

Mengungkap Hubungan Sosio-Ekonomi dan Akses Sanitasi Layak di Indonesia dengan SRGCNN

Kurniawan Chandra Wijaya¹, Dzulfahmi Dzakia Ahmad², Annas Banu Rusyidi³

¹Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada, Yogyakarta, email: kurniawanchandrawijaya0204@mail.ugm.ac.id

²Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada, Yogyakarta, email: dzulfahmidzakiaahmad@mail.ugm.ac.id

³Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada, Yogyakarta, email: annas.ban2002@mail.ugm.ac.id

Corresponding Author: Kurniawan Chandra Wijaya

ABSTRAK — Sanitasi yang layak merupakan salah satu aspek krusial dalam meningkatkan kualitas hidup manusia. Meskipun demikian, masih terdapat berbagai tantangan yang dihadapi untuk mewujudkannya. Penelitian ini bertujuan untuk mengeksplorasi pengaruh faktor-faktor sosio-ekonomi terhadap variabel target berupa persentase rumah tangga yang memiliki akses terhadap sanitasi layak. Algoritma *deep learning* berupa *Spatial Regression Graph Convolutional Neural Network* (SRGCNN) serta SRGCNN-Geographically Weighted (SRGCNN-GW) digunakan untuk memprediksi variabel target karena dapat memperhitungkan faktor sosio-ekonomi serta menangkap interaksi antarwilayah dengan memodelkan data dalam bentuk graf dan memperhitungkan data spasial antarwilayah. Selain itu, metode SRGCNN juga tidak memerlukan asumsi parametrik yang sering kali menjadi kendala pada metode konvensional. Hasil prediksi model SRGCNN dan SRGCNN-GW dibandingkan dengan beberapa model prediksi lainnya menggunakan metrik evaluasi *Mean Absolute Percentage Error* (MAPE) dan R^2 . Model prediksi yang dijadikan pembandingan adalah regresi linear, *Decision Tree Regressor*, *XGBoost Regressor*, *CatBoost Regressor*, dan *Multi Layer Perceptron*. Model SRGCNN-GW menjadi model terbaik karena memiliki performa yang jauh lebih baik serta memiliki metrik evaluasi paling optimal daripada model pembandingan. Hal ini ditandai dengan nilai MAPE sebesar 8,50% dan R^2 sebesar 92,14% pada model SRGCNN-GW. Dengan menguji kenormalan *error* dari model SRGCNN-GW menggunakan uji Kolmogorov-Smirnov, diperoleh nilai *p-value* sebesar 0,7151. Artinya, nilai *error* mengikuti distribusi normal. Hal ini mengindikasikan bahwasanya model SRGCNN-GW dapat melakukan prediksi dengan stabil. Dari beberapa variabel independen yang terlibat, pengeluaran per kapita disesuaikan (dalam ribu rupiah/orang/tahun) dan umur harapan hidup menjadi variabel yang memiliki pengaruh paling signifikan dalam model terhadap variabel target. Di sisi lain, variabel persentase penduduk miskin memiliki pengaruh yang paling rendah terhadap variabel target.

KATA KUNCI — SRGCNN, SRGCNN-GW, Sanitasi, Sosio-ekonomi.

I. PENDAHULUAN

A. LATAR BELAKANG

Sanitasi layak adalah kebutuhan mendasar bagi manusia yang berperan penting dalam menjaga kesehatan dan kenyamanan setiap individu. Akses terhadap fasilitas sanitasi yang memadai dapat membantu mencegah penyebaran penyakit dan memastikan lingkungan hidup yang sehat. Menurut World of Health Organization (WHO), kurangnya sanitasi yang layak dapat berdampak negatif terhadap kesejahteraan masyarakat serta dapat memperlambat kemajuan sosial dan ekonomi sebuah negara.

Sayangnya, akses terhadap sanitasi layak masih menjadi tantangan besar di seluruh dunia. Pada tahun 2023, diperkirakan 1,7 miliar orang di dunia tidak memiliki akses ke fasilitas sanitasi dasar, yang mencakup pembuangan kotoran manusia yang aman dan akses ke tempat cuci tangan dengan sabun dan air di rumah [1]. Kondisi ini tentunya sangat memprihatinkan. Di Indonesia, permasalahan akses sanitasi layak juga menunjukkan adanya kesenjangan yang signifikan antara wilayah. Berdasarkan data Badan Pusat Statistik (BPS) Pusat, terdapat disparitas yang cukup besar dalam hal akses rumah tangga terhadap sanitasi yang layak pada tahun 2023. Persentase rumah tangga yang memiliki akses terhadap sanitasi layak di Kota Denpasar menyentuh persentase sebesar 99,92%,

sedangkan Kabupaten Yalimo hanya berada pada angka 0,48%. Kesenjangan ini menunjukkan bahwa tantangan besar dalam penyediaan sanitasi layak di Indonesia masih perlu mendapat perhatian serius dari berbagai pihak.

Sebagai bagian dari upaya global untuk mengatasi permasalahan tersebut, Perserikatan Bangsa-Bangsa melalui *Sustainable Development Goals* (SDGs) telah menempatkan akses terhadap sanitasi layak sebagai salah satu tujuan SDGs. Tujuan 6 SDGs, yaitu Air Bersih dan Sanitasi Layak, secara khusus menargetkan agar seluruh populasi di dunia memiliki akses penuh terhadap air bersih dan sanitasi layak pada tahun 2030. Target ini mencerminkan pentingnya sanitasi dalam mendukung kesehatan masyarakat dan pembangunan berkelanjutan di masa depan.

