



APLIKASI GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) UNTUK PEMETAAN FAKTOR YANG MEMPENGARUHI INDEKS AKTIVITAS LITERASI MEMBACA DI INDONESIA

ALFISYHRINA HAPSERY, DEA TRISHNANTI

Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas PGRI Adi Buana Surabaya

alfisyahrina@unipasby.ac.id

ABSTRAK

Papua memiliki angka indeks literasi terendah yaitu 19,9 dan Jakarta merupakan daerah dengan angka indeks literasi tertinggi yaitu mencapai 58,16. Namun Jakarta dengan angka indeks literasi tertinggi menempati posisi angka indeks literasi sedang karena masih dibawah angka 60. Papua dan Jakarta adalah dua provinsi yang sangat mencolok perbedaannya dari segi geografis dimana Jakarta terletak di pulau Jawa dan merupakan Ibukota negara sedangkan Papua berada dipulau paling timur Indonesia. Oleh karena itu diperlukan suatu metode pemodelan statistik dengan memperhitungkan aspek spasial yaitu menggunakan metode Geographically Weighted Regression (GWR) yang diharapkan mampu menghasilkan model angka indeks literasi di Indonesia tahun 2017 di tiap wilayah. GWR adalah salah satu metode spasial yang menggunakan faktor geografis sebagai variabel bebas yang dapat mempengaruhi variabel dependen. Berdasarkan hasil analisis dan pembahasan dengan menggunakan analisis GWR diketahui bahwa pada pengelompokkan variabel berpengaruh signifikan terhadap indeks aktivitas literasi diperoleh 11 kelompok dimana kelompok 1 hanya terdapat satu variabel yang berpengaruh signifikan terhadap indeks aktivitas literasi yaitu persentase melek huruf latin yang terdapat di Provinsi Papua. Sedangkan, pada kelompok 11 dimana semua variabel independen berpengaruh signifikan terhadap indeks aktivitas literasi terdapat di Provinsi Jambi, Provinsi Sumatera Selatan, dan Provinsi Lampung. Hasil analisis juga menunjukkan bahwa pemodelan terbaik dari beberapa metode regresi adalah model GWR karena memiliki kebaikan model yang lebih besar dari model regresi linier yaitu sebesar 92.46%.

Kata Kunci: angka literasi, efek spasial, GWR

ABSTRACT

Papua has the lowest literacy index, which is 19.9 and Jakarta is the area with the highest literacy index, reaching 58.16. However, Jakarta with the highest literacy index number occupies a moderate literacy index position because it is still below 60. Papua and Jakarta are two provinces with very striking differences in terms of geography where Jakarta is located on the island of Java and is the capital of the country while Papua is on the easternmost island of Indonesia. Therefore, we need a statistical modeling method that takes into account spatial aspects, namely using the Geographically Weighted Regression (GWR) method which is expected to be able to produce a model of the literacy index number in Indonesia in 2017 in each region. GWR is a spatial method that uses geographical factors as independent variables that can affect the dependent variable. Based on the results of the analysis and discussion using GWR analysis, it is known that in grouping the variables that have a significant effect on the literacy activity index, 11 groups are obtained, where in group 1 there is only one variable that has a significant effect on the literacy activity index, namely the percentage of Latin literacy in Papua

Province. Meanwhile, in group 11 where all independent variables have a significant effect on the literacy activity index, there are Jambi Province, South Sumatra Province, and Lampung Province. The results of the analysis also show that the best modeling of several regression methods is the GWR model because it has a higher model goodness than the linear regression model, which is 92.46%.

Keywords: literacy rate, spatial effect, GWR

1 Pendahuluan

Literasi dapat diartikan melek teknologi dan informasi atau peka terhadap perubahan di lingkungan sekitar. Seseorang yang memperoleh banyak informasi dan memahami suatu hal yang dia baca serta menerapkannya dalam kehidupan disebut dengan literat [1]. Budaya literasi masih menjadi permasalahan utama dalam bidang pendidikan. Tingkat literasi masyarakat Indonesia tergolong begitu memprihatinkan. Terbukti dengan hasil beberapa survei yang menunjukkan bahwa Indonesia termasuk dalam jajaran negara dengan tingkat literasi terendah.

Central Connection State University pada tahun 2016 menyelenggarakan World's Most Literate Nation dan menjelaskan peringkat Indonesia dari 61 negara yang dijelaskan berdasarkan lima aspek. Indonesia menduduki peringkat 60 berdasarkan beberapa aspek diantaranya, kepemilikan komputer di tiap rumah, sistem yang diterapkan dalam sekolah, jumlah jam belajar di sekolah, keberadaan perpustakaan, pergerakan Koran dan Koran online, dan aspek yang terpenting adalah assestment membaca [2]. Selain itu, survei *Programme for Internatioal Student Assestment* [3] pada 2015 menmpublikasikan bahwa berdasarkan pemahaman dan keterampilan penggunaan bahan bacaan untuk teks dokumen pada usia 9 hingga 14 tahun Indonesia menduduki peringkat 10 terbawah yaitu peringkat 64 dari 72 negara [4].

