

# Pemodelan Angka Buta Huruf di Provinsi Sumatera Barat Tahun 2014 dengan *Geographically Weighted Regression*

Ratih Maharani dan Wiwiek Setya Winahju

Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

*e-mail:* wiwiek@statistika.its.ac.id

**Abstrak**—Angka Buta Huruf (ABH) merupakan salah satu target dalam *Millennium Development Goals* (MDGs) yaitu menurunkan angka buta huruf menjadi 0%. Faktor yang mempengaruhi ABH berbeda dari suatu wilayah ke wilayah lain sehingga dalam penelitian ini digunakan metode *Geographically Weighted Regression* (GWR) untuk mendapatkan faktor yang berpengaruh tiap kabupaten/kota di Provinsi Sumatera Barat. Berdasarkan hasil pengujian aspek spasial, tidak terdapat dependensi spasial dan tidak terdapat heterogenitas spasial di setiap lokasi pengamatan. Pemilihan pembobot fungsi kernel yang terpilih dengan kriteria *Cross Validation* (CV) minimum adalah fungsi kernel *Gaussian*. Model GWR pada angka buta huruf lebih baik dibandingkan pemodelan dengan regresi linier. Hal ini dikarenakan nilai AIC pada model GWR lebih kecil yaitu sebesar 32,2707 dibandingkan model regresi linier yaitu sebesar 41,6471. Faktor angka partisipasi murni SD, angka partisipasi murni SMP, persentase fasilitas pendidikan SMP dan persentase tenaga pendidik SMP berpengaruh signifikan pada sebagian besar kabupaten/kota di Provinsi Sumatera Barat.

**Kata Kunci**—Angka Buta Huruf, *Geographically Weighted Regression*, Kernel *Gaussian*.

## I. PENDAHULUAN

Salah satu indikator tingkat pendidikan adalah Angka Buta Huruf. Angka Buta Huruf adalah proporsi penduduk usia 15 tahun ke atas yang tidak dapat membaca dan atau menulis huruf latin atau huruf lainnya terhadap penduduk usia 15 tahun ke atas di suatu daerah [1].

Angka buta huruf di Indonesia mengalami penurunan signifikan setiap tahun. Pada tahun 2014 angka buta huruf di Indonesia sebesar 4,88% atau sekitar 6.007.486 orang. Namun, pemerintah Indonesia masih memiliki pekerjaan rumah untuk menuntaskan 6.007.486 penduduk buta huruf. Sedangkan, angka buta huruf di Provinsi Sumatera Barat pada tahun 2010-2014 terjadi penurunan sebesar 1,35%, yaitu 2,91% pada tahun 2010 menjadi 1,56% pada tahun 2014.

Pada penelitian ini dilakukan pemodelan angka buta huruf dan faktor-faktor yang mempengaruhi tiap kabupaten/kota di Provinsi Sumatera Barat. Salah satu metode statistika yang menghubungkan variabel respon dengan variabel prediktor yaitu metode regresi. Metode regresi linier tidak memper- timbangkan aspek lokal yang berbeda-beda antar wilayah. Angka buta huruf tiap kabupaten/kota di Provinsi Sumatera Barat berbeda-beda, dikarenakan kondisi setiap wilayah mempunyai perbedaan karakteristik satu sama lain. Salah satu metode statistik dengan memperhitungkan aspek spasial adalah *Geographically Weighted Regression* (GWR). GWR

adalah pengembangan dari model regresi dimana setiap parameter dihitung pada setiap lokasi, sehingga setiap titik lokasi geografis mempunyai nilai parameter regresi yang berbeda-beda.

Penelitian ini dilakukan pemodelan dan pemetaan angka buta huruf berdasarkan faktor-faktor yang mempengaruhinya tiap Kabupaten/Kota di Provinsi Sumatera Barat menggunakan *Geographically Weighted Regression* (GWR) sehingga penurunan angka buta huruf dapat difokuskan terhadap faktor yang berpengaruh signifikan pada masing-masing kabupaten/ kota di Provinsi Sumatera Barat.

## II. TINJAUAN PUSTAKA

### A. Regresi Linier

Metode regresi linier merupakan metode yang digunakan untuk memodelkan hubungan antara satu variabel respon dan satu atau lebih variabel prediktor. Untuk pengamatan sebanyak  $n$  dengan variabel prediktor ( $x$ ) sebanyak  $p$ , maka model regresi dapat dituliskan dalam bentuk persamaan sebagai berikut [2].