Dalam mencapai target tujuan 6 SDGs di Indonesia, tentunya hal ini tidak hanya bergantung pada penyediaan infrastruktur sanitasi, tetapi juga pada pemahaman yang mendalam terhadap berbagai faktor yang mempengaruhi akses sanitasi layak di setiap kabupaten/kota. Dengan demikian, dibutuhkan analisis lebih mendalam mengenai faktor-faktor yang mempengaruhi akses sanitasi layak di kabupaten/kota seluruh Indonesia. Pada penelitian ini, faktor yang dipertimbangkan hanyalah faktor sosio-ekonomi.

B. TUJUAN

Tujuan dari penelitian ini adalah melakukan eksperimen implementasi pendekatan *deep learning* untuk menganalisis faktor-faktor sosio-ekonomi yang memengaruhi akses terhadap sanitasi layak di berbagai wilayah di Indonesia. Dengan memahami pengaruh dari faktor-faktor ini, penelitian ini diharapkan dapat memberikan wawasan mendalam mengenai kesenjangan akses sanitasi dan memberikan rekomendasi kebijakan yang dapat membantu mencapai target tujuan 6 SDGs.

II. KAJIAN TERKAIT

Banyak penelitian telah mengkaji faktor-faktor yang mempengaruhi akses rumah tangga terhadap sanitasi layak. Beberapa di antaranya mencakup aspek sosial, ekonomi, dan lingkungan. Berdasarkan penelitian pada referensi [2], dengan pendekatan analisis regresi logistik berganda, ditemukan bahwa pengeluaran rumah tangga dan perilaku masyarakat berpengaruh signifikan terhadap ketersediaan sanitasi. Jika pengeluaran rumah tangga meningkat, ketersediaan sanitasi rumah tangga ikut meningkat. Namun, kelemahan penelitian pada referensi [2] adalah tidak dilakukan pemeriksaan asumsi analisis regresi logistik berganda untuk memastikan validitas model regresi yang dibentuk.

Referensi [3] menggunakan metode regresi linear berganda untuk mengeksplorasi hubungan antara berbagai faktor sosio-ekonomi terhadap sanitasi layak di Indonesia. Hasil penelitian menunjukkan bahwa tingkat penyelesaian pendidikan SMA/Sederajat dan kepemilikan jamban sendiri signifikan meningkatkan akses sanitasi layak. Di samping itu, menurut penelitian pada referensi [4], dengan pendekatan metode regresi linear berganda, ditemukan bahwa indeks pembangunan manusia (IPM) memiliki korelasi positif yang signifikan terhadap akses sanitasi.

Namun, penelitian-penelitian tersebut memiliki keterbatasan dalam memperhitungkan dimensi spasial dari sanitasi layak di berbagai kabupaten/kota Indonesia. Keterkaitan spasial antarwilayah belum dieksplorasi lebih lanjut, padahal wilayah yang berdekatan cenderung memiliki kesamaan karakteristik dan sering kali saling memengaruhi satu sama lain.

Sebagai tindak lanjut dari beberapa penelitian tersebut, metode *Spatial Regression Graph Convolutional Neural Network* (SRGCNN) diusulkan dalam penelitian ini. SRGCNN mampu menangkap interaksi antarwilayah dengan memodelkan data dalam bentuk graf dan memperhitungkan hubungan spasial antarwilayah. Metode ini memungkinkan analisis yang lebih komprehensif terhadap akses sanitasi layak di berbagai kabupaten/kota di Indonesia.

III. SOLUSI DAN USULAN

A. ALGORITMA METODE SPATIAL REGRESSION GRAPH CONVOLUTIONAL NEURAL NETWORK (SRGCNN)

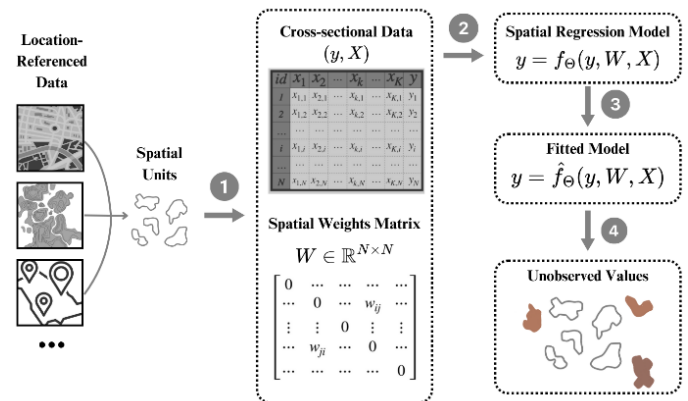
Dalam upaya mengetahui pengaruh faktor-faktor sosio-ekonomi terhadap akses sanitasi layak di wilayah di Indonesia, penelitian ini mengusulkan penerapan metode *Spatial Regression Graph Convolutional Neural Network* (SRGCNN).

SRGCNN merupakan model *deep learning* yang dirancang untuk mengintegrasikan data spasial dan hubungan antarwilayah melalui representasi graf. Selain memperhitungkan pengaruh sosio-ekonomi, metode ini juga dapat memperhitungkan pengaruh data spasial dalam memprediksi akses sanitasi suatu kabupaten/kota, sehingga menghasilkan prediksi yang lebih akurat dibandingkan metode konvensional. Selain itu, metode SRGCNN juga tidak memerlukan asumsi parametrik yang sering kali menjadi kendala pada metode konvensional, sehingga penelitian ini dapat dilakukan tanpa perlu mengkhawatirkan mengenai asumsi tersebut.

Metode SRGCNN mengadopsi metode analisis regresi linear dengan bobot spasial. Bentuk umum dari model tersebut adalah sebagai berikut.

$$\mathbf{y} = \mathbf{f}_{\theta}(\mathbf{y}, \mathbf{W}, \mathbf{X}) + \epsilon. \quad (3.1)$$

Pembentukan model diawali dengan pembentukan matriks pembobot spasial atau *spatial weights* $\mathbf{W} \in \mathbb{R}^{N \times N}$ untuk observasi spasial sebanyak N yang memiliki variabel independen sebanyak K , yaitu $\mathbf{X} \in \mathbb{R}^{N \times K}$ dan variabel dependen $\mathbf{y} \in \mathbb{R}^{N \times 1}$. Himpunan parameter yang akan diestimasi adalah θ yang berisi efek spasial dan nonspasial pada model. Kerangka kerja model SRGCNN ditunjukkan pada gambar 1 [5].



