

Strategi Peningkatan Akses Sanitasi dan Air Minum Aman Indonesia untuk Pencapaian SDGs 6

Penelitian dengan pendekatan *Convolutional Neural Network (CNN)* untuk *forecasting*, *Gaussian Mixture Model (GMM)* untuk *clustering*, dan *Geographically Weighted Panel Regression (GWPR)*

1st Erlin Shofiana
Department of Mathematics
Universitas Gadjah Mada
Yogyakarta, Indonesia
erlinshofiana0804@mail.ugm.ac.id

2nd Aulia Mirfah Setyo Ayu Damayanti
Department of Mathematics
Universitas Gadjah Mada
Yogyakarta, Indonesia
auliamirfahsetyoayudamayanti@mail.ugm.ac.id

3rd Fitri Hartanti
Department of Mathematics
Universitas Gadjah Mada
Yogyakarta, Indonesia
fitrihartanti@mail.ugm.ac.id

Abstract—Sanitasi yang layak dan akses terhadap air minum aman merupakan kebutuhan mendasar yang berperan krusial dalam kesehatan masyarakat dan pertumbuhan ekonomi. Di Indonesia, tantangan besar dalam sektor sanitasi masih terlihat, terutama jika dibandingkan dengan negara-negara tetangga seperti Thailand, Malaysia, dan Singapura yang telah mencapai 100% akses sanitasi. Menurut data Joint Monitoring Program (JMP) WHO-Unicef 2017, Indonesia menempati posisi kedua terendah di antara negara-negara G20 dan ASEAN dalam capaian sanitasi. Kondisi ini tak hanya memengaruhi kesehatan dengan meningkatnya prevalensi penyakit seperti diare dan tingginya angka stunting—di mana 1 dari 3 anak Indonesia mengalami stunting—tetapi juga memberikan dampak ekonomi yang signifikan, dengan kerugian negara mencapai Rp56,7 triliun per tahun akibat buruknya sanitasi. Penelitian ini melakukan pemodelan prediksi menggunakan metode *Convolutional Neural Network (CNN)* untuk memproyeksikan akses air minum aman di Indonesia dari tahun 2000 hingga 2022. Hasilnya menunjukkan bahwa target 45% akses air minum aman pada tahun 2030 masih jauh dari realisasi. Selanjutnya, metode *clustering Gaussian Mixture Model (GMM)* mengidentifikasi tiga kluster utama terkait kondisi air di Indonesia, dengan perbedaan akses air minum layak menjadi faktor signifikan dalam pembentukan kluster tersebut. Analisis lanjutan dengan *Geographically Weighted Panel Regression (GWPR)* mengungkapkan faktor-faktor regional yang mempengaruhi akses terhadap air minum layak di berbagai wilayah Indonesia. Temuan penelitian ini juga menyoroti kontribusi besar sanitasi yang baik dalam menurunkan angka stunting hingga 27% dan mengurangi kejadian diare sebesar 94% melalui implementasi Sanitasi Total Berbasis Masyarakat (STBM). Oleh karena itu, intervensi strategis yang lebih terfokus dan berbasis wilayah sangat diperlukan untuk mempercepat peningkatan akses sanitasi dan air minum aman di Indonesia, demi mencapai target nasional dan meningkatkan kualitas kesehatan serta ekonomi masyarakat.

Keywords—Sanitasi, Air minum, Forecasting, Clustering, Geographically Weighted Panel Regression

I. PENDAHULUAN

A. Latar Belakang

Sanitasi adalah usaha untuk menjaga kebersihan lingkungan dan kesehatan masyarakat dengan mengelola faktor-faktor lingkungan fisik yang berpotensi menimbulkan penyakit. *Water, Sanitation, and Hygiene (WASH)* merupakan program global yang berfokus pada penyediaan akses air bersih, sanitasi layak, dan praktik kebersihan yang benar guna mencegah penyebaran penyakit menular [1].

Indikator utama dalam WASH mencakup ketersediaan air bersih, fasilitas sanitasi yang aman, dan praktik mencuci tangan dengan sabun, yang berperan penting dalam menjaga kesehatan masyarakat [2].

Sejalan dengan *Sustainable Development Goals (SDGs)* poin ke-6, sanitasi yang aman dan terkelola dengan baik sangat penting bagi kesejahteraan masyarakat dan lingkungan. Sanitasi yang buruk menyebabkan pencemaran sumber air bersih dan menjadi faktor utama penyebaran penyakit yang berkaitan dengan air seperti diare, kolera, dan tifus. Di Indonesia, sekitar 25 juta orang masih buang air besar sembarangan, yang tidak hanya menurunkan martabat manusia, tetapi juga meningkatkan risiko kesehatan, terutama pada anak-anak. Penyakit diare, yang merupakan salah satu penyebab utama kematian anak di Indonesia, berhubungan erat dengan kualitas sanitasi yang buruk.

Selain menimbulkan dampak kesehatan, buruknya sanitasi juga memberikan efek negatif secara sosial dan ekonomi. Air yang tercemar limbah manusia dapat memicu siklus kemiskinan, di mana masyarakat miskin lebih rentan terhadap penyakit, kekurangan akses pendidikan, dan menurunnya produktivitas akibat kesehatan yang terganggu. Sebuah survei di Yogyakarta menunjukkan bahwa 89 persen sumber air dan 67 persen air minum rumah tangga terkontaminasi bakteri tinja, yang menggambarkan betapa seriusnya masalah sanitasi di Indonesia, bahkan di kota yang tergolong makmur [3].

Oleh karena itu, mencapai target SDG 6 membutuhkan strategi inklusif dan berkelanjutan untuk menyediakan akses sanitasi aman, terutama bagi kelompok rentan. Penelitian ini akan mengevaluasi posisi Indonesia dalam hal WASH (*Water, Sanitation, and Hygiene*) di Asia Tenggara dan memperkirakan waktu yang diperlukan untuk mencapai sanitasi layak. Kesenjangan akses terhadap sanitasi yang baik masih besar, terutama di kalangan masyarakat miskin. Penelitian ini juga akan menganalisis faktor-faktor yang memengaruhi buruknya sanitasi, guna merumuskan langkah strategis untuk menciptakan lingkungan yang sehat bagi semua lapisan masyarakat.

B. Rumusan Masalah

1. Bagaimana kondisi akses sanitasi dan air minum aman di Indonesia dibandingkan dengan negara-negara Asia Tenggara lainnya?

2. Bagaimana tren dan proyeksi akses air minum aman di Indonesia?
3. Bagaimana perbedaan kualitas sanitasi di berbagai wilayah di Indonesia?
4. Faktor apa saja yang mempengaruhi proporsi air minum layak di Indonesia?
5. Bagaimana peran pemerintah untuk memperbaiki permasalahan air bersih dan sanitasi?