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i; i = 1, 2, \dots, n \quad (1)$$

dimana:

$y_i$  : nilai observasi variabel respon pada pengamatan ke- $i$   
 $x_{ik}$  : nilai observasi variabel prediktor ke- $k$  pengamatan ke-  $i$

$\beta_0$  : nilai *intersep* model regresi

$\beta_k$  : koefisien regresi variabel prediktor ke- $k$

$\varepsilon_i$  : error pada pengamatan ke- $i$  dengan asumsi ( $\varepsilon_i \sim \text{IIDN}(0, \sigma^2)$ ).

Salah satu syarat dalam pembentukan model regresi adalah tidak ada kasus multikolinieritas atau tidak terdapat korelasi antara satu variabel prediktor dengan variabel prediktor yang lain. Salah satu pendeteksian kasus multikolinieritas dilakukan menggunakan kriteria nilai *Varian Inflation Factor* (VIF) [3].

Jika nilai VIF lebih besar dari 10 menunjukkan adanya multikolinieritas antar variabel prediktor. Nilai VIF dinyatakan sebagai berikut.

$$VIF = \frac{1}{1 - R_j^2}; j = 1, 2, \dots, k \quad (2)$$

dengan  $R_j^2$ , adalah koefisien determinasi antara variabel prediktor ke- $j$ .

Penaksiran parameter model regresi dilakukan dengan menggunakan metode *Ordinary Least Square* (OLS)

dengan tujuan untuk meminimumkan jumlah kuadrat error.

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (3)$$

Pengujian signifikan parameter regresi linier terdiri dari uji serentak dan uji parsial. Uji signifikansi secara serentak menggunakan analisis varians (ANOVA) dengan hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0 ; k = 1, 2, \dots, p$$

Statistik uji:

$$F_{hit} = \frac{MSR}{MSE} \quad (4)$$

Tolak  $H_0$  jika  $F_{hit} > F_{\alpha, p, n-(p+1)}$ , yang berarti minimal ada satu parameter yang berpengaruh secara signifikan. Kemudian dilakukan pengujian parameter secara parsial dengan hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \beta_k \neq 0 ; k = 1, 2, \dots, p$$

Statistik uji:

$$t_{hit} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \quad (5)$$

Tolak  $H_0$  jika  $|t_{hit}| > t_{\frac{\alpha}{2}, n-(p+1)}$

Regresi linier mempunyai asumsi yang harus dipenuhi yaitu residual berdistribusi normal, identik, dan independen. Uji asumsi residual berdistribusi normal dapat dilakukan menggunakan uji *Kolmogorov-Smirnov*. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : F(e) = F_0(e)$$

$$H_1 : F(e) \neq F_0(e)$$

Statistik Uji:

$$D = \max |F_n(e) - F_0(e)| \quad (6)$$

Tolak  $H_0$  jika  $|D| > D_{(1-\alpha), n}$  dimana D merupakan nilai berdasarkan tabel *Kolmogorov-Smirnov*.

Apabila varians residual bersifat heteroskedastisitas, berarti residual tidak identik. Salah satu uji yang digunakan adalah uji *glejser* [4]. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2 ; i = 1, 2, \dots, n$$

Statistik uji:

$$F_{hit} = \frac{\left[ \sum_{i=1}^n (\hat{e}_i - |\hat{e}_i|)^2 \right] / p}{\left[ \sum_{i=1}^n (e_i - |\hat{e}_i|)^2 \right] / n - (p+1)} \quad (7)$$

Tolak  $H_0$  jika  $F_{hit} > F_{\alpha, p, n-(p+1)}$  atau  $p\text{-value} < \alpha$ .

Uji asumsi residual independen dilakukan untuk mengetahui adanya korelasi antar residual. Salah satu uji yang dapat digunakan adalah uji *Durbin-Watson*. Hipotesis pengujian yang digunakan adalah sebagai berikut.

$$H_0 : \rho = 0 \text{ (residual independen)}$$

$$H_1 : \rho \neq 0 \text{ (residual tidak independen)}$$

Statistik uji:

$$d = \frac{\sum_{i=2}^n (\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^n \hat{e}_i^2} \quad (8)$$

Tolak  $H_0$  jika  $d \leq d_{L, \alpha/2}$  atau  $d_{L, \alpha/2} \leq (4 - d) \leq d_{U, \alpha/2}$ .

