatat relatif sedikit dengan persentase di bagian bawah grafik, yang mengindikasikan bahwa kondisi gizi buruk dan gizi lebih jarang terjadi di antara provinsi. Panel Gizi Kurang (%) menampilkan distribusi yang cenderung lebih terpusat di sekitar nilai tertentu (misalnya sekitar 20%), yang mencerminkan kondisi gizi kurang yang lebih sering terjadi namun dengan variasi yang terbatas di antara daerah. Format histogram yang menggunakan bin dengan lebar 1 persen memungkinkan visualisasi yang detail dan spesifik dari distribusi setiap status gizi. Selain itu, pemisahan ke dalam panel-panel (facet_wrap) dengan skala sumbu y yang disesuaikan secara independen memperjelas pola dan perbedaan antara kategori status gizi, sehingga memudahkan analisis kondisi gizi balita di tiap provinsi secara komparatif.

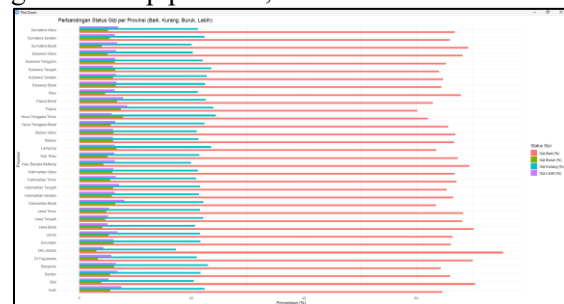


Gambar 3 Visualisasi histogram

4.2.4 Visualisasi Bar Plot

Bar plot horizontal pada gambar tersebut menyajikan perbandingan empat kategori status gizi—yaitu Gizi Baik (%), Gizi Kurang (%), Gizi Buruk (%), dan Gizi Lebih (%)—di setiap provinsi. Setiap provinsi ditampilkan sebagai baris pada sumbu y, sedangkan pada sumbu x tertera persentase masing-masing kategori. Setiap kategori diberi warna yang berbeda sehingga memudahkan dalam mengidentifikasi dan membandingkan kondisi gizi antar

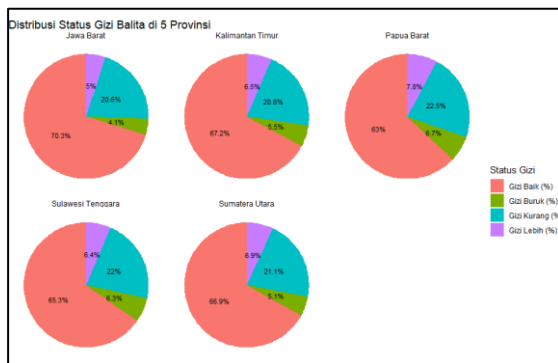
provinsi. Karena penggunaan fungsi reorder(Provinsi, Persentase), provinsi diurutkan berdasarkan nilai persentasenya, memberikan gambaran yang lebih jelas mengenai perbedaan, misalnya, provinsi dengan persentase Gizi Baik yang tinggi atau Gizi Kurang/Buruk yang menonjol. Dengan tata letak horizontal serta pengaturan posisi (dodge) pada bar plot, visualisasi ini menyederhanakan perbandingan langsung antara kategori status gizi di setiap provinsi,



Gambar 4 Visualisasi bar plot

4.2.5 Visualisasi Pie Chart

Pie chart pada gambar berikut menggambarkan distribusi status gizi balita di lima provinsi pilihan, yaitu Sumatera Utara, Jawa Barat, Kalimantan Timur, Sulawesi Tenggara, dan Papua Barat, dengan masing-masing pie chart menampilkan empat kategori status gizi: Gizi Baik, Gizi Kurang, Gizi Buruk, dan Gizi Lebih. Dalam setiap pie chart, segmen-segmen berwarna mewakili persentase balita yang termasuk dalam masing-masing kategori, sehingga mudah untuk melihat proporsi relatif dari setiap status gizi di provinsi tersebut. Dengan menggunakan metode facet wrap, setiap provinsi memiliki tampilan pie chart tersendiri, yang memungkinkan perbandingan langsung antar provinsi. Label persentase yang ditampilkan pada setiap segmen membuat interpretasi data menjadi lebih detail dan informatif, sehingga pengguna dapat dengan cepat mengidentifikasi apakah mayoritas balita di suatu provinsi memiliki status gizi baik atau ada proporsi signifikan yang berada pada kategori gizi kurang, gizi buruk, atau gizi lebih.