C. Tujuan dan Manfaat

Tujuan dari penelitian ini adalah sebagai berikut :

1. Membandingkan akses sanitasi dan air minum aman di Indonesia dengan negara Asia Tenggara
2. Menganalisis perkembangan serta memperkirakan tren masa depan terkait akses air minum aman di Indonesia
3. Mengidentifikasi pola distribusi dan ketimpangan kualitas sanitasi di Indonesia
4. Mengidentifikasi faktor yang memengaruhi proporsi air minum layak di Indonesia
5. Mengevaluasi peran pemerintah dalam memperbaiki permasalahan air bersih dan sanitasi

Adapun manfaat dari penelitian ini mencakup beberapa aspek:

1. Akademik, penelitian ini memberikan kontribusi penting bagi literatur akademik terkait akses sanitasi dan air minum aman, khususnya terkait metodologi *forecasting*, *clustering*, dan *geographically weighted panel regression*
2. Praktis, pemerintah dapat memanfaatkan hasil penelitian ini untuk merumuskan kebijakan yang lebih efektif dan tepat sasaran untuk memperluas akses air minum aman dan sanitasi di seluruh wilayah Indonesia

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan baik dalam bidang akademik maupun praktis, untuk mendapatkan pemahaman yang lebih mendalam tentang tantangan dan peluang dalam menyediakan air minum aman dan sanitasi yang layak bagi seluruh penduduk Indonesia.

II. PEMBAHASAN

A. Landasan Teori

1. WASH

WASH (*Water, Sanitation, and Hygiene*) adalah konsep yang menekankan pentingnya pengelolaan air, sanitasi, dan kebersihan untuk kesehatan serta kesejahteraan masyarakat. Keberhasilan pengelolaan air dan sanitasi yang aman hanya dapat dicapai jika disertai dengan praktik kebersihan yang baik. Tanpa akses yang memadai ke WASH, kesejahteraan, martabat, dan peluang hidup masyarakat, terutama perempuan dan anak perempuan, terancam. Selain itu, akses ke air bersih dan sanitasi merupakan hak asasi manusia yang harus dipenuhi oleh pemerintah [4].

Ketiadaan akses terhadap WASH berdampak besar, terutama di negara-negara berpenghasilan rendah, di mana jutaan orang tidak memiliki akses ke air bersih, sanitasi yang aman, atau fasilitas mencuci tangan yang memadai. Hal ini menyebabkan tingginya angka penyakit menular, seperti diare dan kolera, yang berkontribusi pada tingginya angka

kematian bayi dan malnutrisi. Ketidaksetaraan dalam akses WASH juga memperparah kerentanan bagi perempuan, anak perempuan, orang tua, dan penyandang disabilitas, yang sering kali tidak dapat berpartisipasi sepenuhnya di ruang publik atau pendidikan.

2. SDGs

Sustainable Development Goals (SDGs) adalah kerangka kerja global yang bertujuan untuk mencapai masa depan yang lebih baik dan berkelanjutan bagi semua. Terdiri dari 17 tujuan yang saling terhubung, SDGs dirancang untuk mengatasi tantangan global, termasuk kemiskinan, ketidaksetaraan, perubahan iklim, dan degradasi lingkungan. SDGs poin ke-6 secara khusus berfokus pada penyediaan akses terhadap air bersih, sanitasi, dan kebersihan yang aman dan terjangkau bagi semua orang, mengingat bahwa akses ini merupakan hak asasi manusia yang fundamental untuk kesehatan dan kesejahteraan masyarakat [5].

Pada tahun 2030, diperkirakan miliaran orang masih akan kekurangan akses terhadap layanan dasar ini kecuali kemajuan yang signifikan terjadi. Permintaan akan air meningkat seiring pertumbuhan populasi dan urbanisasi, dan banyak orang mengalami kekurangan air yang parah. Dalam menghadapi tantangan ini, penting untuk melakukan investasi dalam infrastruktur, melindungi ekosistem terkait air, serta meningkatkan pendidikan tentang kebersihan. Meskipun ada kemajuan, masih ada 2,2 miliar orang yang tidak memiliki akses terhadap air minum yang aman dan 3,5 miliar orang yang tidak memiliki sanitasi yang dikelola dengan baik [6].

3. Forecasting

Convolutional Neural Network (CNN) pada awalnya dikenal untuk aplikasi *Computer Vision* (CV) dan *Natural Language Processing* (NLP), tetapi lebih baru, CNN juga diterapkan dalam *Time Series Forecasting*, baik untuk data multivariat maupun univariat. CNN memberikan hasil yang sering kali lebih baik daripada model autoregresif tradisional seperti *Vector Autoregression* (VAR) [7]. *Forecasting* pada *time series* bertujuan untuk memprediksi nilai masa depan dari satu atau lebih variabel berdasarkan observasi di masa lalu.

Persamaan umumnya adalah:

$$E[X_{t+d}|X_{t-1}, X_{t-2}, \dots, X_0] = f(X_{t-1}, X_{t-2}, \dots, X_0), \quad d = 0, 1, 2, \dots \quad (1)$$

Di mana $X_i = (x_{1i}, x_{2i}, \dots, x_{ni})$ adalah vektor observasi dari seluruh variabel yang dipertimbangkan pada waktu i .

Untuk *forecasting* pada data multivariat, CNN bekerja dengan menggunakan *convolutional layer* yang beroperasi pada data *time series* multi-dimensi. Sebagai contoh, jika dimiliki beberapa variabel (misalnya data cuaca atau ekonomi), model CNN dapat menangkap pola-pola kompleks dari hubungan antar variabel. Dalam arsitektur CNN, urutan lapisan yang sering digunakan adalah:

Input -> Convolution -> Leaky ReLU -> Pooling -> Convolution -> Leaky ReLU -> Dense

Pada bagian *convolutional layer*, kernel digunakan untuk memindai input data *time series* dalam urutan tertentu. Setiap kernel menangkap pola lokal dalam data, misalnya korelasi antara nilai waktu t dengan nilai-nilai di masa lalu. Setelah itu, dilakukan batch normalization untuk menstabilkan distribusi aktivasi, dan Leaky ReLU digunakan sebagai fungsi aktivasi untuk mencegah masalah vanishing gradient. Lapisan pooling kemudian mereduksi dimensi data dengan menjaga fitur-fitur penting dari convolutional layer. Hasil dari convolutional dan pooling layer diratakan menggunakan flatten layer, yang kemudian diumpukan ke dense layer untuk menghasilkan prediksi nilai time series masa depan.