## B. Pengujian Aspek Spasial

Pengujian aspek spasial terdiri dari pengujian heterogenitas spasial dan dependensi spasial. Pengujian heterogenitas spasial dilakukan untuk melihat apakah terdapat kekhasan pada setiap lokasi pengamatan. Sedangkan Pengujian dependensi spasial bertujuan untuk mengetahui apakah pengamatan di suatu lokasi berpengaruh terhadap pengamatan di lokasi lain yang letaknya berdekatan. Pengujian heterogenitas spasial dilakukan menggunakan pengujian *Breusch-Pagan* [5] dengan hipotesis sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{Minimal ada satu } \sigma_i^2 \neq \sigma^2, i = 1, 2, \dots, n$$

Statistik uji:

$$BP = \left( \frac{1}{2} \right) \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \quad (9)$$

dengan elemen vektor f adalah  $f_i = \frac{e_i^2}{\hat{\sigma}^2} - 1$

dimana:

$e_i$  : error dari metode *Ordinary Least Square* (OLS)

$e_i = y_i - \hat{y}_i$ , ( $\hat{y}_i$  diperoleh dari metode OLS)

$\sigma^2$  : varians dari y

$e_i^2$  : kuadrat sisaan untuk pengamatan ke-i

$\mathbf{Z}$  : matriks berukuran  $n \times (p+1)$  yang berisi vektor yang sudah di normal bakukan (z) untuk setiap pengamatan.

Tolak  $H_0$  jika statistik uji BP  $> \chi_{\alpha, p}^2$

Pengujian dependensi spasial dapat dilakukan dengan uji *Moran's I*. Dengan hipotesis sebagai berikut.

$$H_0 : I = 0$$

$$H_1 : I \neq 0$$

Statistik uji:

$$Z_I = \frac{\hat{I} - E(\hat{I})}{\sqrt{\text{var}(\hat{I})}} \quad (10)$$

Tolak  $H_0$  jika, nilai  $|Z_{I \text{ hit}}| > Z_{\alpha/2}$  yang artinya terdapat dependensi spasial.

## C. Geographically Weighted Regression (GWR)

Model *Geographically Weighted Regression* adalah pengembangan dari model regresi dimana setiap parameter dihitung pada setiap lokasi, sehingga setiap titik lokasi geografis mempunyai nilai parameter regresi yang berbeda-beda [6]. Variabel respon (y) merupakan variabel random kontinu. Model GWR adalah sebagai berikut.

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p x_{ik} \beta_k(u_i, v_i) + \varepsilon_i ; i = 1, 2, \dots, n \quad (11)$$

Persamaan estimasi parameter model GWR untuk setiap lokasi adalah.

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} \quad (12)$$

dimana  $\hat{\beta}$  merupakan estimasi dari  $\beta$  dan  $\mathbf{W}(u_i, v_i)$  adalah matrik diagonal pembobot yang elemen diagonalnya menunjukkan pembobot yang bervariasi dari setiap prediksi parameter pada lokasi i.

Penentuan *bandwidth* optimum menggunakan metode *Cross Validation* (CV) adalah sebagai berikut.

$$CV(b) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(b))^2 \quad (13)$$

Untuk mendapatkan nilai  $b$  yang optimal maka diperoleh dari  $b$  yang menghasilkan nilai *Cross Validation* (CV) mini-mum

Pemilihan pembobot spasial digunakan untuk menentukan besarnya pembobot masing-masing lokasi berbeda dimana pembobot yang digunakan adalah dengan fungsi kernel sebagai berikut.

1. Fungsi *Gaussian*

$$w_j(u_i, v_i) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \tag{14}$$

2. Fungsi *Bisquare*

$$w_j(u_i, v_i) = \begin{cases} (1 - (d_{ij}/b)^2)^2, & \text{untuk } d_{ij} \leq b \\ 0, & \text{untuk } d_{ij} > b \end{cases} \tag{15}$$

3. Fungsi *Tricube*

$$w_j(u_i, v_i) = \begin{cases} (1 - (d_{ij}/b)^3)^3, & \text{untuk } d_{ij} \leq b \\ 0, & \text{untuk } d_{ij} > b \end{cases} \tag{16}$$

dengan  $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$  adalah jarak *eucliden* antara lokasi  $(u_i, v_i)$  ke lokasi  $(u_j, v_j)$  dan  $b$  adalah nilai *bandwidth* optimum pada tiap lokasi.