Gambar 1. Kerangka Kerja Model SRGCNN

Mekanisme *backpropagation* dalam SRGCNN serupa dengan model *multi-layer GCNN* sebagai berikut.

1. Pembuatan *fully-connected graph* dalam GCNN dengan matriks pembobot spasial \mathbf{W} dan data (\mathbf{X}, \mathbf{y}) . Graf $G = (V, E)$ dibentuk untuk merepresentasikan struktur spasial dari N observasi spasial. Setiap titik spasial dianggap sebagai *node* $v_i \in V$ pada graf G dan direpresentasikan dengan fitur *node* (\mathbf{X}, \mathbf{y}) . Matriks pembobot spasial \mathbf{W} menunjukkan ukuran hubungan dalam graf yang direpresentasikan oleh *edge* graf E , dengan $e_{ij} = (v_i, v_j, w_{ij}) \in E$ merupakan *edge* antara v_i dan v_j dengan bobot $w_{ij} \in \mathbf{W}$.
2. Model GCNN memperhitungkan efek *lag* spasial pada *forward propagation* dari fitur graf *input*. Implementasi GCNN dengan aturan *layer-wise neural network forward propagation*, yaitu

$$\begin{aligned} \mathbf{X}^{l+1} &= \sigma(\mathbf{Z}^{l+1}) = \sigma(\mathbf{W}_L \mathbf{X}^l \boldsymbol{\theta}^l) \\ &= \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{W}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}^l \boldsymbol{\theta}^l\right), \end{aligned} \quad (3.2)$$

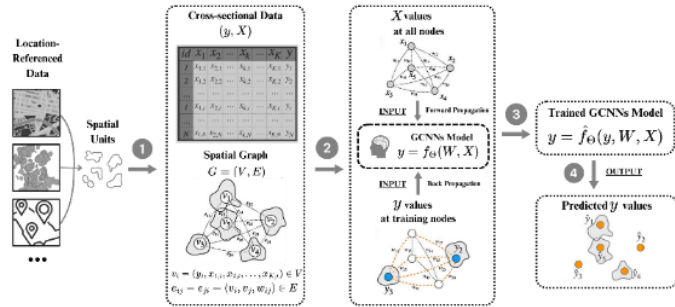
dengan \mathbf{X}^l merupakan matriks fitur aktivasi pada *layer* ke- l , $\sigma(\cdot)$ merupakan fungsi aktivasi nonlinear, dan Θ^l merupakan matriks parameter dalam *layer*. Bentuk $\mathbf{W}_L = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{W}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}^l$ merupakan matriks renormalisasi Laplace dari $\tilde{\mathbf{W}} = \mathbf{W} + \mathbf{I}$.

3. Parameter Θ pada *neural network* diestimasi dengan algoritma *backpropagation* dan *gradient descent*.

Model SRGCNN dapat disederhanakan menjadi persamaan berikut.

$$\mathbf{X}^{(l+1)} = \mathbf{W} \times \mathbf{X}^{(l)} \times \Theta^{(l)}. \quad (3.3)$$

Diasumsikan terdapat observasi spasial sebesar N , fitur *input* dengan C_{in} *channels* dan fitur *output* dengan C_{out} *channels*, didapatkan $\mathbf{X}^{(l+1)} \in \mathbb{R}^{N \times C_{out}}$, $\mathbf{X}^{(l)} \in \mathbb{R}^{N \times C_{in}}$, $\mathbf{W} \in \mathbb{R}^{N \times N}$, dan $\Theta^{(l)} \in \mathbb{R}^{C_{in} \times C_{out}}$. Kerangka kerja SRGCNN sebagai suatu model *deep learning* ditunjukkan pada gambar 2 [5].



Gambar 2. SRGCNN sebagai *Deep Learning*

Parameter $\Theta^{(l)}$ sama untuk setiap *neuron* GCNN yang menghasilkan regresi spasial global dengan struktur graf *local-connected*. Oleh karena itu, model dapat dikembangkan menjadi SRGCNN-Geographically Weighted atau SRGCNN-GW. Model tersebut memiliki pendekatan yang serupa dengan *Geographically Weighted Regression* (GWR), yaitu melakukan ekspansi regresi spasial global menjadi regresi spasial lokal.

$$\mathbf{X}^{(l+1)} = \mathbf{W} \times (\mathbf{X}^{(l)} \otimes \Theta_{\text{lokal}}^{(l)}) \times \Theta^{(l)}, \quad (3.4)$$

dengan $\Theta_{\text{lokal}}^{(l)} \in \mathbb{R}^{N \times C_{in}}$ mengandung parameter *geographically weighted* yang berbeda pada tiap daerah. Berbeda dengan GWR, metode SRGCNN-GW dapat digunakan untuk memprediksi daerah baru dengan memperhitungkan kembali matriks pembobot spasialnya.