CNN pada Data Univariat Pada data univariat (hanya satu variabel), struktur CNN lebih sederhana namun tetap efektif dalam menangkap pola temporal yang signifikan dari data satu dimensi.

Input -> Conv1D -> ReLU -> Max Pooling -> Flatten -> Dense

Lapisan *convolutional* menggunakan kernel kecil untuk mendeteksi pola lokal pada data urutan satu dimensi (*univariate time series*), seperti tren dan pola musiman. Setelah tahap *convolution*, lapisan *pooling* (misalnya *max pooling*) berfungsi untuk menyaring dan mereduksi dimensi data, hanya mempertahankan elemen-elemen paling relevan. Selanjutnya, *flatten layer* meratakan output dari pooling layer menjadi vektor satu dimensi agar siap untuk diproses oleh *dense layer* [8]. Lapisan dense ini kemudian menafsirkan fitur yang diekstraksi dan melakukan prediksi terhadap nilai time series di masa depan.

4. *Clustering*

Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. Dalam data mining terdapat dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* [9]

Pada penelitian ini teknik *clustering* yang digunakan adalah *Gaussian Mixture Models* dengan metrik evaluasi *silhouette score*. *Gaussian Mixture Model* (GMM) merupakan teknik *clustering* yang membentuk *cluster* berdasarkan *probability density function*. Setiap *cluster* memiliki tiga parameter, yaitu bobot, mean, dan variansi [10]. Beberapa keunggulan *Gaussian Mixture Model* diantaranya :

- Menggunakan pendekatan probabilistik yang memberikan fleksibilitas dalam mengidentifikasi batas-batas kluster yang tidak kaku,
- Mampu memisahkan kluster dengan ukuran, orientasi, dan distribusi yang berbeda, sehingga lebih efektif dalam menangani kluster yang tumpang tindih atau kompleks.

5. *Geographically Weighted Panel Regression*

Metode GWPR adalah metode yang menggabungkan antara model *geographically weighted regression* (GWR) dan regresi data panel. Pada model GWR menggunakan data spasial yaitu pendekatan titik koordinat garis lintang (*latitude*) dan garis bujur (*longitude*), sehingga memungkinkan parameter model bervariasi di setiap lokasi [11]. Melalui model GWPR, faktor-faktor yang memengaruhi variabel respon dapat dianalisis secara lokal untuk setiap lokasi, memungkinkan identifikasi pengaruh yang berbeda di tiap wilayah.

Model umum regresi data panel adalah sebagai berikut

$$y_{it} = \beta_{it} + \beta'_{it}x_{it} + \varepsilon_{it} \quad (2)$$

dengan:

y_{it} = variabel respon ke- i pada waktu ke- t

β_{it} = parameter dari lokasi ke- i pada waktu ke- t yang merupakan skalar

β'_{it} = vektor konstanta dengan ukuran $1 \times k$ dari lokasi ke- i dan waktu ke- t

x_{it} = vektor prediktor berukuran $1 \times p$

ε_{it} = error pada lokasi ke- i dan waktu ke- t

B. *Metode Penelitian*

1. *Software*

Software yang digunakan dalam analisis data adalah Python dan R. Python digunakan untuk analisis *forecasting* dan *clustering*, sementara R digunakan untuk analisis GWPR (*Geographically Weighted Panel Regression*). Selain itu, untuk visualisasi data, Power BI digunakan untuk menampilkan posisi Indonesia dibandingkan dengan negara lain di Asia Tenggara pada indikator WASH, dan Tableau digunakan untuk visualisasi *clustering* berdasarkan WASH.

2. *Data Overview*

Terdapat empat dataset yang digunakan dalam analisis ini, dengan masing-masing variabel dan sumber data adalah sebagai berikut.

Tabel 1. *Dataset*

Analisis	Variabel	Sumber
Visualisasi	Presentase air minum aman, presentase fasilitas cuci tangan, dan presentase sanitasi aman untuk negara-negara di Asia Tenggara.	<i>United Nations</i> (UN)
Forecasting: Convolutional Neural Network (CNN)	<i>Proportion of population safely managed drinking water services</i> untuk Indonesia.	<i>United Nations</i> (UN)
Clustering: Gaussian Mixture	34 provinsi di Indonesia, persentase rumah tangga yang	Badan Pusat Statistika (BPS)

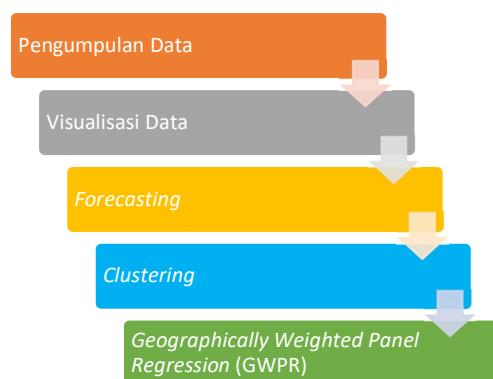
Model (GMM)	memiliki akses terhadap sumber air minum layak, presentase rumah tangga yang memiliki akses sanitasi layak, dan proporsi rumah tangga yang memiliki fasilitas cuci tangan dengan sabun dan air menurut daerah tempat tinggal.
Geography ally Weighted Panel Regression (GWPR)	34 provinsi, tahun, jumlah penduduk, air bersih disalurkan (ribu m ³), PDRB per kapita, persentase sanitasi layak, realisasi PMDN, realisasi PMA, dan persentase rumah tangga yang memiliki akses terhadap sumber air minum layak.
	Badan Pusat Statistika (BPS)

3. Metode Analisis Data

Metode analisis data yang digunakan penulis dalam analisis ini adalah *time series forecasting* metode *Convolutional Neural Network* (CNN), *clustering* metode *Gaussian Mixture Model* (GMM), dan *Gaussian Mixture Model* (GMM).

4. Alur Analisis

Penelitian ini dilaksanakan mengikuti langkah-langkah yang telah direncanakan, penelitian ini bertujuan untuk memastikan bahwa setiap tahapan berjalan sesuai dengan tujuan awal dan meminimalkan risiko penyimpangan dari standar yang telah ditetapkan. Dengan demikian, penelitian dapat memberikan hasil yang lebih akurat dan relevan. Diagram alur yang menggambarkan proses penelitian ini dapat dilihat sebagai berikut:



Gambar 1. Alur Analisis

Pengumpulan Data

Pengumpulan data dalam penelitian dilakukan beberapa kali untuk menyesuaikan dengan analisis yang akan dilakukan. Visualisasi menggunakan data dari United Nations (UN) terkait persentase air minum aman, fasilitas cuci tangan, dan sanitasi aman di kawasan Asia Tenggara. Selain itu, *forecasting* dengan *Convolutional Neural Network* (CNN) diterapkan untuk memprediksi proporsi populasi Indonesia yang menggunakan layanan air minum aman di masa depan, juga berdasarkan data dari UN [12].