Pengujian model GWR terdiri dari pengujian kesesuaian model GWR dan pengujian signifikansi parameter model GWR. Pengujian kesesuaian model dilakukan dengan hipotesis sebagai berikut.

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

$$H_1: \text{Minimal ada satu } \beta_k(u_i, v_i) \neq \beta_k$$

Statistik uji:

$$F_{hit} = \frac{SSE(H_1)/df_2}{SSE(H_0)/df_1} \tag{17}$$

Daerah penolakan: Tolak  $H_0$  jika  $F_{hitung} < F_{(\alpha, df_1, df_2)}$

Pengujian parameter model GWR dilakukan untuk mengetahui parameter mana yang berpengaruh secara signifikan terhadap variabel respon secara parsial. Hipotesis yang digunakan adalah sebagai berikut.

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

$$H_1: \beta_k(u_i, v_i) \neq 0; i = 1, 2, \dots, n; k = 1, 2, \dots, p$$

Statistik uji:

$$t_{hit} = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma} \sqrt{c_{kk}}} \tag{18}$$

Tolak  $H_0$  jika  $|t_{hit}| > t_{\alpha/2, (\delta_1^2/\delta_2)}$  yang artinya parameter variabel prediktor berpengaruh signifikan terhadap variabel respon.

D. *Pemilihan Model Terbaik*

Kriteria pemilihan model terbaik dapat dilakukan dengan melihat nilai *Akaike's Information Criterion* (AIC). AIC dapat digunakan sebagai kriteria pemilihan model terbaik dimana parameter  $\hat{\beta}$  merupakan estimator maksimum likelihood. Model terbaik adalah model regresi yang mempunyai nilai AIC terkecil [6].

$$AIC = 2n \log_{\epsilon}(\hat{\sigma}) + n \log_{\epsilon}(2\pi) + n + tr(\mathbf{S}) \tag{19}$$

E. *Angka Buta Huruf*

Angka Buta Huruf (ABH) adalah proporsi penduduk usia 15 tahun ke atas yang tidak dapat membaca dan atau menulis huruf latin atau huruf lainnya terhadap penduduk usia 15 tahun ke atas di suatu daerah. Angka buta huruf dapat dihitung melalui rumus sebagai berikut.

$$ABH = \frac{\text{Jumlah penduduk buta huruf usia 15+ tahun}}{\text{Jumlah penduduk usia 15+ tahun}} \times 100\%$$

Angka buta huruf digunakan untuk mengukur keberhasilan program pemberantasan buta huruf, terutama di daerah pedesaan di Indonesia dimana masih tinggi jumlah penduduk yang tidak pernah bersekolah atau tidak tamat SD. Selain itu, juga digunakan untuk menunjukkan kemampuan penduduk dalam menyerap informasi serta menunjukkan kemampuan untuk berkomunikasi secara lisan dan tertulis.

III. METODOLOGI PENELITIAN

A. *Sumber Data*

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik yaitu data Provinsi Sumatera Barat Dalam Angka Tahun 2015 berupa data angka buta huruf dan faktor-faktor yang mempengaruhi tiap kabupaten/kota di Provinsi Sumatera Barat tahun 2014. Dalam penelitian ini juga digunakan data letak astronomi yaitu letak lintang dan letak bujur tiap kabupaten/ kota di Provinsi Sumatera Barat sebagai faktor pembobot geografis. Unit observasi yang digunakan adalah 19 kabupaten /kota di Provinsi Sumatera Barat.

B. *Variabel Penelitian*

Variabel yang digunakan dalam penelitian ini sebagai berikut.

1. Angka Buta Huruf tiap kabupaten/kota di Sumatera Barat (Y)
2. Persentase daerah berstatus kota (X<sub>1</sub>)
3. Persentase penduduk miskin (X<sub>2</sub>)
4. Angka Partisipasi Murni SD (X<sub>3</sub>)
5. Angka Partisipasi Murni SMP (X<sub>4</sub>)
6. Persentase fasilitas pendidikan SD (X<sub>5</sub>)
7. Persentase fasilitas pendidikan SMP (X<sub>6</sub>)
8. Persentase tenaga pendidik SD (X<sub>7</sub>)
9. Persentase tenaga pendidik SMP (X<sub>8</sub>)
10. Lintang (*longitude*) kabupaten/kota ke- $i$  ( $u_i$ )
11. Bujur (*latitude*) kabupaten/kota ke- $i$  ( $v_i$ )

C. *Langkah Analisis*

Langkah analisis untuk menyelesaikan permasalahan ada-lah sebagai berikut.