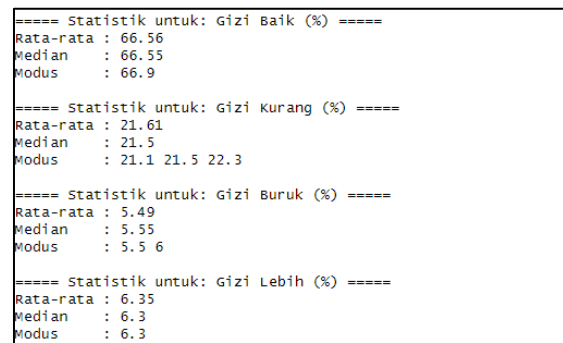
Gambar 5 Visualisasi pie chart

4.3 Statistik Deskriptif

Statistik deskriptif atau statistik deduktif adalah bagian dari statistik mempelajari cara pengumpulan data dan penyajian data sehingga mudah dipahami. Statistik deskriptif hanya berhubungan dengan hal menguraikan atau memberikan keterangan-keterangan mengenai suatu data atau keadaan atau fenomena [20].

4.3.1 Statistik Deskriptif Mean, Median, dan Modus

Analisis statistik deskriptif yang ditampilkan pada gambar berikutnya memberikan gambaran mendalam mengenai tiga ukuran pemusatan data—rata-rata (mean), median, dan modus—untuk empat kategori status gizi, yaitu Gizi Baik (%), Gizi Kurang (%), Gizi Buruk (%), dan Gizi Lebih (%). Dalam kode yang digunakan, dilakukan perulangan (loop) terhadap masing-masing kolom data sehingga secara sistematis dihitung nilai rata-rata untuk mengetahui kecenderungan umum nilai, median sebagai nilai tengah yang membagi data ke dalam dua bagian yang sama, serta modus yang merupakan nilai yang paling sering muncul, memungkinkan pengidentifikasian tren dominan pada distribusi data. Dengan melakukan perhitungan ini, kita bisa dengan mudah melihat apakah distribusi data simetris atau condong ke arah tertentu, serta mengidentifikasi outlier atau nilai yang memiliki frekuensi tinggi.



Gambar 6 Statistik deskriptif mean, median, dan modus

4.3.2 Statistik Deskriptif Range, Varians, dan Standar Deviasi

Gambar selanjutnya menampilkan analisis statistik deskriptif untuk empat kategori status gizi balita yaitu Gizi Baik, Gizi Kurang, Gizi Buruk, dan Gizi Lebih. Dalam analisis ini, tiga ukuran sebaran data—range, varians, dan standar deviasi—dihitung untuk memberikan gambaran mendalam mengenai distribusi setiap kategori. Range menunjukkan perbedaan antara nilai maksimum dan minimum; misalnya, Gizi Baik memiliki range sebesar 15.3 yang mengindikasikan variasi yang cukup lebar antar nilai, sedangkan Gizi Lebih memiliki range 4.2 yang menunjukkan penyebaran yang lebih homogen. Varians mengukur seberapa jauh data tersebar dari nilai rata-rata, di mana kategori Gizi Baik kembali menunjukkan varians tertinggi sebesar 8.12, yang mengindikasikan adanya perbedaan signifikan di antara nilai-nilainya, sementara Gizi Lebih memiliki varians serendah 0.71. Sedangkan standar deviasi—sebagai akar kuadrat dari varians—memberikan ukuran penyimpangan data dari rata-rata dalam satuan yang sama, dengan nilai tertinggi 2.85 pada Gizi Baik dan terendah 0.85 pada Gizi Lebih.

```

> kolom_gizi <- c("Gizi Baik (%)", "Gizi Kurang (%)", "Gizi Buruk (%)", "Gizi Lebih (%)")
> for (kolom in kolom_gizi) {
+   nilai <- gizi[[kolom]]
+   nilai_range <- range(nilai, na.rm = TRUE)
+   cat("\nrange", kolom, ":", round(nilai_range[2] - nilai_range[1], 2), "\n")
+ }

Range Gizi Baik (%) : 15.3
Range Gizi Kurang (%) : 7.1
Range Gizi Buruk (%) : 4.8
Range Gizi Lebih (%) : 4.2
> for (kolom in kolom_gizi) {
+   nilai <- gizi[[kolom]]
+   varian <- var(nilai, na.rm = TRUE)
+   cat("\nvarians", kolom, ":", round(varian, 2), "\n")
+ }

varians Gizi Baik (%) : 8.12
varians Gizi Kurang (%) : 1.76
varians Gizi Buruk (%) : 1.19
varians Gizi Lebih (%) : 0.71
> for (kolom in kolom_gizi) {
+   nilai <- gizi[[kolom]]
+   stddev <- sd(nilai, na.rm = TRUE)
+   cat("\nstandar deviasi", kolom, ":", round(stddev, 2), "\n")
+ }

Standar Deviasi Gizi Baik (%) : 2.85
Standar Deviasi Gizi Kurang (%) : 1.33
Standar Deviasi Gizi Buruk (%) : 1.09
Standar Deviasi Gizi Lebih (%) : 0.85