Untuk *clustering* dengan *Gaussian Mixture Model* (GMM) digunakan untuk mengelompokkan 34 provinsi yang ada di Indonesia berdasarkan akses rumah tangga terhadap air minum layak [13], sanitasi layak [14], dan fasilitas cuci tangan [15], menggunakan data dari Badan Pusat Statistik (BPS). Terakhir, metode *Geographically Weighted Panel Regression* (GWPR) menganalisis faktor-faktor sosial, ekonomi, dan lingkungan di provinsi-provinsi tersebut, dengan mempertimbangkan variabel seperti jumlah penduduk [16], PDRB per kapita [17], realisasi PMDN [18], realisasi PMA [19] dan air bersih disalurkan [20], guna memahami hubungan antara faktor-faktor ini terhadap akses air bersih dan sanitasi.

Visualisasi Data

Visualisasi, pada awal analisis, dilakukan tahapan visualisasi data untuk memberikan gambaran awal mengenai distribusi dan pola dari variabel-variabel yang dianalisis, seperti persentase air minum aman, fasilitas cuci tangan, dan sanitasi aman, guna melihat perbandingan antar wilayah serta mengidentifikasi potensi kesenjangan atau anomali dalam akses terhadap layanan dasar ini.

Forecasting

Melihat hasil visualisasi indikator sanitasi di negara-negara Asia Tenggara, analisis ini akan berfokus pada presentase air minum aman di Indonesia. Dilakukan analisis *forecasting* terhadap persentase air minum layak di Indonesia dari tahun 2000 hingga 2022, menggunakan *Convolutional Neural Network* (CNN) dan proyeksi hingga tahun 2030. Tujuan dari analisis ini adalah untuk mengidentifikasi tren peningkatan akses air minum dan menentukan apakah Indonesia dapat mencapai target Sustainable Development Goals (SDGs) untuk air minum aman sebesar 45% pada tahun 2030.

Proses forecasting dimulai dengan persiapan data. Dengan menggunakan fungsi *split_sequence* untuk membagi data deret waktu menjadi input dan output berdasarkan jumlah langkah (*n_steps*) yang ditentukan, di mana setiap pola input terdiri dari *n_steps* nilai, sedangkan outputnya adalah nilai berikutnya. Data kemudian diubah bentuknya agar sesuai dengan format yang dibutuhkan oleh CNN, dan dinormalisasi menggunakan *MinMaxScaler* untuk mengurangi variasi, sehingga model dapat belajar dengan lebih efektif.

Selanjutnya, analisis dilakukan dengan mengimpor *library* yang diperlukan dan

mendefinisikan model CNN. Model ini dibangun menggunakan Keras, dengan arsitektur yang melibatkan lapisan Convolutional1D, *pooling*, *flatten*, dan *dense*. Data dibagi menjadi beberapa lipatan (*folds*) menggunakan *k-fold cross-validation* untuk evaluasi yang lebih komprehensif. Pada setiap *fold*, model dilatih dengan data pelatihan dan diuji dengan data pengujian, menggunakan fungsi kerugian *Mean Squared Error (MSE)* sebagai metrik evaluasi. Setelah pelatihan selesai, dihitung kerugian pada data pengujian untuk analisis lebih lanjut.

Untuk memprediksi masa depan, diambil input terakhir dari data yang telah dinormalisasi, dan model digunakan untuk meramalkan nilai berikutnya. Proses ini diulang untuk menghasilkan delapan periode prediksi ke depan, dan hasil prediksi dinormalisasi kembali ke skala aslinya agar dapat dibandingkan dengan data historis.

Akhirnya, dilakukan evaluasi dengan menghitung rata-rata kerugian dari semua *fold* yang telah dievaluasi. Rata-rata kerugian ini dihitung dengan rumus:

$$\text{Mean Loss} = \frac{1}{n} \sum_{i=1}^n \text{loss}_i \quad (3)$$

di mana n adalah jumlah *fold* dan loss_i adalah nilai kerugian pada *fold* ke- i . Hasil rata-rata kerugian ini memberikan gambaran menyeluruh tentang kinerja model dalam melakukan prediksi, sekaligus menunjukkan potensi Indonesia untuk mencapai target air minum aman di masa depan.

Clustering

Berdasarkan analisis *forecasting*, diperoleh proyeksi persentase air minum layak di Indonesia. Untuk mengidentifikasi kelompok wilayah yang memiliki akses paling buruk terhadap air bersih maupun sanitasi, dilakukan analisis *clustering*. Analisis ini dapat membantu menentukan wilayah mana yang membutuhkan perhatian khusus, serta menyusun strategi yang lebih terfokus untuk mempercepat peningkatan akses air bersih dan sanitasi di wilayah-wilayah tersebut.

Sebelum melakukan analisis *clustering*, data proporsi rumah tangga dengan akses air minum layak, sanitasi layak, dan fasilitas cuci tangan dengan sabun distandarisasi terlebih dahulu untuk menghilangkan perbedaan skala antar variabel, sehingga setiap variabel memiliki kontribusi yang setara dalam proses *clustering*. Selanjutnya, untuk menentukan jumlah *cluster* yang optimum, digunakan metrik *Silhouette Score*, di mana semakin tinggi nilai *Silhouette Score* menunjukkan semakin baik pemisahan antar *cluster* dan semakin homogen objek-objek dalam setiap *cluster*, yang menandakan *clustering* lebih optimal. Hasil analisis menunjukkan bahwa jumlah *cluster* yang optimum adalah tiga *cluster*.

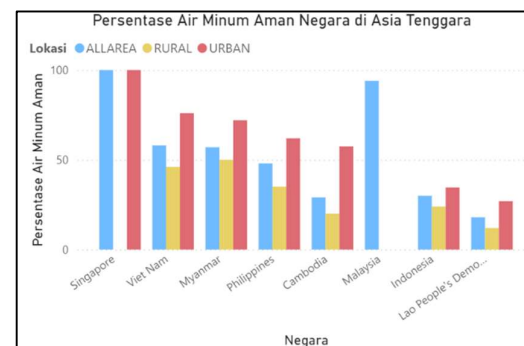
Hasil *clustering* tersebut kemudian dianalisis untuk mengidentifikasi karakteristik setiap *cluster* guna menentukan variabel-variabel yang menunjukkan ketimpangan antar *cluster*. Analisis ini bertujuan untuk mengungkap perbedaan signifikan dalam akses air minum layak, sanitasi layak, dan fasilitas cuci tangan dengan sabun di setiap *cluster*.