1. Mendeskripsikan variabel angka buta huruf dan faktor-faktor yang mempengaruhi tiap kabupaten/kota di Provinsi Sumatera dengan menggunakan peta tematik.
2. Pengujian multikolinieritas antar variabel prediktor dengan menggunakan kriteria nilai VIF.
3. Mengidentifikasi pola hubungan antara variabel respon dengan masing-masing variabel prediktor
4. Mendapatkan model regresi linier.
5. Pengujian aspek spasial pada data.
6. Melakukan pemodelan dengan menggunakan GWR.

IV. ANALISIS DAN PEMBAHASAN

A. *Karakteristik Angka Buta Huruf di Provinsi Sumatera Barat Tahun 2014*

Kabupaten/kota di Sumatera Barat yang mempunyai ABH tertinggi pada tahun 2014 adalah Kabupaten Sijunjung sebesar 4,68% yang berarti bahwa dari 100 penduduk terdapat 5 penduduk yang buta huruf. ABH terendah adalah Kota Sawahlunto sebesar 0%. Berikut

penyebaran angka buta huruf tiap Kabupaten/Kota di Provinsi Sumatera Barat.



**Gambar 1.** Penyebaran angka buta huruf di Provinsi Sumatera Barat

Gambar 1 menunjukkan bahwa wilayah dengan ABH tertinggi (2,51-4,68 persen) terdapat pada wilayah Kep. Mentawai dan Kabupaten Sijunjung. Hal ini diduga angka putus sekolah di daerah tersebut cenderung lebih tinggi dibandingkan kabupaten/kota lainnya di Sumatera Barat dan kinerja pemerintah dalam program pemberantasan buta huruf pada daerah tersebut belum efektif. Wilayah dengan ABH terendah (0-0,58 persen) yaitu Kabupaten Agam, Pasaman, Lima Puluh Kota, Kota Padang, Sawahlunto, Bukittinggi, Payakumbuh, Solok, dan Pariaman. Hal ini menunjukkan bahwa makin meningkat kesadaran orang tua akan pentingnya pendidikan.

**B. Pengujian Multikolinieritas**

Berikut adalah salah satu cara mendeteksi multikolinieritas dengan kriteria nilai VIF.

**TABEL 1.** NILAI VIF SETIAP VARIABEL PREDIKTOR

Variabel	Nilai VIF	Variabel	Nilai VIF
X <sub>1</sub>	5,4939	X <sub>5</sub>	<b>62,4990</b>
X <sub>2</sub>	2,5002	X <sub>6</sub>	<b>37,8059</b>
Variabel	Nilai VIF	Variabel	Nilai VIF
X <sub>3</sub>	2,3592	X <sub>7</sub>	<b>138,800</b>
X <sub>4</sub>	3,2454	X <sub>8</sub>	5,7931

Tabel 1 menunjukkan nilai VIF dari masing-masing variabel prediktor X<sub>1</sub> hingga X<sub>8</sub>. Terdapat 3 variabel prediktor dengan nilai VIF lebih besar dari 10 yaitu variabel X<sub>5</sub>, X<sub>6</sub>, dan X<sub>7</sub>. Hal ini mengindikasikan bahwa variabel X<sub>5</sub>, X<sub>6</sub>, dan X<sub>7</sub> mempunyai korelasi dengan variabel prediktor yang lain.

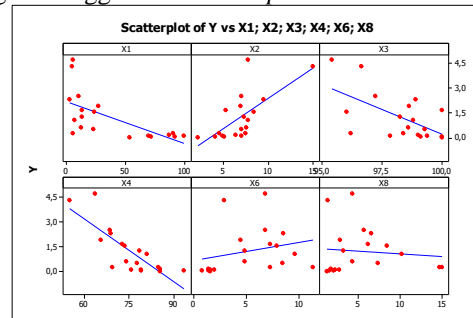
Untuk mengatasi hal tersebut maka variabel yang mempunyai nilai VIF yang paling besar dari variabel yang lain dikeluarkan secara bertahap. Berdasarkan kriteria nilai VIF, maka variabel yang dieliminasi adalah variabel X<sub>7</sub> yaitu persentase tenaga pendidikan SD. Setelah variabel X<sub>7</sub> dikeluarkan maka dilakukan pengujian nilai VIF kembali terhadap variabel yang terpilih untuk memastikan bahwa tidak ada multikolinieritas antar variabel yang terpilih. Setelah mengeluarkan variabel X<sub>7</sub> masih terdapat 2 variabel prediktor dengan nilai VIF yang lebih dari 10 yaitu variabel X<sub>5</sub> dan X<sub>6</sub>. Dari kedua variabel