B. PEMROSESAN DATA

Penelitian ini menggunakan data sekunder yang bersumber dari hasil publikasi Badan Pusat Statistik (BPS) pada tahun 2023 [7]. Ruang lingkup penelitian ini mencakup 514 kabupaten/kota di Indonesia. Variabel dependen yang dianalisis adalah persentase rumah tangga yang memiliki akses terhadap sanitasi layak, sedangkan variabel independen yang dianalisis adalah

persentase penduduk miskin, rata-rata lama sekolah (tahun), pengeluaran per kapita disesuaikan (dalam ribu rupiah/orang/tahun), pengeluaran per kapita/bulan, umur harapan hidup, harapan lama sekolah, indeks pembangunan manusia, tingkat pengangguran terbuka, dan persentase rumah tangga yang memiliki akses terhadap air minum layak. Variabel-variabel independen tersebut dijadikan sebagai representasi faktor-faktor sosio-ekonomi. Selain itu, untuk merepresentasikan wilayah geografis setiap kabupaten/kota, ditambahkan data poligon yang diperoleh dari situs Dinas Kependudukan dan Catatan Sipil [6]. Data poligon ini mencakup batas administrasi resmi kabupaten/kota di seluruh Indonesia.

Dalam mengimplementasikan model SRGCNN terhadap data, digunakan perangkat lunak Python dengan berbagai pustaka untuk memfasilitasi pengolahan data graf secara efisien. Infrastruktur yang digunakan adalah Google Colab karena terdapat unit pengolahan grafis (GPU) yang dapat mempercepat proses komputasi.

Sebelum melakukan analisis dengan menggunakan metode SRGCNN, penting untuk dipastikan bahwa setiap tahap pemrosesan data dilakukan dengan teliti. Berikut ini merupakan tahapan dalam persiapan data.

1. Menghimpun data variabel independen dan variabel dependen dari setiap hasil publikasi BPS Provinsi pada tahun 2023. Keseluruhan data tersebut digabung menjadi satu dengan 514 kabupaten/kota sebagai subjek penelitian.
2. Mengubah setiap data yang dinyatakan dalam bentuk persentase menjadi bentuk desimal.
3. Mengintegrasikan data poligon setiap kabupaten/kota dengan data keseluruhan variabel sehingga memungkinkan model SRGCNN untuk menganalisis hubungan spasial antarwilayah.
4. Melakukan normalisasi data variabel independen dengan menggunakan metode *Min-Max Scaling*. Normalisasi dilakukan untuk menghindari dominasi variabel dengan nilai rentang yang lebih besar dibandingkan variabel lainnya, sehingga dapat memastikan data berada pada rentang yang sama. Perhitungan normalisasi dilakukan dengan rumus berikut.

$$x_{\text{scaled}} = \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (3.5)$$

5. Membentuk graf dengan menghubungkan setiap titik pusat atau *centroid* poligon dengan titik-titik di sekitarnya berdasarkan *k-Nearest Neighbor* (*k-NN*). Nilai k yang dipilih adalah 5. Gambar 3 menunjukkan bagaimana graf terbentuk dengan pendekatan *k-NN* pada data yang digunakan.

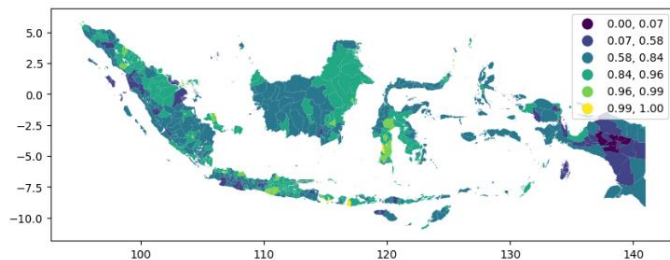
Gambar 3. Graf Kabupaten/Kota dengan k -NN

6. Memecah data menjadi data latih, data validasi, dan data uji dengan rasio secara berurutan sebesar 70%, 15%, dan 15%. Pemecahan ini dilakukan secara acak.

IV. HASIL EKSPERIMEN DAN PENGUJIAN

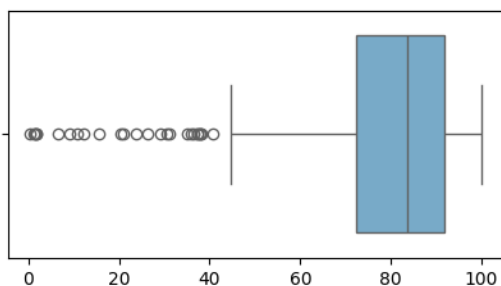
A. ANALISIS DATA EKSPLORATIF

Analisis data eksploratif merupakan langkah penting dalam memahami karakteristik dan pola yang terdapat dalam dataset sebelum melakukan analisis lebih lanjut. Dengan melakukan eksplorasi terhadap data, dapat diidentifikasi tren, pola, serta *outlier* yang mungkin tersembunyi.



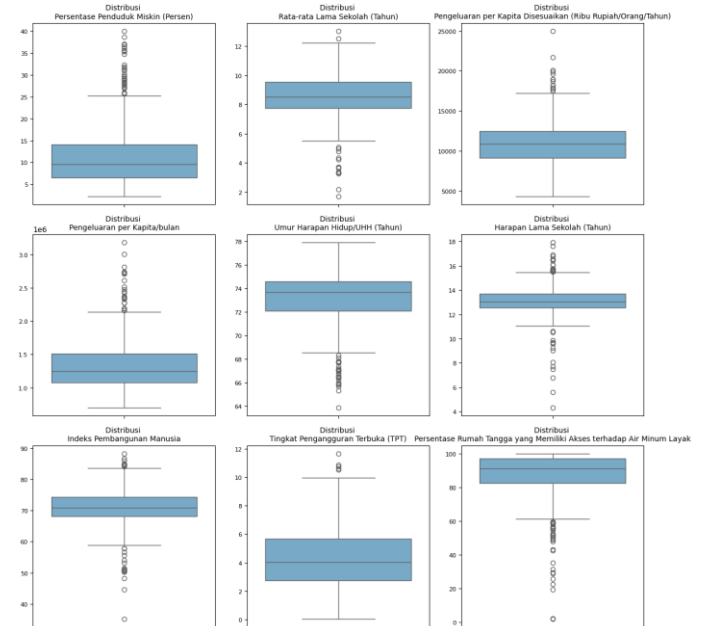
Gambar 4. Peta Distribusi Persentase Akses Sanitasi Layak

Gambar 4 menampilkan peta distribusi persentase rumah tangga yang memiliki akses terhadap sanitasi layak di seluruh kabupaten/kota Indonesia. Gambar ini mengungkapkan adanya kesenjangan antara wilayah bagian barat dan timur Indonesia. Kabupaten/kota di wilayah bagian barat Indonesia cenderung memiliki persentase akses sanitasi layak yang lebih tinggi, khususnya di daerah Pulau Jawa. Hal ini berkebalikan dengan kabupaten/kota di wilayah bagian timur Indonesia yang memiliki persentase akses sanitasi layak yang cenderung rendah.