Geographically Weighted Panel Regression (GWPR)

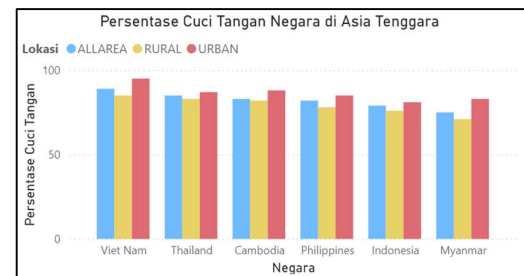
Dalam rangka memahami lebih mendalam faktor-faktor yang memengaruhi proporsi air minum layak di setiap provinsi, dilakukan analisis *Geographically Weighted Panel Regression (GWPR)*. Analisis GWPR diawali dengan pengumpulan data *cross section*, yaitu data setiap provinsi untuk setiap variabel, dan data *time series*, yakni data dari tahun 2017 hingga 2020. Data yang terkumpul kemudian diuji untuk menentukan model panel yang paling sesuai, serta untuk memastikan adanya heterogenitas spasial. Selanjutnya, analisis dilakukan dengan menentukan bobot optimum berdasarkan metrik evaluasi, yaitu R^2 tertinggi dan *residual sum of squares (RSS)* terendah. Setelah bobot optimum diperoleh, dilakukan uji signifikansi terhadap parameter-parameter dalam model untuk memastikan kontribusi setiap faktor secara statistik.

C. Hasil dan Pembahasan

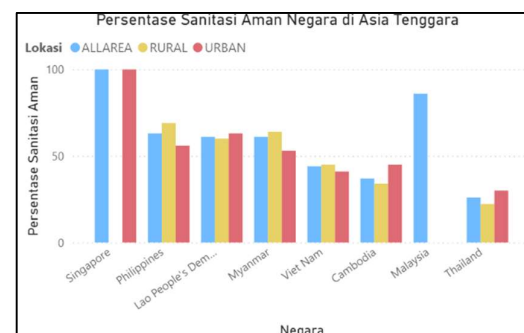
1. Visualisasi Data



Gambar 2. Presentase Air Minum Aman Negara di Asia Tenggara



Gambar 3. Presentase Cuci Tangan Aman Negara di Asia Tenggara



Gambar 4. Presentase Sanitasi Aman Negara di Asia Tenggara

Visualisasi data di atas menunjukkan tiga indikator penting terkait dengan akses air minum aman, sanitasi aman, dan kebiasaan mencuci tangan di negara-negara Asia Tenggara. Masing-masing indikator ini memiliki nilai berdasarkan area masing-masing, yaitu keseluruhan wilayah (ALLAREA), daerah pedesaan (RURAL), dan perkotaan (URBAN). Berikut adalah penjelasan secara lebih rinci.

- **Persentase Air Minum Aman**

Gambar 2 menunjukkan perbandingan akses air minum yang aman di berbagai negara Asia Tenggara. Singapura memiliki akses air minum aman yang sangat tinggi di semua. Negara-negara seperti Vietnam dan Filipina menunjukkan perbedaan antara akses air minum di daerah perkotaan dan pedesaan. Di sisi lain, Indonesia berada di posisi terendah baik di perkotaan maupun pedesaan, menunjukkan bahwa akses air minum aman di Indonesia sangat memprihatinkan.

- **Persentase Cuci Tangan**

Gambar 3 menunjukkan persentase kebiasaan cuci tangan di beberapa negara Asia Tenggara. Secara umum, kebiasaan mencuci tangan di negara-negara ini cukup merata, dengan Vietnam dan Thailand memiliki persentase yang sedikit lebih tinggi dibandingkan negara lain. Indonesia kembali berada di peringkat bawah dalam hal ini, mengindikasikan bahwa kampanye kebersihan dasar, termasuk cuci tangan, perlu lebih ditekankan.

- **Persentase Sanitasi Aman**

Gambar 4 menunjukkan persentase akses terhadap sanitasi yang aman di negara-negara Asia Tenggara. Singapura berada di peringkat teratas dengan sanitasi aman yang hampir universal di semua area (ALLAREA, RURAL, dan URBAN). Filipina dan Malaysia juga menunjukkan hasil yang cukup baik, meskipun ada perbedaan antara area perkotaan dan pedesaan. Negara-negara seperti Vietnam, Kamboja, dan Myanmar memiliki tingkat sanitasi aman yang lebih rendah, terutama di daerah pedesaan. Indonesia tidak termasuk dalam visualisasi ini, namun negara-negara lain yang ditampilkan menunjukkan adanya kesenjangan antara wilayah perkotaan dan pedesaan dalam akses terhadap sanitasi aman, yang bisa menjadi perhatian bersama di kawasan ini.

Secara keseluruhan, dari visualisasi ini dapat diketahui bahwa Indonesia berada di peringkat terendah untuk akses air minum aman dan kebiasaan cuci tangan. Ini menunjukkan bahwa implementasi program WASH (*Water, Sanitation, and Hygiene*) sangat penting untuk memperbaiki kondisi kesehatan masyarakat di Indonesia. Program ini juga berkaitan langsung dengan tujuan SDGs poin

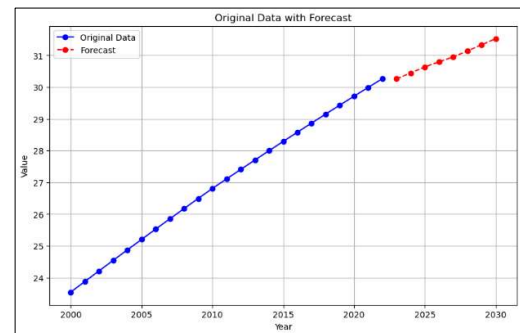
ke-6, yaitu memastikan ketersediaan dan pengelolaan air bersih serta sanitasi yang berkelanjutan untuk semua. Dengan kondisi saat ini, Indonesia harus lebih serius mengatasi tantangan dalam akses air bersih dan sanitasi untuk mencapai target ini pada tahun 2030.