yaitu variabel X<sub>5</sub> dan X<sub>6</sub>, variabel X<sub>5</sub> yaitu variabel persentase fasilitas pendidikan SD yang memiliki nilai VIF paling tinggi. Sehingga, variabel X<sub>5</sub> dikeluarkan dari variabel penelitian.

Setelah variabel X<sub>5</sub> dikeluarkan maka tidak ada nilai VIF yang lebih dari 10. Jadi dapat dikatakan bahwa tidak terjadi kasus multikolinieritas antara variabel prediktor setelah mengeluarkan variabel X<sub>5</sub> dan X<sub>7</sub>. Sehingga variabel terpilih yaitu variabel X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>6</sub> dan X<sub>8</sub> dapat digunakan pada pemodelan angka buta huruf dengan menggunakan metode analisis regresi linier berganda.

**C. Identifikasi Pola Hubungan Antar Variabel**

Untuk mengetahui pola hubungan antara angka buta huruf dan faktor-faktor yang mempengaruhi dapat dilihat dengan menggunakan *scatterplot*.



**Gambar 2.** Pola hubungan antar angka buta huruf dan faktor-faktor yang mempengaruhi

Berdasarkan Gambar 2 dapat dijelaskan bahwa empat variabel prediktor yaitu persentase daerah berstatus kota (X<sub>1</sub>), angka partisipasi murni SD (X<sub>3</sub>), angka partisipasi murni SMP (X<sub>4</sub>) dan persentase tenaga pendidikan SMP (X<sub>8</sub>) berkorelasi negatif terhadap angka buta huruf. Hal ini berarti jika terjadi peningkatan pada variabel tersebut, maka akan menurunkan terjadinya angka buta huruf. Variabel persentase penduduk miskin (X<sub>2</sub>) dan persentase fasilitas pendidikan SMP (X<sub>6</sub>) memiliki hubungan korelasi positif dengan terjadinya angka buta huruf, sehingga apabila terjadi peningkatan pada variabel tersebut akan meningkatkan terjadinya angka buta huruf.

**D. Pemodelan Regresi Linier Angka Buta Huruf**

Pemodelan angka buta huruf adalah sebagai berikut  $\hat{Y} = 43,0973 - 0,0102X_1 + 0,0323X_2 - 0,3839X_3 - 0,0509X_4 + 0,1913X_6 - 0,2166X_8$

Persentase angka buta huruf akan mengalami penurunan sebesar 0,0102 jika persentase daerah berstatus kota meningkat sebesar satu persen dengan syarat variabel prediktor yang lain konstan. Sebaliknya persentase angka buta huruf akan mengalami kenaikan sebesar 0,0323 jika persentase penduduk miskin meningkat sebesar satu persen dengan syarat variabel prediktor yang lain konstan.

Pengujian signifikansi parameter secara serentak pada pemodelan regresi linier diperoleh nilai F<sub>hitung</sub> sebesar 14,62 dan p-value sebesar 0,000 dengan F<sub>(0,1;6;12)</sub> sebesar 2,3310 dan taraf signifikansi (α) sebesar 0,1 sehingga diputuskan Tolak H<sub>0</sub> karena F<sub>hitung</sub> > F<sub>tabel</sub> dan p-value < α. Hal ini berarti bahwa minimal ada satu parameter yang berpengaruh signifikan secara serentak terhadap model.

Pengujian parameter secara parsial diperoleh t<sub>(0,05;12)</sub> = 1,782 menunjukkan bahwa variabel angka partisipasi murni SD (X<sub>3</sub>), angka partisipasi murni SMP (X<sub>4</sub>), persentase fasilitas pendidikan SMP (X<sub>6</sub>), dan persentase tenaga pendidikan SMP (X<sub>8</sub>) signifikan terhadap model secara individu.