Gambar 5. Distribusi Persentase Akses Sanitasi Layak

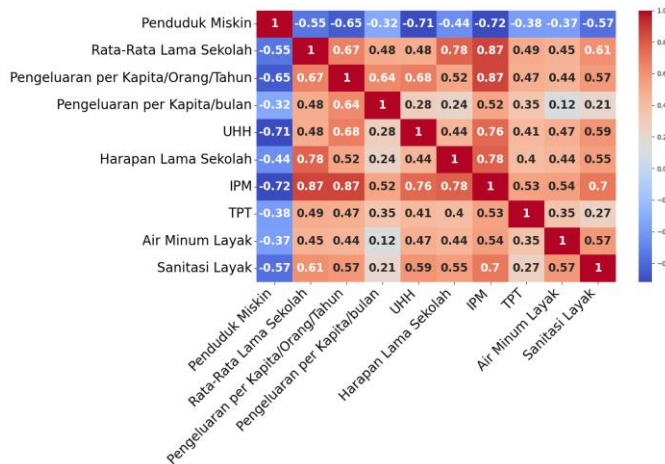
Gambar 5 merupakan visualisasi *boxplot* dari variabel dependen. *Boxplot* adalah grafik yang digunakan untuk menggambarkan distribusi data berdasarkan kuartilnya. Berdasarkan *boxplot* tersebut, menguatkan bukti adanya ketimpangan yang cukup besar dalam akses sanitasi layak di seluruh kabupaten/kota. Gambar ini menunjukkan bahwa lebih banyak kabupaten/kota memiliki persentase akses sanitasi layak yang lebih tinggi dibandingkan yang lebih rendah. Wilayah dengan akses yang sangat rendah menciptakan beberapa *outliers* ke arah bawah yang mengindikasikan wilayah tersebut sangat tertinggal dalam hal akses terhadap sanitasi layak.



Gambar 6. Distribusi Variabel Independen

Gambar 6 merupakan *boxplot* yang menunjukkan distribusi berbagai indikator sosio-ekonomi dari seluruh kabupaten/kota Indonesia. Berdasarkan gambar tersebut, terlihat bahwa terdapat kesenjangan pada beberapa indikator sosio-ekonomi, seperti pada persentase penduduk miskin. Terdapat pencilon yang menandakan sangat tertinggalnya kondisi beberapa kabupaten/kota di Indonesia dibandingkan wilayah lainnya. Hal

ini mengindikasikan adanya ketimpangan yang signifikan antara wilayah maju dan tertinggal.



Gambar 7. Analisis Korelasi Antarvariabel

Gambar 7 merupakan visualisasi *heatmap correlation* dari kesepuluh variabel. *Heatmap correlation* adalah visualisasi yang dapat digunakan sebagai analisis korelasi guna mengetahui eratny hubungan antara variabel independen dan variabel dependen. Berdasarkan *heatmap correlation* dapat diketahui bahwa variabel independen yang memiliki hubungan terkuat terhadap variabel dependen adalah variabel indeks pembangunan manusia dengan nilai korelasi sebesar 0,7. Nilai positif menunjukkan hubungan yang searah. Sementara, variabel pengeluaran per kapita/bulan menjadi variabel independen dengan hubungan terlemah terhadap variabel dependen.

B. PEMODELAN DATA METODE SRGCNN

Pada penelitian ini, digunakan beberapa parameter dalam model SRGCNN. Berikut penjelasan parameter-parameter yang digunakan.

1. $n_features$ adalah parameter yang mewakili jumlah fitur *input* atau jumlah variabel independen yang digunakan dalam model. Jumlah fitur yang dilibatkan pada penelitian ini sebanyak 9 variabel.
2. n_labels adalah parameter yang mengacu pada jumlah variabel dependen yang diestimasi secara bersamaan dalam model. Pada penelitian ini, digunakan kelas target yang diprediksi adalah 1 karena model hanya melakukan prediksi satu variabel dependen atau target *output*.
3. *Epoch* adalah parameter yang menunjukkan jumlah iterasi selama proses pelatihan data. Selama pelatihan data menggunakan metode SRGCNN ini digunakan iterasi sebanyak 18.000 kali.
4. *Learning rate* (η) adalah parameter yang digunakan untuk menentukan seberapa besar perubahan bobot pada setiap iterasi selama pelatihan. Nilai η yang digunakan adalah sebesar 0,01 untuk model SRGCNN dan 0,001 untuk model SRGCNN-GW.
5. *Hidden layer* adalah parameter yang digunakan untuk menunjukkan jumlah *layer* serta banyaknya *nodes* setiap *layer* yang digunakan pada bagian *hidden layer*. Penelitian

ini menggunakan 3 *hidden layer* dengan jumlah *nodes* setiap *layer*-nya sebagai berikut.