Sebuah penelitian dilakukan oleh Puslitjakdikbud (2017) melalui Indeks Aktivitas Literasi Membaca (Alibaca) menyebutkan aktivitas literasi dipengaruhi oleh empat dimensi, yaitu Dimensi Kecakapan, Dimensi Akses, Dimensi Alternatif, dan Dimensi Budaya. Empat dimensi tersebut melatarbelakangi terbentuknya indikator – indikator yang mampu merepresentasikan tingkat literasi membaca di Indonesia. Berdasarkan penelitian tersebut, provinsi DKI Jakarta menempati posisi tertinggi untuk Indeks Alibaca yaitu sebesar 58,16; sedangkan indeks terendah diperoleh Provinsi Papua sebesar 19,90. Perbedaan kondisi wilayah dan ketersediaan akses diduga menjadi penyebab terjadinya kesenjangan indeks tersebut.

Kondisi literasi masyarakat Papua pernah di rangkum oleh Aleks Giyai seorang Ketua Taman Baca Gerakan Papua Mengajar dalam HARIAN NASIONAL [5] bahwa tidak bisa dipungkiri, saat ini kondisi literasi Papua masih sangat terkebelakang apalagi jika dibandingkan daerah lain. Jika disebut ada anak yang sudah duduk di bangku SMP, tetapi membaca saja masih terbata-bata, itu benar. Saat ini boleh dibilang Papua masih mengalami darurat literasi. Padahal literasi faktor kunci bagi masyarakat untuk bebas dari belenggu kebodohan dan ketertinggalan dengan kota lain di pulau Jawa yang merupakan pusat pertumbuhan ekonomi dengan ketersediaan infrastruktur yang maju. Sehingga dalam penelitian ini, penulis ingin mengidentifikasi berbagai faktor yang diduga mempengaruhi kualitas Alibaca di Indonesia menggunakan metode yang melibatkan tata letak geografis daerah di Indonesia.

Pendekatan dengan metode spasial dianggap beralasan, dengan melihat terjadinya kesenjangan yang jauh antara capaian Indeks Alibaca Provinsi DKI Jakarta dan Papua. GWR merupakan metode spasial dengan melibatkan kondisi geografis masing masing wilayah sebagai salah satu faktor yang diduga mempengaruhi variabel dependen [6]. Dalam penulisan ini, metode GWR akan di aplikasikan untuk mencari tahu variable manakah yang berpengaruh terhadap angka literasi baca di wilayah Indonesia dengan memperhatikan letak daerahnya untuk mengestimasi parameter modelnya sehingga mampu memberikan hasil yang tepat untuk merepresentasikan model dan memetakan Aktivitas Literasi Membaca di tiap wilayah di

Indonesia sehingga diharapkan hasil penelitian ini dapat digunakan sebagai rujukan untuk pemerataan pembangunan di Indonesia.

2 Tinjauan Pustaka

2.1 Analisis Regresi Linier

Analisis regresi linier merupakan suatu metode yang membantu mengetahui pengaruh satu atau lebih variabel dan hubungan diantara variabel. Metode ini terbagi menjadi dua yaitu regresi linier dan berganda dimana luaran yang dihasilkan dari metode ini adalah sebuah model yang dapat menjelaskan pengaruh variabel (Gujarati, 2004). Model pada metode ini secara umum dapat dituliskan seperti pada rumus (1) berikut :

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i \quad (1)$$

Dengan : y_i : Variabel respon pengamatan ke-i

x_{ik} : Koefisien regresi ke-k pengamatan ke-i

β_k : koefisien regresi pada x_k

β_0 : Intersep

ε_i : Residual ke-i

i : unit observasi; $i = 1, 2, \dots, n$

n : jumlah observasi

k : unit variabel prediktor $k = 1, 2, \dots, p$

p : jumlah variabel prediktor

Metode ini hanya melibatkan variabel respon dan prediktor tanpa mempertimbangkan adanya pengaruh lain seperti wilayah atau geografis. Dalam notasi matriks, model regresi dapat dituliskan sebagai berikut. [7]

$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times k}{\mathbf{X}} \underset{k \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\boldsymbol{\varepsilon}}, \quad (2)$$

dengan :

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

2.2 Geographically Weighted Regression (GWR)

Fotheringham et. al. (2002) menjelaskan bahwa model yang dihasilkan dari metode GWR merupakan suatu model yang dikembangkan dari metode regresi linier. GWR mengembangkan dengan menambahkan geografid pada setiap titik lokasi untuk setiap parameter. Pengembangan ini didasarkan pada konsep dari regresi non parametrik yang diterapkan pada model regresi secara umum. Model GWR yang diperoleh akan digunakan untuk memprediksi besarnya variabel respon dengan parameter yang dihasilkan dimana setiap parameter didapat dari lokasi yang menjadi objek. Rumus (4) menjelaskan penulisan secara umum untuk Model GWR [8].