2. Forecasting

Berdasarkan hasil *forecasting* menggunakan *Convolutional Neural Network* (CNN) dengan data dari *United Nations* (UN), *Indicator 6.1.1: Proportion of population using safely managed drinking water services* untuk Indonesia dari tahun 2000 hingga 2022, prediksi untuk persentase populasi yang menggunakan air minum aman hingga tahun 2030 menunjukkan tren peningkatan yang stabil. Berikut adalah prediksi dari tahun 2023 hingga 2030:

Tabel 2. Hasil *Forecasting*

Tahun	Forecast	Tahun	Forecast
2023	30,26%	2027	30,95%
2024	30,45%	2028	31,14%
2025	30,63%	2029	31,33%
2026	30,79%	2030	31,53%



Gambar 5. Plot Data Asli dan Hasil *Forecasting*

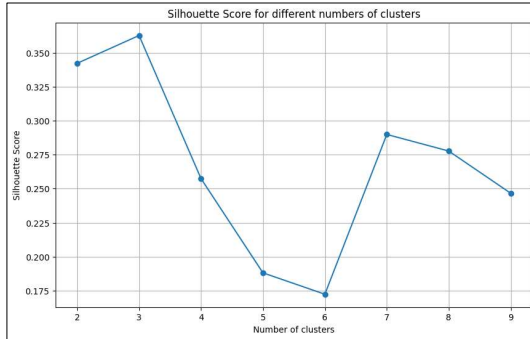
Prediksi ini menunjukkan bahwa meskipun ada peningkatan bertahap dalam akses air minum aman, laju kenaikan ini tidak cukup signifikan untuk mencapai target *Sustainable Development Goals* (SDGs) dari Rencana Pembangunan Jangka Menengah (RPJM) sebesar 45% pada tahun 2030. Dengan proyeksi terakhir sebesar 31,53% pada tahun 2030, Indonesia masih akan berada di bawah target yang ditetapkan.

Adapun *mean loss* dari model CNN sebesar 0,00001897, yang menunjukkan bahwa kesalahan rata-rata antara prediksi model dan data aktual sangat kecil. Ini mengindikasikan bahwa model memiliki tingkat akurasi yang cukup baik dalam memprediksi tren akses air minum aman, namun laju peningkatan akses air yang diproyeksikan masih jauh dari target yang diharapkan untuk mencapai SDGs.

Oleh karena itu, diperlukan intervensi lebih lanjut dan upaya yang lebih intensif dari pemerintah serta pihak terkait untuk mempercepat peningkatan akses air minum aman agar target SDGs dapat tercapai pada tahun 2030.

3. Clustering

Dalam proses *clustering*, jumlah *cluster* optimum ditentukan menggunakan nilai *Silhouette Score*. Nilai *silhouette score* yang lebih tinggi menunjukkan pemisahan antar *cluster* yang lebih baik dan homogenitas yang lebih kuat di dalam setiap *cluster*.



Gambar 6. Grafik *Silhouette Score*

Gambar 6 adalah grafik *silhouette score* yang menunjukkan bahwa nilai tertinggi tercapai pada 3 *cluster*, sehingga jumlah *cluster* optimum ditetapkan sebanyak 3.



Gambar 7. Visualisasi Hasil Clustering

Tabel 3. Hasil *Clustering*

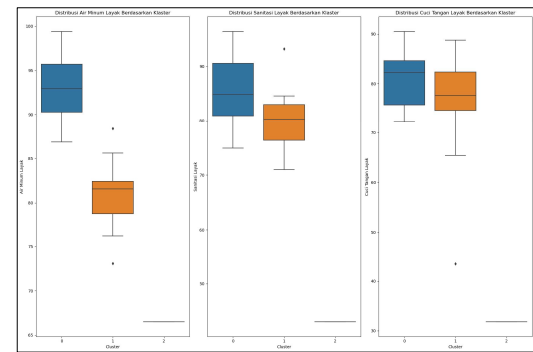
Cluster	Anggota
0	Aceh, Sumatera Utara, Riau, Sumatera Selatan, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara
1	Sumatera Barat, Jambi, Bengkulu, Lampung, Kepulauan Bangka Belitung, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Barat, Papua Barat
2	Papua

Dari hasil *clustering* menggunakan metode *gaussian mixture models* dengan jumlah 3 *cluster*,

didapatkan hasil seperti yang ditampilkan pada Gambar 6 dan Tabel 3 Hasil ini menggambarkan distribusi data dalam tiga *cluster* berdasarkan probabilitas keanggotaan tiap data, di mana setiap *cluster* merepresentasikan kelompok wilayah dengan karakteristik akses air minum layak, sanitasi layak, dan fasilitas cuci tangan yang berbeda-beda. Hasil karakteristik rata-rata fitur berdasarkan *cluster* ditampilkan sebagai berikut:

Tabel 4. Karakteristik Setiap *Cluster*

Cluster	Air Minum Layak	Sanitasi Layak	Cuci Tangan Layak
0	92,861	85,479	80,560
1	80,816	80,352	75,798
2	66,490	43,000	31,780



Gambar 8. Boxplot Karakteristik Setiap *Cluster*

Berdasarkan hasil klasterisasi, ditemukan tiga kelompok utama yang dibedakan berdasarkan presentase akses rumah tangga terhadap air minum layak, sanitasi layak, dan fasilitas cuci tangan layak di berbagai wilayah Indonesia. Setiap *cluster* menunjukkan karakteristik yang berbeda secara signifikan, terutama dalam akses terhadap air minum aman.

Cluster 0 merupakan kelompok dengan akses tertinggi terhadap fasilitas dasar. Di wilayah yang termasuk dalam *cluster* ini, persentase rumah tangga dengan akses air minum layak mencapai 92,861%, sanitasi layak sebesar 85,479%, dan fasilitas cuci tangan layak sebesar 80,560%. *Cluster* ini menunjukkan bahwa wilayah-wilayah ini memiliki infrastruktur yang sangat baik, sehingga mayoritas penduduknya dapat menikmati layanan dasar yang berkualitas.

Di sisi lain, *cluster 1* memiliki tingkat akses yang relatif baik, namun masih di bawah klaster pertama. Wilayah dalam *cluster* ini memiliki persentase akses air minum layak sebesar 80,816%, sanitasi layak sebesar 80,352%, dan fasilitas cuci tangan layak sebesar 75,798%. Meskipun tergolong baik, wilayah-wilayah ini masih memerlukan upaya peningkatan terutama pada akses terhadap air minum, agar dapat mendekati kualitas *cluster 0*.

Yang paling mencolok adalah *cluster* 2, di mana terlihat kesenjangan yang sangat besar dalam akses terhadap fasilitas dasar. Persentase rumah tangga dengan akses air minum layak di *cluster* ini hanya mencapai 66,490%, sedangkan akses terhadap sanitasi layak berada pada 43,000%, dan fasilitas cuci tangan layak hanya sebesar 31,780%. Kesenjangan yang luar biasa ini menunjukkan bahwa *cluster* 2 berada jauh di belakang *cluster* lainnya, sehingga memerlukan perhatian dan intervensi yang serius dari pemerintah dan pihak terkait untuk memperbaiki kualitas hidup di wilayah tersebut.

Melihat bahwa gap terbesar di antara ketiga *cluster* ini terjadi pada akses terhadap air minum layak, dilakukan analisis *Geographically Weighted Panel Regression* (GWPR). Tujuan dari analisis ini adalah untuk memahami lebih dalam faktor-faktor yang memengaruhi persentase rumah tangga yang memiliki akses terhadap air minum layak di setiap provinsi.