- *Layer* pertama terdapat $1 \times n_features$ atau 9 *nodes*.
 - *Layer* kedua terdapat $4 \times n_features$ atau 36 *nodes*.
 - *Layer* ketiga terdapat $8 \times n_features$ atau 72 *nodes*.
6. *Dropouts* adalah parameter yang digunakan untuk mencegah terjadinya *overfitting* dengan menonaktifkan sejumlah *nodes* pada setiap *hidden layer* selama proses pelatihan.
 - Jumlah *nodes* yang dinonaktifkan dari *layer* pertama ke *layer* kedua sebesar 10%.
 - Jumlah *nodes* yang dinonaktifkan dari *layer* kedua ke *layer* ketiga sebesar 30%.
 - Jumlah *nodes* yang dinonaktifkan dari *layer* ketiga ke *output layer* sebesar 50%.
 7. Fungsi aktivasi ReLU adalah fungsi aktivasi yang akan mengaktifkan *nodes* jika nilainya positif dan akan menonaktifkan *nodes* jika nilainya negatif (nilainya diubah menjadi 0). Penelitian ini menggunakan fungsi aktivasi ReLU pada semua *layer*. Fungsi aktivasi ReLU dapat dihitung dengan menggunakan persamaan berikut.

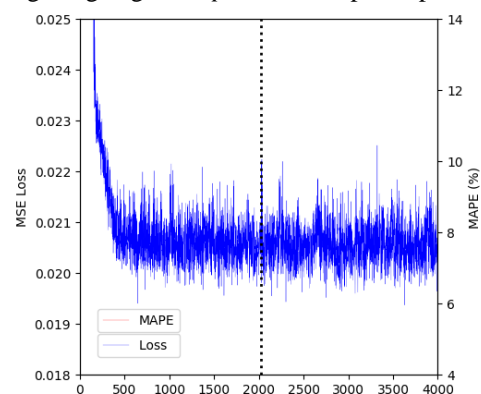
$$\text{ReLU}(x) = \max(0, x). \quad (4.1)$$
 8. Fungsi aktivasi sigmoid adalah fungsi yang dapat mengubah nilai *output* ke dalam rentang nilai 0 dan 1. Penelitian ini menggunakan fungsi aktivasi sigmoid karena untuk menyesuaikan data variabel dependen yang memiliki nilai desimal pada rentang 0 dan 1. Fungsi aktivasi sigmoid ini hanya diterapkan pada *output layer*. Perhitungan fungsi aktivasi sigmoid dilakukan dengan persamaan berikut.

$$\sigma(X) = \frac{1}{1 + e^{-X}}. \quad (4.2)$$
 9. *Loss function* berupa *mean squared error*, yaitu

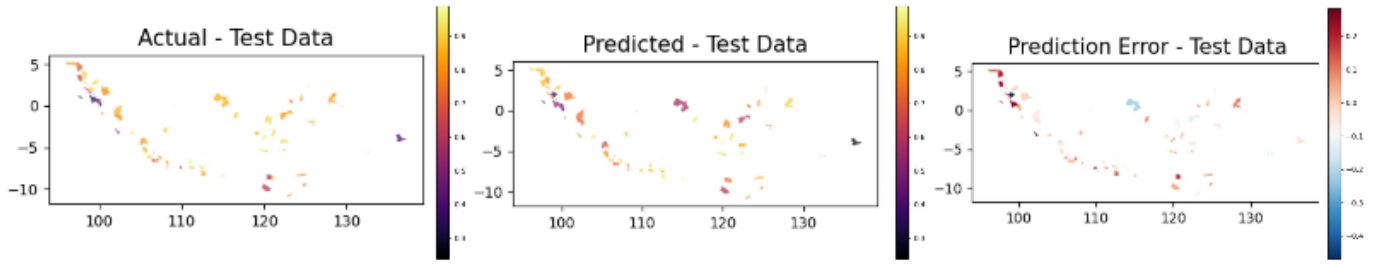
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (4.3)$$
 10. Algoritma *early stopping* untuk mencegah terjadinya *overfitting* pada data. *Early stopping* dilakukan dengan *patience* sebesar 5.000. Algoritma ini akan menghentikan proses iterasi apabila terdeteksi adanya *overfitting*.

C. EVALUASI KINERJA MODEL

Pemodelan dengan metode SRGCNN dilakukan dengan data yang telah dipecah. Gambar 8 menunjukkan bagaimana model berkembang dengan grafik *epoch* terhadap *loss* pada data latih.

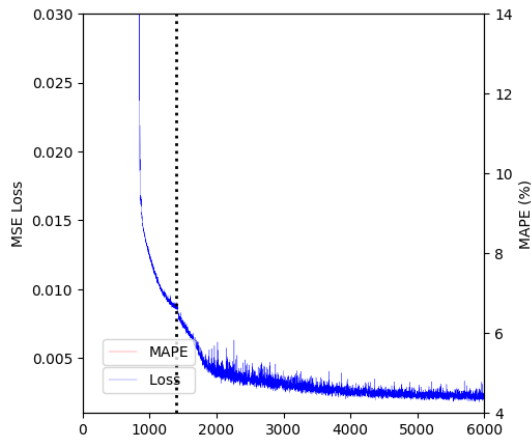


Gambar 8. Metrik Kinerja SRGCNN



Gambar 9. Peta Nilai Aktual, Prediksi, dan Error dari Data Tes Model SRGCNN-GW

Di sisi lain, gambar 10 menunjukkan bagaimana model SRGCNN-GW berkembang.



Gambar 10. Metrik Kinerja SRGCNN-GW

Dalam mengevaluasi kinerja model, akan digunakan dua metrik evaluasi, yaitu *Mean Absolute Percentage Error* (MAPE) dan R^2 . MAPE digunakan untuk mengukur akurasi prediksi dengan menghitung rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual, sedangkan R^2 digunakan untuk mengukur seberapa baik model mampu menjelaskan variabilitas dari variabel dependen oleh variabel-variabel independen. Rumus kedua metrik tersebut dinyatakan dalam persamaan-persamaan berikut.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|; \quad (4.4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (4.5)$$

Tabel I menunjukkan nilai kedua metrik evaluasi pada data uji menggunakan model SRGCNN, SRGCNN-GW, dan ditambahkan beberapa model pembandingan.