```

Gambar 7 Statistik deskriptif range, varians, dan standar deviasi

4.3.3 Statistik Deskriptif Fungsi Summary

Gambar berikutnya menampilkan output dari fungsi summary() yang memberikan ringkasan statistik deskriptif untuk dataset gizi. Fungsi ini menghitung dan menampilkan statistik dasar dari masing-masing kolom numerik—yaitu Gizi Baik (%), Gizi Kurang (%), Gizi Buruk (%), dan Gizi Lebih (%)—yang mencakup nilai minimum, kuartil pertama, median, rata-rata, kuartil ketiga, dan nilai maksimum. Misalnya, pada kolom Gizi Baik (%), nilai terendah adalah 60.20 dan nilai tertinggi mencapai 75.50, dengan median sebesar 66.55 dan rata-rata hampir sama yaitu 66.56, yang menunjukkan distribusi data yang relatif simetris. Output serupa juga diberikan untuk kolom-kolom lainnya; Gizi Kurang (%) memiliki rentang nilai dari 17.20 hingga 24.30, sedangkan Gizi Buruk (%) dan Gizi Lebih (%) menunjukkan variasi masing-masing dengan pola sebaran yang juga dapat dibandingkan antara kuartil dan nilai ekstrim yang muncul.

```

> # Ringkasan statistik dasar semua kolom numerik
> summary(gizi)

```

Provinsi	Gizi Baik (%)	Gizi Kurang (%)	Gizi Buruk (%)	Gizi Lebih (%)
Length:34	Min. :60.20	Min. :17.20	Min. :3.000	Min. :4.300
Class :character	1st Qu.:64.72	1st Qu.:20.95	1st Qu.:4.850	1st Qu.:6.000
Mode :character	Median :66.55	Median :21.50	Median :5.550	Median :6.300
	Mean :66.56	Mean :21.61	Mean :5.488	Mean :6.347
	3rd Qu.:68.15	3rd Qu.:22.30	3rd Qu.:6.175	3rd Qu.:6.675
	Max. :75.50	Max. :24.30	Max. :7.800	Max. :8.500

Gambar 8 Statistik deskriptif fungsi summary

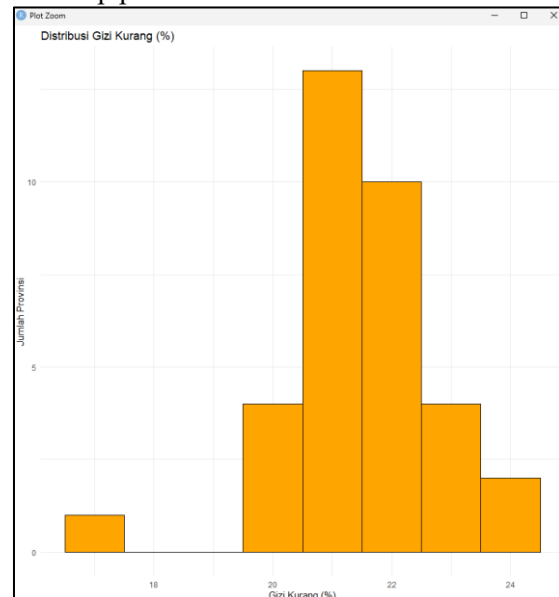
4.4 Analisis Data

Analisis data adalah bagian penting dari proses penelitian karena setelah pengolahan data pelatihan analisis data dengan software RStudio

dapat meningkatkan pengetahuan tentang pengolahan data [21].