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad (3)$$

dengan : y_i : nilai observasi variabel respon ke-i

x_{ik} : nilai observasi variabel prediktor ke-k di lokasi pengamatan ke-i

(u_i, v_i) : menyatakan titik koordinat letak geografis lokasi pengamatan ke-i

$\beta_0(u_i, v_i)$: konstanta/intersep pada pengamatan ke-i

- $\beta_k(u_i, v_i)$: koefisien regresi variabel prediktor ke-k di lokasi pengamatan ke-i
 ε_i : error pengamatan ke-i diasumsikan identik, prediktor dan berdistribusi normal dengan mean nol dan varian konstan σ^2
 i : unit observasi; $i = 1, 2, \dots, n$
 n : jumlah observasi
 k : unit variabel prediktor $k = 1, 2, \dots, p$
 p : jumlah variabel prediktor

2.3 Uji Moran's I

Moran's I merupakan metode yang digunakan untuk mengidentifikasi suatu wilayah terdapat dependensi spasial atau tidak. Hipotesis yang digunakan adalah sebagai berikut:

H_0 : $\lambda = 0$ (tidak terdapat dependensi spasial)

H_1 : $\lambda \neq 0$ (terdapat dependensi spasial)

statistik uji :

$$Z_{hitung} = \frac{I - I_0}{\sqrt{var(I)}} \quad (4)$$

dengan

$$I = \frac{e'W^*e}{e'e} \quad (5)$$

dimana e merupakan residual regresi OLS dan W^* merupakan matriks pembobot spasial yang sudah di standarisasi. Kriteria penolakan yaitu Tolak H_0 jika $Z(I) > Z_{\alpha/2}$.

2.4 Keragaman Spasial

Keragaman spasial adalah fenomena dalam model spasial. Keberagaman karakteristik suatu lokasi dan letak geografisnya menyebabkan terjadinya keberagaman spasial [9]. Keberagaman lokasi tidak akurat sebelum dilakukan pengujian. Anselin (2013) menyebutkan uji keberagaman lokasi dimana objek yang diamati berupa wilayah dapat diketahui melalui uji keberagaman yang disebut *Breusch-Pagan* (BP). Hipotesisnya adalah sebagai berikut

H_0 : ragam residual (*error*) pada model homogen

H_1 : ragam residual (*error*) pada model tidak homogen

Statistik uji *Breusch-Pagan* (BP) adalah

$$BP = \left(\frac{1}{2} \right) \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \sim X_{(p)}^2, \quad (6)$$

dimana nilai elemen vektor \mathbf{f}^T diperoleh dari $f = \frac{e_i^2}{\sigma^2} - 1$. Nilai e_i^2 merupakan *least square residual* untuk observasi ke-i. \mathbf{Z} adalah matriks yang telah dinormal standarkan dengan ukuran matrik $n \times (p+1)$ disetiap observasi. Pengambilan keputusan didasarkan pada nilai statistik uji yang diperoleh yaitu ketika bila $BP > \chi^2(p)$, maka keputusan yang diambil adalah Tolak H_0 .

2.5 Estimasi Parameter Model GWR dengan Fungsi Kernel

Estimasi parameter pada model GWR menggunakan fungsi kernel untuk mendapatkan nilai pembobot yang dapat mewakili letak antar observasi [6]. Pemilihan pembobot spasial yang digunakan dalam GWR yaitu fungsi *fixed kernel* dan fungsi *adaptive kernel*.

a. Fungsi *Fixed Kernel Gaussian*

$$w_j(u_i, v_i) = \exp \left\{ -\frac{1}{2} \left(\frac{d_{ij}}{h} \right)^2 \right\} \quad (7)$$

b. Fungsi *Adaptive Kernel Gaussian* (8)

$$w_j(u_i, v_i) = \exp \left\{ -\frac{1}{2} \left(\frac{d_{ij}}{h_i} \right)^2 \right\}$$

c. Fungsi *Fixed Kernel Bisquare*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h} \right)^3 \right)^3, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases} \quad (9)$$

d. Fungsi *Adaptive Kernel Bisquare*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h_i} \right)^3 \right)^3, & \text{untuk } d_{ij} \leq h_i \\ 0, & \text{untuk } d_{ij} > h_i \end{cases} \quad (10)$$

dimana $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ merupakan rumus untuk menghitung jarak *euclidia*. Perhitungan jarak ini digunakan untuk mengetahui nilai parameter antara lokasi (u_i, v_i) ke lokasi (u_j, v_j) . Sedangkan h adalah parameter yang disebut parameter penghalus (*bandwidth*).

Apabila sebuah titik berada pada suatu lingkungan dan dianggap memiliki pengaruh pada radius tersebut disebut dengan *bandwidth*. Proses pemodelan data untuk membentuk suatu model GWR dengan pendekatan spasial yang ketepatan data sangat diinginkan dapat dipermudah dengan *bandwidth*. Pada proses pengolahan untuk mendapatkan *bandwidth* yang optimum diperlukan metode pembagian data menggunakan *Cross Validation* (CV) dengan mengatur varians dari model [8]. Secara matematis dapat didefinisikan sebagai berikut:

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2, \quad (11)$$

$\hat{y}_{\neq i}(h)$ merupakan hasil estimasi dari y_i ketika prosesnya menghilangkan pengamatan lokasi (u_i, v_i) . Nilai h yang optimal akan didapatkan dari h yang menghasilkan CV yang paling minimum.