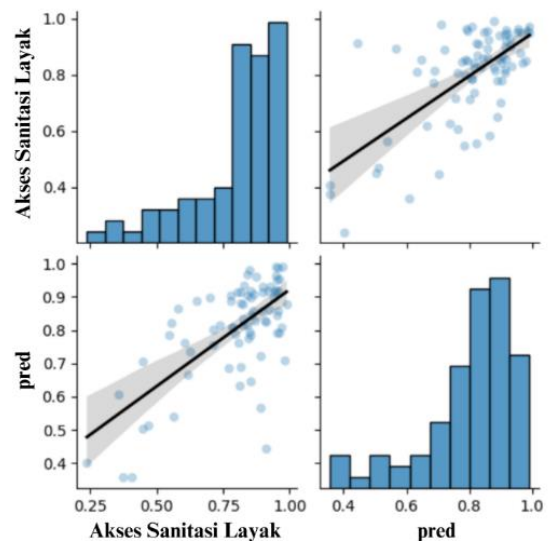
TABEL I
PERBANDINGAN METRIK EVALUASI

Model	MAPE	R^2
SRGCNN	11,60%	90,98%
SRGCNN-GW	8,50%	92,14%
Regresi Linear	15,89%	61,74%
Decision Tree Regressor	16,01%	36,28%
XGBoost Regressor	22,21%	-44,98%
CatBoost Regressor	14,22%	47,59%
Multi Layer Perceptron	21,00%	-6,48%

Terlihat bahwa kedua model SRGCNN memiliki performa yang jauh lebih baik daripada model-model pembandingan lainnya. Model SRGCNN-GW memiliki metrik evaluasi paling optimal apabila dibandingkan dengan model lainnya. Oleh karena itu, model SRGCNN-GW dianggap sebagai model terbaik. Hal ini juga mempertimbangkan bagaimana model SRGCNN-GW dapat memperhitungkan parameter-parameter lokal pada data. Oleh karena itu, analisis selanjutnya akan digunakan model SRGCNN-GW.

V. ANALISIS HASIL EKSPERIMEN DAN PENGUJIAN

Gambar 9 merupakan pemetaan dari data aktual, data hasil prediksi, dan nilai error pada data tes dengan model SRGCNN-GW. Berdasarkan gambar 9, dapat dilihat bahwa pada peta nilai aktual, distribusi persentase rumah tangga yang memiliki akses terhadap sanitasi layak di wilayah data uji bervariasi. Pada peta nilai prediksi, dapat dilihat bahwa terdapat warna wilayah yang mendekati aktual. Hal ini menunjukkan bahwa model prediksi cukup bekerja dengan baik dalam memperkirakan akses sanitasi layak di sebagian besar wilayah. Pada peta error, dapat dilihat distribusi dari kesalahan prediksi. Sebagian besar wilayah memiliki warna yang mendekati putih. Hal ini menunjukkan bahwa kesalahan prediksi kecil atau hampir tidak ada.

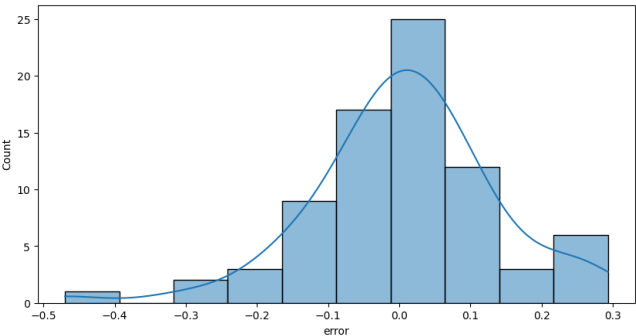


Gambar 11. Pair Plot Data Asli dan Prediksi Model SRGCNN-GW

Gambar 11 merupakan *pair plot* yang membandingkan nilai aktual variabel target dengan nilai prediksi model SRGCNN-GW. Berdasarkan histogram data aktual, diperoleh bahwa sebagian besar kabupaten/kota memiliki persentase akses

sanitasi layak yang relatif tinggi. Jika melihat histogram nilai prediksi, diketahui bahwa model SRGCNN-GW sebagian besar memprediksi dengan akurasi yang baik terhadap nilai akses sanitasi layak yang sebenarnya, ditandai dengan bentuk distribusi hasil prediksi yang mirip dengan distribusi data aktual. Berdasarkan *scatter plot*, dapat ditunjukkan adanya hubungan linier positif yang kuat antara nilai aktual dan prediksi, artinya model mampu memprediksi data dengan baik.

Kestabilan model SRGCNN-GW dalam memprediksi dapat dilihat dengan mengecek apakah distribusi *error* mengikuti distribusi normal atau tidak. Kenormalan *error* mengindikasikan kestabilan model dalam memprediksi. Hal ini berarti model tidak memprediksi nilai variabel dependen dengan kecenderungan lebih tinggi atau lebih rendah. Distribusi *error* dapat dilihat pada gambar 12 berikut.