**TABEL 2. ESTIMASI PARAMETER MODEL REGRESI LINIER BERGANDA**

Parameter	Estimasi	T	P-value	Keputusan
$\beta_0$	43,0973	3,126	0,0088	Tolak $H_0$
$\beta_1$	-0,0102	-1,384	0,1917	Gagal Tolak $H_0$
$\beta_2$	0,03253	0,384	0,7077	Gagal Tolak $H_0$
$\beta_3$	-0,3839	-2,665	0,0206	Tolak $H_0$
$\beta_4$	-0,0509	-1,827	0,0927	Tolak $H_0$
$\beta_6$	0,1913	1,832	0,0919	Tolak $H_0$
$\beta_8$	-0,2166	-3,255	0,0069	Tolak $H_0$

Asumsi residual berdistribusi normal, identik, dan independen harus dipenuhi dalam analisis regresi linier. Pengujian asumsi distribusi normal dilakukan dengan menggunakan uji *Kolmogorov-Smirnov* diperoleh nilai  $D(0,132) < D_{(0,9;19)}(0,271)$  dan  $p\text{-value} > 0,150$  sehingga keputusannya adalah gagal tolak  $H_0$ . Hal ini berarti asumsi residual berdistribusi normal telah terpenuhi. Pengujian asumsi residual identik menggunakan uji *Glejser* dengan meregresikan nilai mutlak dari residual dengan semua variabel prediktor didapatkan  $p\text{-value} < \alpha$  maka asumsi residual identik terpenuhi. Pengujian asumsi residual independen menggunakan *Durbin-Watson* diperoleh  $d$  sebesar 1,3763 dengan nilai  $d_L=0,649$ . Sehingga diputuskan untuk gagal tolak  $H_0$  karena nilai  $d > d_L$ . Hal ini berarti tidak terjadi korelasi antar residual atau asumsi residual independen terpenuhi.

**E. Pengujian Aspek Spasial**

Pengujian *Breusch-Pagan* digunakan untuk melihat heterogenitas spasial setiap lokasi, sedangkan pengujian *Moran's I* digunakan untuk pengujian dependensi spasial.

Hasil pengujian *Breusch-Pagan* dengan menggunakan taraf signifikansi ( $\alpha$ ) sebesar 10% menunjukkan  $p\text{-value} > \alpha$  sehingga diputuskan gagal tolak  $H_0$  artinya tidak terdapat perbedaan karakteristik antara satu titik pengamatan dengan titik pengamatan yang lainnya. Sedangkan untuk uji *Moran's I* menunjukkan bahwa  $p\text{-value} > \alpha$  sehingga diputuskan gagal tolak  $H_0$  artinya tidak terdapat dependensi spasial antar kabupaten/kota.

**TABEL 3. PENGUJIAN ASPEK SPASIAL**

Pengujian	Nilai Signifikansi	Keputusan
Breusch-Pagan	0,6508	Gagal Tolak $H_0$
Moran's I	0,8690	Gagal Tolak $H_0$

**F. Pemodelan Angka Buta Huruf dengan Geographically Weighted Regression (GWR)**

Menentukan nilai *bandwidth* (b) dengan menggunakan kriteria nilai *Cross Validation* (CV) minimum. Pada Tabel 4. diperoleh bahwa nilai CV minimum yaitu sebesar 15,6711 terdapat pada pembobot fungsi kernel *Fixed Gaussian* dengan nilai *bandwidth* sebesar 2,6181.

**TABEL 4. CROSS VALIDATION (CV) DAN BANDWIDTH**

Model	CV	b
Gaussian*	15,6711	2,6181
Bisquare	26,2834	1,5539
Tricube	27,7256	1,5903

Keterangan: \*) Pembobot terbaik

Matriks pembobot spasial yang diperoleh untuk tiap-tiap lokasi kemudian digunakan untuk membentuk model sehingga tiap-tiap lokasi memiliki model yang berbeda.

**TABEL 5. ESTIMASI PARAMETER MODEL GWR**

Parameter	Min	Max
$\beta_0$	42,2100	43,7700
$\beta_1$	-0,0108	-0,0099
$\beta_2$	0,0215	0,0415
$\beta_3$	-0,3900	-0,3762

$\beta_4$	-0,0511	-0,0495
$\beta_6$	0,1891	0,1951
$\beta_8$	-0,2187	-0,2166

Pengujian kesesuaian model GWR diharapkan lebih baik jika dibandingkan dengan pemodelan menggunakan regresi linier. Hipotesis yang digunakan adalah sebagai berikut.