2.6 Pengujian Parameter Model GWR

Pengujian parameter secara serentak dan parsial digunakan untuk mengetahui faktor yang signifikan dalam model yang sesuai. Hipotesis yang digunakan adalah

$$H_0 : \beta_k(u_i, v_i) = \beta_k \text{ untuk setiap } k=0,1,2,\dots,p, \text{ dan } i=1,2,\dots,n$$

(tidak ada perbedaan yang signifikan antara model regresi global dan GWR)

$$H_1 : \text{Paling sedikit ada satu } \beta_k(u_i, v_i) \neq \beta_k, k=0,1,2,\dots,p$$

(ada perbedaan yang signifikan antara model regresi global dan GWR).

Dengan rumus statistik uji yang digunakan sebagai berikut [8]

$$F_{hitung} = \frac{RSS(H_1) / \left(\frac{\delta_1^2}{\delta_2} \right)}{RSS(H_0) / (n-p-1)}, \quad (12)$$

untuk signifikansi parameter pada masing-masing model di setiap wilayah sebagai berikut

$$H_0 : \beta_k(u_i, v_i) = 0$$

$$H_1 : \beta_k(u_i, v_i) \neq 0 \text{ dengan } k=1,2,\dots,p$$

dimana rumus yang sesuai untuk digunakan sebagai berikut

$$t_{hit} = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma} \sqrt{c_{kk}}} \quad (13)$$

3 Metodologi Penelitian

3.1 Variabel Penelitian dan Sumber

Pada penelitian ini digunakan data sekunder dari Pusat Penelitian Kebijakan Pendidikan Dan Kebudayaan oleh Kementerian Pendidikan Dan Kebudayaan 2019. Observasi penelitian adalah provinsi di Indonesia tahun 2017 di sebanyak 34 provinsi. Berikut adalah penjelasan variabel yang disajikan pada Tabel 1 [4].

Tabel 1. Variabel dan Skala Data Penelitian

Indikator	Variabel	Skala
Y	Angka Indeks Aktivitas Literasi	Rasio
X ₁	Persentase Melek Huruf Latin	Rasio
X ₂	Rata-rata Lama Sekolah (RLS) penduduk usia lebih dari 25	Rasio
X ₃	Persentase perpustakaan umum	Rasio
X ₄	Persentase sekolah yang memiliki jaringan akses internet	Rasio
X ₅	Persentase penduduk lebih dari 5 tahun dan dalam 3 Bulan terakhir pernah mengakses internet	Rasio
X ₆	Persentase penduduk lebih dari 5 tahun dan dalam 3 bulan terakhir pernah menggunakan komputer	Rasio

3.2 Teknik Analisis Data

Tahapan analisis data untuk memodelkan angka literasi membaca di 34 provinsi Indonesia adalah

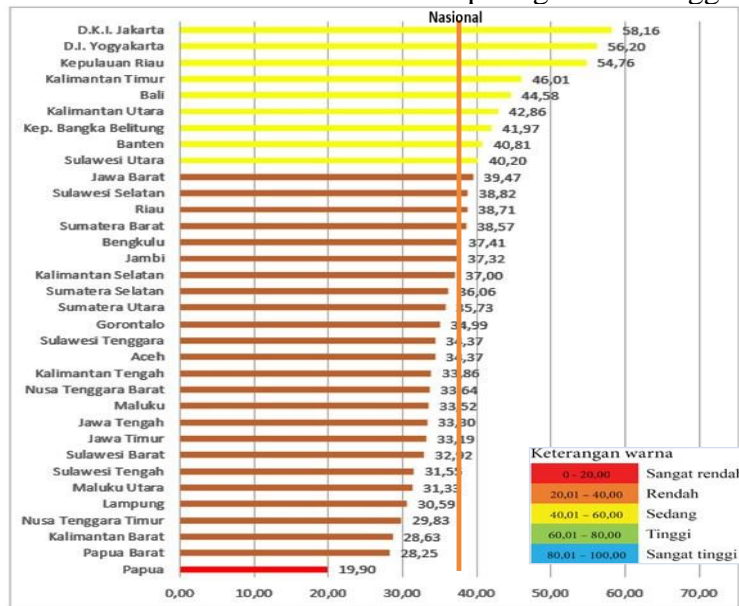
1. Melakukan eksplorasi data secara visual untuk mempermudah menjelaskan informasi dalam data.
2. Memodelkan kasus angka indeks aktivitas literasi menggunakan analisis regresi linier dengan meminimumkan jumlah kuadrat *error* (OLS).
3. Menguji residual dari model pada point 2 sesuai dengan asumsi IIDN.
4. Menguji dependensi spasial menggunakan *moran's I*.
5. Pemodelan Spasial dengan GWR sebagai berikut :
 - a. Untuk mengetahui kejelasan dari efek wilayah, maka digunakan uji *Breush – Pagan* yaitu uji homogenitas pada residual.
 - b. Menentukan nilai *bandwidth* optimum dengan pengulangan hingga mendapatkan nilai minimum.
 - c. Matriks pembobot optimum ditentukan dengan menggunakan fungsi kernel.
 - d. Melakukan estimasi parameter dari hasil nilai *bandwidth* yang optimum.
6. Melakukan perbandingan kedua model dengan kriteria R²
 - a. Uji kesesuaian model untuk melihat apa faktor geografi berpengaruh terhadap angka indeks aktivitas literasi membaca.
 - b. Melakukan uji parameter secara serentak dan parsial.
7. Melakukan visualisasi data dan interpretasi dari model yang sesuai.

4. HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Kondisi Angka Indeks Literasi Indonesia

Gambar 2 berikut merupakan hasil perhitungan indeks provinsi yang menunjukkan 3 provinsi dengan nilai indeks tertinggi, yaitu Provinsi DKI Jakarta dengan angka indeks 58,16; D.I. Yogyakarta dengan angka 56,20; dan Kepulauan Riau dengan angka 54,76. Meskipun

demikian, tiga provinsi tersebut belum mencapai kategori aktivitas literasi tinggi karena indeks ketiganya belum melampaui angka 60,01. Indeks dibawah 60,01 masih berada di level aktivitas literasi sedang. Berikut Indeks Alibaca Provinsi menurut peringkat dari tinggi ke rendah:



Gambar 1. Indeks Alibaca Provinsi menurut Peringkat dari Tinggi ke Rendah

Provinsi Papua berada diperingkat terendah dengan angka indeks 19,90 dan termasuk kategori aktivitas literasi sangat rendah (berada pada rentang angka 0 – 20,00); kemudian Papua Barat dengan angka 28,25 dan Kalimantan Barat dengan angka 28,63 – keduanya termasuk kategori rendah (rentang 20,01 – 40,00).

4.2 Pemodelan Regresi Linier

Hasil dan pembahasan model regresi linier dapat digunakan untuk melihat hubungan antara angka indeks literasi Indonesia dengan faktor-faktor yang diduga mempengaruhi. Variabel independen yang digunakan dalam model regresi tidak saling multikolinieritas. Penelitian ini menggunakan nilai *Variance Inflation Factors (VIF)* sebagai kriteria untuk mengetahui adanya multikolinieritas antar variabel independen. Nilai *VIF* yang lebih besar dari 10 menunjukkan adanya kolinieritas antar variabel independen.

Tabel 2. Nilai VIF Variabel Prediktor dari Faktor yang Diduga Mempengaruhi Literasi Membaca

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Nilai VIF	1.914	3.754	1.872	1.965	6.607	6.354

Tabel 2 menunjukkan antar variabel independen tidak saling berkorelasi, sehingga semua variabel independen yang diduga mempengaruhi angka indeks literasi Indonesia dapat digunakan dalam pembentukan model regresi linier. Pengujian parameter secara serentak dilakukan untuk melihat apakah variabel independen (setidaknya satu) ada yang memberi pengaruh terhadap variable dependen. Pengujian serentak dengan model awal sebagai berikut:

$$y = -8.868 + 0.2626x_1 + 1.563x_2 + 0.006671x_3 - 4.617 \times 10^{-7}x_4 + 0.2747x_5 + 1.253x_6$$

Hipotesis yang digunakan untuk uji serentak adalah

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0$$

Tabel 3. Hasil Uji Serentak Regresi Linear pada Faktor yang mempengaruhi Literasi Membaca

df	F-hitung	F-tabel	Pvalue	R ²
6,27	20.72	2.459	6.142x10 ⁻⁹	0.7819

Tabel 3 menunjukkan hasil pengujian secara serentak model regresi linier dengan estimasi OLS menggunakan taraf signifikansi sebesar 10% keputusan menolak H₀ karena nilai $F_{hitung} = 20.72$ lebih besar dari nilai tabel statistik dengan derajat kebebasan (df) $F_{(0,1;6;27)} = 2,459$ dan *p-value*

$= 6.142 \times 10^{-9} < 0.1$. Sehingga kesimpulan yang diperoleh minimal ada salah satu variable independen yang mempengaruhi angka indeks literasi Indonesia. Nilai koefisien determinasi (R^2) adalah 78.19% yang berarti bahwa model regresi dapat menjelaskan variabilitas angka indeks literasi Indonesia sebesar 78.19% sedangkan sisanya 21.81% dijelaskan oleh variabel lainnya di luar model

4.3 Model Geographically Weighted Regression (GWR)

Model Geographically Weighted Regression (GWR) adalah pengembangan dari model regresi dimana setiap parameter dihitung pada setiap titik lokasi, sehingga setiap titik lokasi geografis mempunyai nilai parameter regresi yang berbeda-beda. Tahapan penyusunan *Model Geographically Weighted Regression (GWR)* dimulai dari uji heteroskedastisitas pada model, bobot optimum, dan pengujian parameter.

a. Uji Heteroskedastisitas

Diagnosis ini untuk mengetahui apakah ada heterogenitas spasial. Hal ini penting dilakukan untuk menentukan tindakan selanjutnya, yaitu menentukan model spasial manakah yang akan digunakan untuk memodelkan angka indeks literasi Indonesia. Uji Heteroskedastisitas dilakukan untuk mengetahui kehomogenan ragam pada residual dengan uji yang digunakan adalah *Breusch-Pagan*. *Model Geographically Weighted Regression (GWR)* dikatakan baik apabila terdapat heteroskedastisitas. Besarnya *p-value* yang diperoleh 0.05885 dimana nilai tersebut lebih kecil dari α ($10\%=0.1$), maka keputusan yang diperoleh menolak hipotesis H_0 . Keputusan ini menunjukkan bahwa ragam residual (*error*) pada model tidak homogen atau ada pengaruh area dalam kasus angka indeks literasi Indonesia. Untuk mengatasi masalah heterogenitas spasial dalam model regresi linier, maka digunakan metode GWR.

b. Moran's I

Pada Tabel 4 dapat dilihat bahwa nilai *moran's I* sebesar -0.034 dan nilai *p-value* sebesar 0.996 dimana lebih kecil dari α ($10\%=0.1$), maka gagal tolak H_0 artinya tidak terdapat dependensi spasial dalam model.