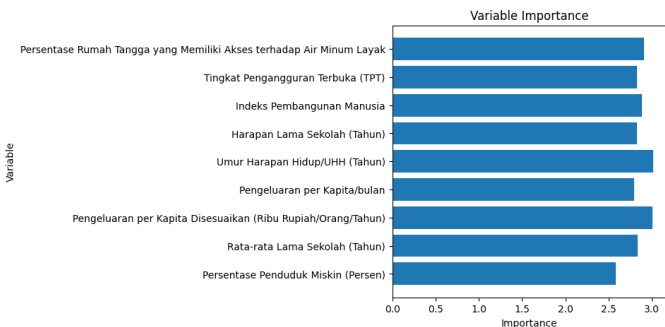
Gambar 12. Distribusi Error Model SRGCNN-GW

Berdasarkan gambar 12, terlihat bahwa *error* cenderung membentuk distribusi normal. Untuk memastikan kesimpulan, statistika inferensial dapat digunakan. Uji yang digunakan untuk melihat normalitas data adalah uji Kolmogorov-Smirnov. Uji ini cocok digunakan untuk melihat normalitas data untuk ukuran data yang besar. Hipotesis uji ini adalah H_0 : data berdistribusi normal lawan H_1 : data tidak berdistribusi normal.

TABEL II
UJI NORMALITAS DATA ERROR

Komponen Uji	Nilai
Statistik Uji Kolmogorov-Smirnov	0,0770
Tingkat signifikansi (α)	0,05
<i>p-value</i>	0,7151
Kesimpulan: Gagal menolak H_0 , terdapat cukup bukti untuk mengatakan bahwa <i>error</i> berdistribusi normal.	

Sebab *error* ditunjukkan sudah mengikuti distribusi normal, model SRGCNN-GW dapat dikatakan memberikan hasil prediksi yang stabil.



Gambar 13. Variable Importance

Gambar 13 merupakan diagram batang dari pemetaan dari *variable importance*. Diagram ini menunjukkan kontribusi masing-masing variabel model SRGCNN dalam memprediksi akses sanitasi layak. Berdasarkan gambar 13, dapat dilihat bahwa pengeluaran per kapita yang disesuaikan dan umur harapan hidup memiliki pengaruh yang paling signifikan dalam model terhadap variabel target. Selain itu, dapat dilihat bahwa variabel persentase penduduk miskin memiliki pengaruh yang paling rendah terhadap variabel target apabila dibanding variabel lainnya.

VI. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa metode *Spatial Regression Graph Convolutional Neural Network–Geographically Weighted* (SRGCNN-GW) efektif dalam menganalisis faktor-faktor sosio-ekonomi yang memengaruhi akses terhadap sanitasi layak di Indonesia. Hasil penelitian ini menunjukkan bahwa faktor-faktor sosio-ekonomi memiliki pengaruh yang signifikan terhadap akses sanitasi layak di Indonesia. SRGCNN-GW mampu menangkap interaksi spasial antarwilayah sehingga menghasilkan prediksi yang lebih akurat apabila dibandingkan metode konvensional yang hanya memperhitungkan data nonspasial. Hasil ini menunjukkan bahwa kesenjangan dalam akses sanitasi tidak hanya dipengaruhi oleh faktor ekonomi semata, tetapi juga oleh karakteristik spasial antarwilayah.

Secara praktis, hasil penelitian ini dapat diterapkan untuk membantu pemerintah sebagai dasar pengambilan keputusan untuk menyusun kebijakan guna mengurangi ketimpangan sanitasi di wilayah Indonesia. Implementasi SRGCNN-GW dirasa perlu karena Indonesia memiliki keberagaman wilayah dan kondisi sosio-ekonomi yang berbeda-beda. Hasil prediksi dari model SRGCNN-GW dapat membantu pemerintah dalam mengidentifikasi wilayah dengan akses sanitasi yang rendah. Dengan informasi tersebut, program pemerintah, seperti *Sanitasi Total Berbasis Masyarakat (STBM)*, dapat diperkuat dan difokuskan ke wilayah-wilayah yang paling membutuhkan intervensi sanitasi layak.

Selain itu, hal yang dapat disarankan untuk penelitian selanjutnya adalah terkait pengembangan lebih lanjut dari metode ini, seperti penambahan variabel-variabel lingkungan dan perilaku masyarakat untuk memberikan gambaran yang lebih menyeluruh mengenai faktor-faktor yang memengaruhi akses terhadap sanitasi layak serta mempercepat tujuan 6 SDGs.

REFERENSI

[1] The United Nations World Water Development Report 2023. 2023. doi: 10.18356/9789210026208.

[2] E. R. Budi, “Analisis kondisi sosial ekonomi masyarakat terhadap ketersediaan sanitasi,” *Economics Development Analysis Journal*, vol. 6, no. 2, pp. 147–154, Mar. 2018, doi: 10.15294/edaj.v6i2.22211.

[3] D. Widyastuti, H. N. Jamaluddin, R. Arisanti, and F. Kartiasih, “Analisis Pengaruh Faktor Sosial Ekonomi Terhadap Akses Sanitasi Layak di Indonesia Tahun 2021,” *Seminar Nasional Official Statistics*, vol. 2023, no. 1, pp. 105–116, Oct. 2023, doi: 10.34123/semnasoffstat.v2023i1.1853.

[4] I. Y. Septiariwa, W. Prayogo, and I. W. K. Suryawan, “Sanitation access in the developing country: the impact of human development, income inequality, and economic growth,” *REGION: Jurnal Pembangunan Wilayah Dan Perencanaan Partisipatif*, vol. 19, no. 2, Art. no. 10.20961, Jul. 2024, doi: 10.20961/region.v19i2.83327.

[5] D. Zhu, Y. Liu, X. Yao, and M. M. Fischer, “Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions,” *GeoInformatica*, vol. 26, no. 4, pp. 645–676, Nov. 2021, doi: 10.1007/s10707-021-00454-x.

- [6] "Batas Administrasi Provinsi, Kabupaten/Kota, Kecamatan, Desa tahun 2019." Dinas Kependudukan dan Pencatatan Sipil Kementerian Dalam Negeri, 2019. [Online]. Available: <https://www.indonesia-geospasial.com/2020/04/download-shapefile-shp-batas.html>
- [7] "Data Sosial Ekonomi 37 Provinsi di Indonesia." Badan Pusat Statistik, 2023. [Online]