4.4.1 Analisis Univariate

Analisis univariate pada gambar selanjutnya berfokus pada pemeriksaan distribusi data dari variabel "Gizi Kurang (%)" di berbagai provinsi. Visualisasinya menggunakan histogram dengan bin lebar 1, di mana setiap bar berwarna oranye dengan garis pembatas hitam, mempermudah pengamatan frekuensi provinsi yang memiliki persentase gizi kurang pada kisaran tertentu. Dari histogram ini, terlihat bahwa sebagian besar provinsi berpusat pada rentang nilai tertentu, dengan sebagian provinsi menunjukkan nilai di bawah, di sekitar, atau di atas titik konsentrasi utama distribusi. Untuk melengkapi visualisasi, analisis numerik dilakukan dengan menghitung nilai rata-rata, median, dan standar deviasi. Rata-rata memberikan gambaran umum mengenai nilai tengah distribusi, sedangkan median mengidentifikasi nilai tengah yang lebih tahan terhadap pencilan.

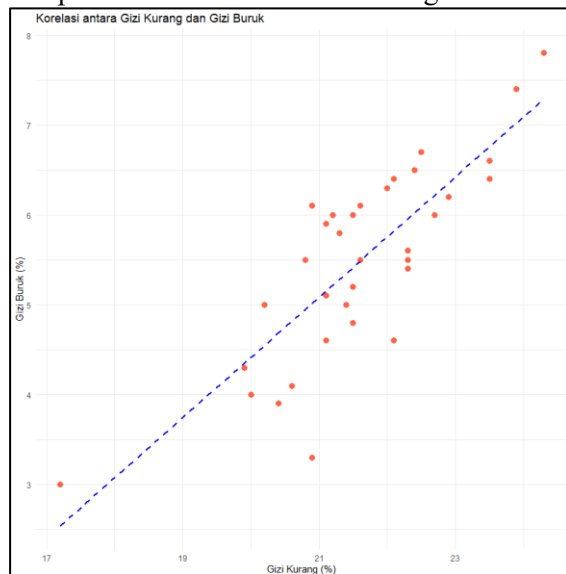


Gambar 9 Analisis univariate – 1 variabel

4.4.2 Analisis Bivariate

Analisis bivariate pada gambar berikutnya mengeksplorasi hubungan antara dua variabel kunci, yaitu persentase Gizi Kurang (%) dan Gizi Buruk (%) di berbagai provinsi. Visualisasinya menggunakan scatter plot yang menampilkan titik-titik data berwarna tomat, yang masing-masing mewakili nilai dua variabel per provinsi. Untuk menilai arah dan

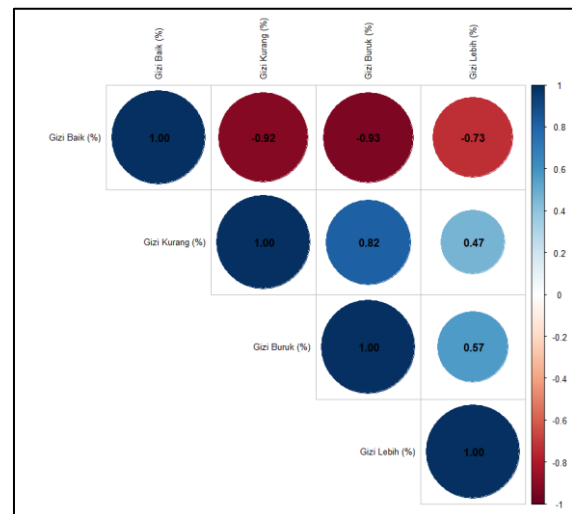
kekuatan hubungan tersebut, garis regresi linear ditambahkan dengan metode lm yang ditampilkan sebagai garis putus-putus biru, sehingga memudahkan pengamatan kecenderungan peningkatan atau penurunan antara kedua variabel. Selain itu, nilai korelasi dihitung secara numerik menggunakan fungsi cor, yang memberikan ukuran statistik tentang berapa kuat kedua variabel ini saling berkaitan.