Tabel 4. *Moran's I*

<i>Moran's I</i>	<i>p-value</i>
-0,034	0,996

Berdasarkan hasil *moran's I*, maka dapat dilakukan pengujian *Geographically Weighted Regression* karena tidak terdapat dependensi spasial.

c. Pemilihan Bobot Optimum

Model GWR menggunakan pembobot berdasarkan letak geografis setiap provinsi. Langkah pertama yang dilakukan adalah dengan menentukan letak geografis (*longitude* dan *latitude*) tiap provinsi di Indonesia, Selanjutnya menghitung jarak euclidean berdasarkan letak geografis untuk setiap provinsi di Indonesia. Suatu wilayah dapat ditentukan urutan wilayah-wilayah lain yang berdekatan berdasarkan jarak *euclidian* sehingga akan didapatkan urutan tetangga terdekat untuk seluruh wilayah pengamatan. Nilai jarak *Euclidian* yang nantinya akan menjadi pembobot pada model GWR, dimana kriteria pembobot yang baik yaitu minimum dari nilai CV. Pemilihan pembobot akan berakibat pada model yang dihasilkan yaitu kasar (*under smoothing*) atau sangat halus (*oversmoothing*). Model *under smoothing* akan diperoleh ketika nilai *bandwith* sangat kecil sehingga nilai varians menjadi besar karena berkurangnya jumlah pengamatan pada radius *b*. Hal ini berlaku sebaliknya, ketika nilai *bandwith* semakin besar nilai varians mengecil dikarenakan jumlah pengamatan dalam radius *b* semakin meningkat. Ada tiga fungsi kernel yang dapat membantu dalam pemilihan pembobot yang optimum *Gaussian*, *Bisquare*, dan *Tricube* dimana masing-masing memiliki cabang fungsi *fixed* artinya nilai seluruh lokasi memiliki nilai *bandwith* yang sama dan *adaptive* artinya nilai seluruh lokasi berbeda. Untuk memilih fungsi kernel terbaik, dilakukan pembuatan model pada masing-masing fungsi pembobot sehingga diperoleh nilai *Cross Validation (CV)* minimum. Pada Tabel 5 dapat dilihat hasil CV masing-masing pembobot. Pemilihan pembobot optimum dengan

menggunakan kriteria nilai CV dari tiap-tiap pembobot yang dapat dilihat pada Tabel 4. Pembobot dapat dikatakan optimum apabila nilai CV yang kecil. Tabel 4 menunjukkan bahwa bandwidth *Addaptive Gaussian* merupakan pembobot optimum untuk dilakukan pemodelan GWR karena memiliki nilai CV yang lebih kecil dibanding pembobot lainnya. Hasil estimasi diperoleh nilai bandwidth dan *Cross Validation*, pembobot optimum atau *bandwith* terbaik diketahui dari nilai CV terkecil.

Tabel 5. Pemilihan Pembobot Optimum Melalui Cross Validation

Fungsi Kernel	Cross Validation (CV)	Bandwith
<i>Adaptive Gaussian</i>	1012.107	0,6470676
<i>Fixed Gaussian</i>	1059.839	43,16534
<i>Adaptive Bi-square</i>	1064.614	0,9999339
<i>Fixed Bi-square</i>	427.1299*	18,59176

*Nilai *Bandwith* optimum karena nilai CV paling minimum

d. Pengujian Kesesuaian Model

Uji kesesuaian model digunakan untuk menguji signifikansi dari faktor geografis yang merupakan inti dari model GWR. Pengujian model GWR dilakukan dengan tujuan untuk mengetahui adanya pengaruh faktor lokasi di provinsi di Indonesia terhadap indeks aktivitas literasi. Berikut adalah hipotesis yang digunakan.

$H_0: \beta_k(u_i, v_i) = 0$ (tidak terdapat perbedaan yang antara model regresi dengan model GWR)

$H_1: \beta_k(u_i, v_i) \neq \beta_k; k=1,2,\dots,n$ (terdapat perbedaan yang antara model regresi dengan model GWR)

Tabel 6. Kesesuaian Model

Model	SS	F	P-value
Model OLS	372.925	0.629	0.139
Model GWR	157.526		

Pada taraf signifikansi (α) sebesar 10% keputusan gagal tolak H_0 karena nilai *p-value* sebesar 0.139 dimana lebih besar dari nilai α , sehingga dapat disimpulkan bahwa tidak terdapat perbedaan yang signifikan antara model regresi global dengan model GWR.