4. *Geographically Weighted Panel Regression* (GWPR)

Analisis GWPR diawali dengan menentukan model panel yang sesuai dengan data. Maka dari itu, dilakukan uji chow, uji haussman, serta uji breusch pagan.

Tabel 4. Uji Chow dan Uji Haussman

Uji	P-value
Chow	5.916e-14
Haussman	< 2.2e-16

Berdasarkan Tabel 4, Pada uji chow didapatkan $p - value$ $5.916e-14 < \alpha = 0,05$. Dengan H_0 adalah model common effect lebih sesuai dibandingkan model fixed effect. Maka dari uji chow didapatkan model yang cocok adalah fixed effect. Pada uji Haussman didapatkan $p - value < 2,2e - 16 < \alpha = 0,05$. Dengan H_0 adalah bahwa model random effect lebih sesuai dibandingkan model fixed effect. Maka dari uji chow didapatkan mode yang cocok adalah fixed effect.

Selanjutnya dilakukan uji breusch pagan untuk melihat melihat efek individu dan waktu.

Tabel 5. Uji Breusch Pagan

Uji	P-value
Efek 2 arah	< 2.2e-16
Efek 1 arah (individu)	1.264e-11
Efek 1 arah (waktu)	2.343e-10

Berdasarkan Tabel 5 dan hipotesis awal (H_0) masing-masing secara berurutan adalah tidak terdapat efek dua arah; tidak terdapat efek individu; dan tidak terdapat efek waktu. Diperoleh kesimpulan bahwa model yang cocok adalah model fixed effect dengan pengaruh individu dan waktu.

Untuk memastikan terdapat heterogenitas spasial pada data, dilakukan uji Breusch-Pagan (bptest), dengan hipotesis nol (H_0) bahwa tidak terdapat heterogenitas spasial. Dari uji tersebut, didapatkan nilai $p-value$ sebesar 0,01587, sehingga dapat disimpulkan bahwa terdapat heterogenitas spasial pada data. Oleh karena itu, analisis *Geographically Weighted Panel Regression* (GWPR) dapat dilakukan. Selanjutnya, akan ditentukan fungsi pembobot terbaik untuk analisis ini.

Tabel 6. Fungsi Pembobot Kernel

Fungsi Pembobot	R ²	RSS
Fixed Bisquare	0.8457202	2819.5747
Fixed Gaussian	0.9743126	469.4566
Fixed Exponential	0.9585474	757.5767

Berdasarkan Tabel 6 didapatkan bahwa fungsi pembobot terbaik adalah *fixed gaussian*.

Hasil analisis GWPR dengan fungsi pembobot *fixed gaussian* didapatkan bahwa variabel yang signifikan memengaruhi persentase rumah tangga yang memiliki akses terhadap sumber air minum layak diantaranya adalah PDRB (produk domestik regional bruto) dan PMDN (penanaman modal dalam negeri).

5. Solusi

Berdasarkan hasil analisis visualisasi untuk indikator sanitasi di negara-negara Asia Tenggara, sangat jelas bahwa Indonesia menghadapi tantangan serius dalam penyediaan air minum aman, dengan persentase hanya mencapai sekitar 30%, jauh di bawah enam negara lainnya di kawasan ini. Hasil forecasting menggunakan CNN juga menunjukkan bahwa jika tidak ada upaya serius dari pemerintah untuk memperbaiki kondisi ini, target peningkatan air minum aman sebesar 45% yang telah ditetapkan dalam RPJM dan SDGs Indonesia pada 2030 tidak akan tercapai. Ditambah lagi, analisis kluster menegaskan bahwa terdapat kesenjangan yang sangat besar, terutama pada kluster 2, di mana akses rumah tangga terhadap air minum layak berada pada tingkat yang jauh lebih rendah dibandingkan kluster lainnya.

Dari hasil analisis *Geographically Weighted Panel Regression* (GWPR), ditemukan bahwa variabel yang paling signifikan memengaruhi persentase rumah tangga dengan akses air minum layak adalah Produk Domestik Regional Bruto (PDRB) dan Penanaman Modal Dalam Negeri (PMDN). Oleh karena itu, solusi untuk permasalahan ini harus berfokus pada peningkatan investasi di sektor air bersih, terutama dalam pembangunan infrastruktur penyediaan air minum yang layak.

Solusi pertama adalah peningkatan realisasi pendanaan untuk Sistem Penyediaan Air Minum (SPAM). Berdasarkan jurnal "Pendanaan dalam

Pencapaian Akses Universal Air Minum di Indonesia", hanya 33,56% dari target pendanaan sebesar 253,85 triliun rupiah yang telah terealisasi. Pendanaan yang berasal dari APBD bahkan hanya mencapai 9,76% dari target 119,3 triliun rupiah. Untuk mencapai target akses 100% air minum layak pada tahun 2028, pendanaan harus ditingkatkan minimal 2,5 kali lipat, dan harus segera diakselerasi dalam beberapa tahun mendatang.

Selain dari segi pendanaan, peningkatan kualitas dan manajemen akses air bersih juga menjadi solusi penting. Tingkat kebocoran air yang tinggi serta kinerja PDAM yang kurang sehat di beberapa daerah menghambat peningkatan akses air minum. Maka, diperlukan peningkatan dalam efisiensi pengelolaan air, perbaikan infrastruktur, serta pemeliharaan jaringan distribusi air.

Upaya berikutnya adalah pengembangan jaringan nonperpipaan, seperti sumur, wadah penampungan air hujan, atau bangunan penangkap mata air, terutama di wilayah perdesaan yang masih tertinggal. Pada tahun 2020, rumah tangga perdesaan yang memiliki akses air minum layak hanya mencapai 82,74%, jauh di bawah capaian di perkotaan. Oleh karena itu, pemerataan pembangunan infrastruktur air minum harus diprioritaskan di wilayah perdesaan.

Terakhir, diperlukan evaluasi menyeluruh terhadap program yang telah dijalankan. Pemerintah perlu menilai efektivitas program-program yang ada, dan melakukan penyesuaian agar lebih tepat sasaran. Ini termasuk peningkatan monitoring dan pengawasan terhadap realisasi pendanaan, serta memastikan bahwa program penyediaan air minum dan sanitasi dilaksanakan dengan standar yang tinggi, baik di kota maupun di desa.