Gambar 10 Analisis bivariate – 2 variabel

4.4.3 Analisis Multivariate

Analisis multivariate pada gambar ini mengeksplorasi hubungan simultan antar empat variabel status gizi—yaitu Gizi Baik (%), Gizi Kurang (%), Gizi Buruk (%), dan Gizi Lebih (%)—dengan menghitung matriks korelasi dari data numerik yang tersedia. Matriks ini mengungkapkan bahwa Gizi Baik (%) memiliki korelasi negatif yang sangat tinggi dengan Gizi Kurang (%) (-0.92) dan Gizi Buruk (%) (-0.93), yang menunjukkan bahwa peningkatan persentase gizi baik di suatu provinsi cenderung diiringi oleh penurunan persentase gizi kurang dan gizi buruk. Di sisi lain, terdapat korelasi positif yang kuat antara Gizi Kurang (%) dan Gizi Buruk (%) (0.82), mengisyaratkan bahwa dua indikator tersebut bergerak seiring dalam kondisi yang serupa. Gizi Lebih (%) menunjukkan hubungan negatif moderat dengan Gizi Baik (%) (-0.73) serta korelasi positif sedang dengan indikator gizi kurang (0.47) dan gizi buruk (0.57), menandakan dinamika yang berbeda dibandingkan indikator lainnya.



Gambar 11 Analisis multivariate

5. KESIMPULAN

- Berdasarkan analisis, distribusi status gizi berbeda menurut provinsi, dengan Papua dan Nusa Tenggara Timur menunjukkan frekuensi gizi kurang dan gizi buruk yang signifikan.
- Gizi kurang dan gizi buruk memiliki hubungan positif satu sama lain, sementara persentase gizi baik memiliki korelasi negatif yang kuat dengan keduanya.
- Menemukan pola distribusi dan pencilan yang sulit dilihat pada tabel numerik menjadi lebih mudah dengan penggunaan visualisasi data.
- Studi ini menyoroti nilai statistik deskriptif dan visualisasi data dalam memahami kerumitan data kesehatan masyarakat.
- Di daerah-daerah yang memiliki masalah gizi yang lebih banyak, hasil penelitian ini dapat membantu memandu saran kebijakan yang bertujuan untuk meningkatkan kondisi gizi balita.

UCAPAN TERIMA KASIH

Dataset “Status Gizi Balita pada tahun 2022,” yang menjadi dasar penelitian ini, disediakan oleh Badan Pusat Statistik (BPS), dan untuk itu penulis berterima kasih. Fakultas Sains dan Teknologi Universitas Islam Negeri Sumatera Utara juga penulis ucapkan terima

kasih karena telah menyediakan fasilitas dan dukungan akademik selama proses penelitian. Secara khusus, kami ingin mengucapkan terima kasih kepada rekan-rekan kami yang telah memberikan kontribusi yang signifikan dalam pembuatan makalah ini.