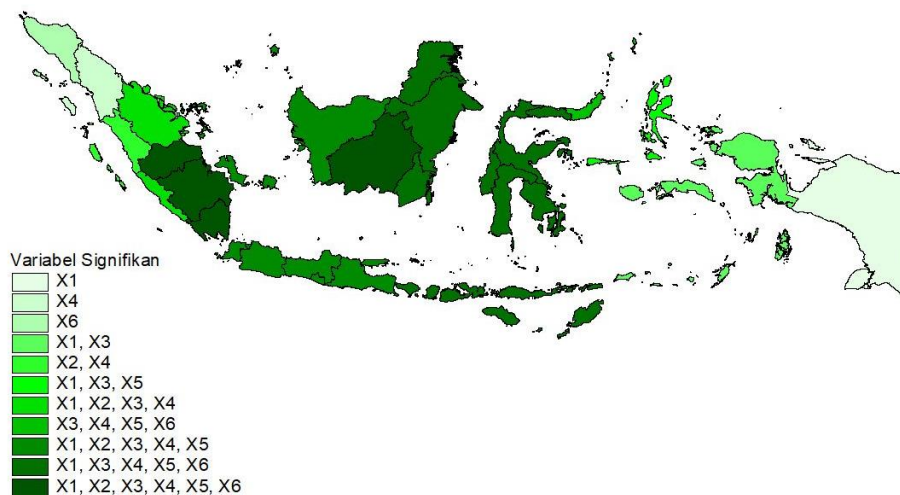
e. Pengelompokkan Variabel Signifikan

Parameter yang dihasilkan pada model GWR bersifat lokal disetiap wilayah tempat data tersebut diamati. Faktor-faktor yang mempengaruhi masalah angka indeks literasi di Indonesia secara spasial sesuai dengan kekurangan yang mendominasi di wilayah tersebut. Indikator pengaruh yang berbeda akan mempengaruhi kebijakan yang diberikan kepada masing-masing wilayah tersebut. Oleh karena itu dilakukan uji parsial untuk setiap parameter disetiap wilayah pengamatan. Berdasarkan uji parameter model GWR secara parsial menggunakan statistik uji t. Pengujian signifikansi parameter model menggunakan hipotesis sebagai berikut.

$H_0: \beta_k(u_i, v_i) = 0$

$H_1: \beta_k(u_i, v_i) \neq \beta_k; k=1,2,\dots,n$

Pada taraf signifikansi (α) 10%, diperoleh nilai $t_{(0,1;28)} = 1,312$. Berikut ini disajikan pengelompokan berdasarkan parameter yang mempengaruhi model. Pada Gambar 2 menunjukkan persebaran variabel-variabel independen yang berpengaruh terhadap indeks aktivitas literasi di setiap Provinsi di Indonesia. Terlihat di setiap provinsi memiliki variabel signifikan yang berbeda.



Gambar 2. Penyebaran Variabel Berpengaruh Signifikan Terhadap Indeks Aktivitas Literasi di Indonesia

Gambar 2 dapat menjelaskan bahwa terdapat 11 kelompok dimana semakin gelap warnanya maka semakin banyak variabel independen yang berpengaruh signifikan terhadap indeks aktivitas literasi di Indonesia. Pada kelompok 1 hanya terdapat satu variabel yang berpengaruh signifikan terhadap indeks aktivitas literasi yaitu persentase melek huruf latin yang terdapat di Provinsi Papua. Sedangkan, pada kelompok 11 dimana semua variabel independen berpengaruh signifikan terhadap indeks aktivitas literasi terdapat di Provinsi Jambi, Provinsi Sumatera Selatan, dan Provinsi Lampung. Berikut variabel-variabel yang berpengaruh terhadap literasi membaca di Indonesia.

Tabel 7. Variabel Berpengaruh Signifikan Terhadap Indeks Aktivitas Literasi di Indonesia

Variabel	Provinsi
X ₁	Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua
X ₂	Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Kalimantan Barat, Kalimantan Tengah
X ₃	Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
X ₄	Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah,

	DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat
X ₅	Jambi, Sumatera Selatan, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara
X ₆	Aceh, Jambi, Sumatera Selatan, Lampung, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat

f. Pembentukan Model GWR

Pemodelan GWR untuk masing-masing Provinsi di Indonesia akan berbeda. Hal ini dapat dilihat dari variabel independen yang berpengaruh terhadap indeks aktivitas literasi di setiap Provinsi di Indonesia berbeda. Salah satu model GWR untuk Provinsi Jambi adalah sebagai berikut:

$$\hat{Y}_5 = 80,398 - 0,690X_{1,5} + 2,077 X_{2,5} + 0,004X_{3,5} - (5,432 \times 10^{-7})X_{4,5} + 0,448X_{5,5} + 0,912X_{6,5}$$