III. PENUTUP

A. Kesimpulan

Berdasarkan hasil analisis, penelitian ini menunjukkan bahwa akses air minum layak di Indonesia diproyeksikan belum mencapai target Sustainable Development Goals (SDGs) dan Rencana Pembangunan Jangka Menengah Nasional (RPJMN) pada tahun 2030. Dari hasil clustering, ditemukan 3 kluster dengan disparitas akses air minum layak dan sanitasi antar wilayah. Provinsi-provinsi yang termasuk dalam kluster 1 dan 2, seperti Sumatera Barat, Jambi, Bengkulu, Lampung, Kepulauan Bangka Belitung, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Barat, Papua Barat, dan Papua, memerlukan perhatian khusus.

Sementara itu, hasil analisis GWPR menunjukkan bahwa variabel yang signifikan memengaruhi persentase rumah tangga dengan akses terhadap sumber air minum layak di antaranya adalah PDRB (Produk Domestik Regional Bruto) dan PMDN (Penanaman Modal Dalam Negeri). Untuk itu, diperlukan evaluasi menyeluruh terhadap program yang telah dijalankan, dengan penilaian efektivitas serta penyesuaian agar program lebih tepat sasaran dan berdampak maksimal.

B. Saran

Berdasarkan hasil penelitian ini, beberapa langkah perbaikan perlu dilakukan guna meningkatkan akses air minum layak dan sanitasi di Indonesia. Pertama, diperlukan peningkatan dalam realisasi pendanaan, terutama untuk Sistem Penyediaan Air Minum (SPAM), serta peningkatan efisiensi dan pengelolaan PDAM untuk mengurangi kebocoran air. Pemerintah juga perlu memprioritaskan pembangunan infrastruktur air di wilayah perdesaan yang tertinggal, mengingat perbedaan akses yang signifikan antara daerah perkotaan dan perdesaan. Selain itu, penting untuk mengembangkan jaringan nonperpipaan di daerah-daerah terpencil.

Dari sisi penelitian, saran untuk penelitian di masa mendatang adalah untuk memperluas eksplorasi sumber data. Salah satu tantangan utama dalam penelitian ini adalah sulitnya dalam mendapatkan data yang relevan dan akurat sesuai kebutuhan analisis. Oleh karena itu, penelitian di masa depan diharapkan dapat menggali lebih dalam dengan menggunakan berbagai sumber data tambahan, baik nasional maupun internasional, serta mempertimbangkan pemanfaatan teknologi dan pendekatan baru dalam pengumpulan data lapangan.

REFERENSI

- [1] UNICEF Indonesia. (n.d.). Air, sanitasi dan kebersihan (WASH): Mewujudkan lingkungan yang bersih untuk hidup, bermain, dan belajar bagi anak-anak. Diakses dari <https://www.unicef.org/indonesia/id/air-sanitasi-dan-kebersihan-wash/>.
- [2] Mardiasuti, A. (2022, September 22). Sanitasi adalah: Pengertian, jenis, manfaat dan contoh penerapannya. detikJabar. Diakses dari <https://www.detik.com/jabar/berita/d-6197674/sanitasi-adalah-pengertian-jenis-manfaat-dan-contoh-penerapannya>.
- [3] World Health Organization. (2021, November 13). WASH dalam pelayanan kesehatan bagi semua: Menghadapi kesetaraan gender dan inklusi sosial. Diakses dari <https://www.who.int/indonesia/id/news/detail/13-11-2021-wash-in-healthcare-services-for-all-addressing-gender-equality-and-social-inclusion>.
- [4] UN Water. (n.d.). *WASH – Water, sanitation and hygiene*. Diakses dari <https://www.unwater.org/water-facts/wash-water-sanitation-and-hygiene>.
- [5] United Nations. (n.d.). *Take action for the Sustainable Development Goals*. Diakses dari <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>.
- [6] United Nations. (n.d.). *Goal 6: Ensure access to water and sanitation for all*. Diakses dari <https://www.un.org/sustainabledevelopment/water-and-sanitation/>.
- [7] Maverick. (2018, November 14). *Apply CNN on Multivariate Time Series Forecasting (I): A process of applying CNN on Multivariate Time Series Forecasting*. Diakses dari <https://smallgum.github.io/2018/11/14/Apply-CNN-on-Multivariate-Time-Series-Forecasting-1/>.
- [8] Brownlee, J. (2020, August 28). *How to develop convolutional neural network models for time series forecasting*. Diakses dari <https://machinelearningmastery.com/how-to-develop-convolutional-neural-network-models-for-time-series-forecasting/>.
- [9] Azizah, R. N. (2017). Landasan Teori: Definisi Clustering. Umpro Repository. Diakses dari <http://eprints.umpo.ac.id>.
- [10] Patel, Eva., Kushwaha, Dharmender S.. 2020. Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. Elsevier
- [11] N. Rahayu, Geographically Weighted Panel Regression untuk Pemodelan Persentase Penduduk Miskin di Provinsi Jawa Tengah (Tesis). Surabaya: FMIPA Institut Teknologi Sepuluh Nopember, 2017.

- [12] United Nations, Department of Economic and Social Affairs, Statistics Division. (n.d.). *SDG indicators database*. Diakses dari <https://unstats.un.org/sdgs/dataportal/database>
- [13] Badan Pusat Statistik. (2023). Persentase Rumah Tangga yang Memiliki Akses terhadap Sumber Air Minum Layak Menurut Provinsi (Persen), 2021-2023. Diakses dari <https://bps.go.id>.
- [14] Badan Pusat Statistik. (2023). Persentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak Menurut Provinsi(Persen), 2017-2023. Diakses dari <https://bps.go.id>.
- [15] Badan Pusat Statistik. (2023). Proporsi Rumah Tangga Yang Memiliki Fasilitas Cuci Tangan Dengan Sabun Dan Air Menurut Provinsi (Persen), 2021-2023. Diakses dari <https://bps.go.id>.
- [16] Badan Pusat Statistik. (2023). Jumlah Penduduk Menurut Provinsi di Indonesia (Ribu Jiwa) 2017-2020. Diakses dari <https://bps.go.id>.
- [17] Badan Pusat Statistik (2024). Produk Domestik Regional Bruto Per Kapita (Ribu Rupiah), 2017-2020. Diakses dari <https://bps.go.id>.
- [18] Badan Pusat Statistik (2024). Realisasi Investasi Penanaman Modal Dalam Negeri Menurut Provinsi (Investasi) (Milyar Rupiah), 2017-2020. Diakses dari <https://bps.go.id>.
- [19] Badan Pusat Statistik (2024). Realisasi Investasi Penanaman Modal Luar Negeri Menurut Provinsi (Juta US\$), 2017-2020. Diakses dari <https://bps.go.id>.
- [20] Badan Pusat Statistik (2024). Jumlah Air Bersih yang Disalurkan Perusahaan Air Bersih (ribu m3), 2017-2020. Diakses dari <https://bps.go.id>.