DAFTAR PUSTAKA

- [1] S. S. Nining Tunggal and M. Retno Heru, Setyorini; Fitriani, "Edukasi Gizi Seimbang Untuk Mencegah Malnutrisi Pada Anak Balanced," vol. 48, no. 2, pp. 39–62, 2021, [Online]. Available: www.ine.es
- [2] S. H. Khalid, S. Patimah, and A. Asrina, "Persepsi Masyarakat mengenai Penyebab dan Dampak Stunting di Kabupaten Majene Sulawesi Barat Tahun 2020," *J. Muslim Community Heal.* 2022, vol. 3, no. 4, pp. 80–94, 2022, [Online]. Available: <https://doi.org/10.52103/jmch.v3i4.1177> JournalHomepage:<https://pasca-umi.ac.id/index.php/jmch>
- [3] R. A. Rahmadani, R. Wahyuni, D. Arda, A. S. Musrah, and R. Sabriana, "Socioeconomic Factors with Nutritional Status of Toddlers," *J. Ilm. Kesehat. Sandi Husada*, vol. 12, no. 2, pp. 445–451, 2023, doi: 10.35816/jiskh.v12i2.1115.
- [4] D. S. Amaruloh, "Penerapan Visualisasi Data Pada Pd. Fokus Bandung," *J. Penelit. Mhs. Tek. dan Ilmu Komput.*, vol. 3, no. 1, pp. 44–52, 2023.
- [5] G. Joelio, L. Nugrah, and M. Saputra, "Penerapan Metode Service Quality Dalam Analisis Persepsi Konsumen Pada Pelayanan Pemesanan Barang E-Commerce Di Universitas Prima Indonesia," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 40–46, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2311.
- [6] M. S. A. Fath and R. A. Darajatun, "Tinjauan perancangan produksi dan kualitas pada produk rak dies di CV Sarana Sejahtera Teknik," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 2, pp. 159–168, 2022, doi: 10.5281/zenodo.6105126.
- [7] Y. D. K. Setyo Wira Rizki, "Analisis Regresi Robust Estimasi-M Dengan Menggunakan Pembobotan Bisquare Tukey Dan Welsch Dalam Mengatasi Data Outlier," *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 8, no. 4, pp. 799–804, 2019, doi: 10.26418/bbimst.v8i4.36199.
- [8] H. D. E. Sinaga, R. Piliang, and D. Anggraeni, "Belajar Membuat Grafik Linier Sederhana Bagi Siswa SMA Panti Budaya Kisaran," *J. IPTEK Bagi Masy.*, vol. 1, no. 3, pp. 139–144, 2022, doi: 10.55537/jibm.v1i3.163.
- [9] D. U. Wutsqa, K. Kismiantini, R. Kusumawati, E. P. Setiawan, and B. Isnaini, "Workshop Visualisasi Data untuk Penelitian Ilmu Sosial dan Pendidikan," *J. Pengabd. Masy. MIPA dan Pendidik. MIPA*, vol. 7, no. 1, pp. 16–25, 2023, doi: 10.21831/jpmmp.v7i1.54412.
- [10] M. Wahyuni, *Statistik Deskriptif Untuk Penelitian Olah Data Manual dan SPSS versi 25*, no. Mi. 2020.
- [11] Y. Akbar, Y. A. Azzahra, S. S. Hartinah, and Z. Arfadhillah, "Monitoring Data Nilai Gizi Balita menggunakan Tableau Public (Studi Kasus: Posyandu Sedap Malam)," *J. Pengabd. Nas. Indones.*, vol. 5, no. 2, pp. 322–331, 2024, doi: 10.35870/jpni.v5i2.674.
- [12] D. Arisandi, T. Sutrisno, and I. Kurniawan, "Klasifikasi Opini Masyarakat Di Twitter Tentang Kebocoran Data Yang Terjadi Di Indonesia Menggunakan Algoritma Svm," *J. Inform. Kaputama*, vol. 7, no. 1, pp. 84–90, 2023, doi: 10.59697/jik.v7i1.10.
- [13] R. Adolph, "VISUALISASI DATA E-PPGBM MENGGUNAKAN GOOGLE DATA STUDIO," vol. 31, pp. 1–23, 2016.
- [14] F. Raudotul Janah, R. Kurniawan, and T. Suprpti, "Analisis Dataset Status Gizi Pada Balita Menggunakan Algoritma K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3602–3609, 2024, doi: 10.36040/jati.v7i6.8220.
- [15] A. R. Permata, "Analisis Data Penelitian Kesehatan: Perbandingan Hasil antara SmartPLS, R dan IBM SPSS Health Research Data Analysis: Comparison of Results between SmartPLS, R and IBM SPSS," *J. Sains Nat.*, vol. 1, no. 1, pp. 17–22, 2023.
- [16] R. Gustiputri, Kismartini, and R. Hanani, "IMPLEMENTASI KEBIJAKAN PERCEPATAN PENURUNAN DAN PENCEGAHAN STUNTING TERINTEGRASI KELURAHAN MARGOMULYO KABUPATEN NGAWI," *Sustain.*, vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI
- [17] R. Maulid, "Teknik Pengolahan Data Deskriptif dengan Statistik Deskriptif," *dqlab.id/*, 2021. <https://dqlab.id/teknik-pengolahan-data-deskriptif-dengan-statistik-deskriptif>
- [18] N. A. Syaripul and M. A. Bachtiar,

- “Visualisasi Data Interaktif Data Terbuka PemProv DKI Jakarta: Topik Ekonomi dan Keuangan Daerah,” *Sist. Inf.*, vol. 12, no. 2, pp. 82–89, 2016.
- [19] D. Aryanti and J. Setiawan, “Visualisasi Data Penjualan dan Produksi PT Nitto Alam Indonesia Periode 2014-2018,” *Ultim. InfoSys*, vol. 9, no. 2, pp. 86–91, 2019, doi: 10.31937/si.v9i2.991.
- [20] L. M. Nasution, “STATISTIK DESKRIPTIF,” *Hikma*, vol. 14, no. 21, pp. 5472–5476, 2017, doi: 10.1021/ja01626a006.
- [21] M. Taqwa and K. Lanani, “Pelatihan Analisis Data Non Parametrik Dengan Aplikasi RStudio Bagi Mahasiswa Prodi Bimbingan Konseling,” pp. 58–62.