Berdasarkan model di atas dapat diartikan, jika persentase melek huruf latin bertambah sebesar 1% dan variabel lain dianggap tetap, maka indeks aktivitas literasi di Provinsi Jambi mengalami penurunan sebesar 0,69. Jika rata-rata lama sekolah penduduk usia lebih dari 25 tahun bertambah sebesar 1 dan variabel lain dianggap tetap, maka indeks aktivitas literasi di Provinsi Jambi mengalami kenaikan sebesar 2,077. Jika persentase perpustakaan umum bertambah sebesar 1% dan variabel lain dianggap tetap, maka indeks aktivitas literasi di Provinsi Jambi mengalami kenaikan sebesar 0,004. Jika persentase sekolah yang memiliki jaringan akses internet bertambah sebesar 1% dan variabel lain dianggap tetap, maka indeks aktivitas literasi di Provinsi Jambi mengalami penurunan sebesar $5,432 \times 10^{-7}$. Jika persentase penduduk lebih dari 5 tahun dan dalam 3 bulan terakhir pernah mengakses internet bertambah sebesar 1% dan variabel lain dianggap tetap, maka indeks aktivitas literasi di Provinsi Jambi mengalami kenaikan sebesar 0,448. Dan jika persentase penduduk lebih dari 5 tahun dan dalam 3 bulan terakhir pernah menggunakan komputer bertambah sebesar 1% dan variabel lain dianggap tetap, maka indeks aktivitas literasi di Provinsi Jambi mengalami kenaikan sebesar 0,864.

Jika model GWR pada Provinsi Papua adalah sebagai berikut:

$$\hat{Y}_{34} = -19,776 + 0,374X_{1,34} + 1,503X_{2,34} + 0,010X_{3,34} - (5,313 \times 10^{-7})X_{4,34} + 0,249X_{5,34} + 1,269X_{6,34}$$

Berdasarkan model di atas dapat diartikan, jika persentase melek huruf latin bertambah sebesar 1% dan variabel lain dianggap tetap, maka indeks aktivitas literasi di Provinsi Papua mengalami kenaikan sebesar 0,69.

g. Pemilihan Model Terbaik

Model terbaik dipilih dengan melihat besarnya nilai kebaikan model. Kebaikan suatu model dapat dilihat dari nilai R^2 dan AIC. Nilai R^2 tertinggi dan nilai AIC terendah menunjukkan bahwa model tersebut lebih baik dibandingkan model lainnya.

Tabel 8. Pemilihan Model Terbaik

Model	AIC	R ²
Model Regresi	193,918	78,19%
Model GWR	162,338	92,46%

Pada Tabel 6 menunjukkan hasil bahwa tidak terdapat perbedaan yang signifikan antara model regresi dengan model GWR. Namun, berdasarkan Tabel 8 menunjukkan bahwa model GWR memiliki nilai AIC terendah sebesar 162.338 dan R² tertinggi sebesar 92.46% dibandingkan dengan model regresi. Hal ini menunjukkan bahwa model GWR lebih baik digunakan untuk memodelkan indeks aktivitas literasi di Indonesia Tahun 2020.

KESIMPULAN

Hasil regresi linier menunjukkan bahwa terdapat pengaruh signifikan dari faktor-faktor angka indeks literasi di Indonesia. Sementara itu model dengan fungsi pembobot yaitu *fix bisquare* kernel digunakan dalam pemodelan GWR pada kasus angka indeks literasi di Indonesia karena memiliki nilai CV terkecil. Pada pengelompokan variabel berpengaruh signifikan terhadap indeks aktivitas literasi diperoleh 11 kelompok dimana kelompok 1 hanya terdapat satu variabel yang berpengaruh signifikan terhadap indeks aktivitas literasi yaitu persentase melek huruf latin yang terdapat di Provinsi Papua. Sedangkan, pada kelompok 11 dimana semua variabel independen berpengaruh signifikan terhadap indeks aktivitas literasi terdapat di Provinsi Jambi, Provinsi Sumatera Selatan, dan Provinsi Lampung. Hasil analisis juga menunjukkan bahwa pemodelan terbaik dari beberapa metode regresi adalah model GWR karena memiliki kebaikan model yang lebih besar dari model regresi linier yaitu sebesar 92.46%.

Daftar Pustaka

- [1] K. Naibaho, "Menciptakan Generasi Literat Melalui Perpustakaan," *Visi Pustaka*, vol. 9, no. No. 3, p. 4, 2007.
- [2] M. Tarigan, "Terburuknya Peringkat Literasi Kita," *Tempo*, 2016. [Online]. Available: koran.tempo.co/read/397142/terpuruknya-peringkat-literasi-kita.
- [3] OECD, "PISA Indonesia 2015," *Ctry. Note*, pp. 4–5, 2015.
- [4] Puslitjakdikbud, "Indeks Aktivitas Literasi Membaca 34 Provinsi," *repositori.kemdikbud.go.id*, 2019. [Online]. Available: repositori.kemdikbud.go.id.
- [5] E. B. Harsono, "Papua (Masih) Darurat Literasi," *Harian Nasional*, 2019. [Online]. Available: <http://www.harnas.co/2019/09/20/papua-masih-darurat-literasi>. [Accessed: 14-Sep-2020].
- [6] A. S. Fotheringham, C. Brunson, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex, England: John Wiley & Sons, Ltd, 2002.
- [7] J. Fox, "Applied Regression Analysis and Generalized Linear Models," *SAGE Publ. Inc*, 2015.
- [8] J. Lee and D. W. S. Wong, *Statistical Analysis with ArcView GIS*. Canada: John Willey & Sons, Inc, 2001.
- [9] M. Charlton and A. S. Fotheringham, *Geographically Weighted Regression*. White paper, National Centre for Geocomputation National University of Ireland Maynooth, 2009.
- [10] L. Anselin, *Spatial Econometrics: Methods and Models*. Berlin: Springer-Verlag, 2013.