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

$$H_1 : \beta_k(u_i, v_i) \neq \beta_k$$

Berdasarkan persamaan (17) diperoleh nilai  $F_{hit}$  sebesar 1,0283 dan  $p\text{-value}$  sebesar 0,483 dengan  $F_{tabel}$  sebesar 2,1660 dan taraf signifikansi ( $\alpha$ ) sebesar 0,1 maka diputuskan gagal tolak  $H_0$ , karena  $F_{hit} < F_{tabel}$  dan  $p\text{-value} > \alpha$  Hal ini berarti bahwa terdapat perbedaan yang signifikan antara model regresi linier dengan model GWR.

Pengujian signifikansi parameter secara parsial dilakukan untuk mengetahui variabel yang signifikan di setiap lokasi Hipotesis yang digunakan adalah sebagai berikut.

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

$$H_1 : \beta_k(u_i, v_i) \neq 0$$

**TABEL 6. PENGELOMPOKKAN KABUPATEN/KOTA**

Kabupaten/Kota	Variabel Signifikan
Kabupaten Agam, Lima Puluh Kota, pasaman, Pasaman Barat, dan Kota Bukittinggi	$X_3, X_6, X_8$
Kabupaten/Kota Kep. Mentawai, Kabupaten Pesisir Selatan, Tanah Datar, Solok, Sijunjung, Tanah Datar, Padang Pariaman, Solok Selatan, Dharmasraya, Kota Padang, Sawahlunto, Solok, Padang Panjang, Payakumbuh, dan Pariaman	$X_3, X_4, X_6, X_8$

Model GWR Kep. Mentawai adalah sebagai berikut.

$$\hat{Y} = 42,2059 - 0,3762 X_3 - 0,0500 X_4 + 0,1901 X_6 - 0,2167 X_8$$

Persentase ABH akan mengalami penurunan sebesar 0,3762 jika angka partisipasi murni SD meningkat sebesar satu persen dengan syarat variabel prediktor yang lain konstan. Berlaku sama pula untuk setiap variabel dalam model GWR.

Pengelompokan kabupaten/kota yang memiliki kesamaan variabel yang berpengaruh terhadap angka buta huruf dapat dilihat dari Gambar 2.



**Gambar 3.** Persebaran variabel signifikan menurut kabupaten/kota di Provinsi Sumatera Barat

Gambar 3 menunjukkan beberapa wilayah yang berdekatan memiliki kesamaan variabel yang signifikan.

#### G. Pemilihan Model Terbaik

Untuk mengetahui model mana yang lebih sesuai untuk menggambarkan angka buta huruf di Provinsi Sumatera Barat, maka dilakukan pemilihan model terbaik antara model regresi linier dan GWR dengan kriteria AIC. Berikut adalah hasil nilai AIC dari masing-masing model. Model yang terbaik adalah menghasilkan nilai AIC terkecil.

Model	AIC
Regresi Linier	41,6471
GWR	32,3707

Tabel 7 memberikan kesimpulan bahwa model sesuai untuk pemodelan angka buta huruf di Provinsi Sumatera Barat adalah model GWR, karena nilai AIC yang dihasilkan lebih kecil dibandingkan model regresi linier.

#### V. KESIMPULAN DAN SARAN

Berdasarkan hasil regresi faktor yang berpengaruh signifikan terhadap angka buta huruf adalah variabel angka partisipasi murni SD ( $X_3$ ), angka partisipasi murni SMP ( $X_4$ ), persentase fasilitas pendidikan SMP ( $X_6$ ), dan persentase tenaga pendidikan SMP ( $X_8$ ). Berdasarkan hasil pengujian aspek spasial, tidak terdapat dependensi spasial dan tidak terdapat heterogenitas spasial. Pemilihan pembobot fungsi kernel yang terpilih dengan kriteria *Cross Validation* (CV) minimum adalah fungsi kernel *Gaussian*. Variabel yang signifikan di tiap kabupaten/kota mengelompokkan Provinsi Sumatera Barat ke dalam dua kelompok. Model *Geographically Weighted Regression* (GWR) pada angka buta huruf lebih baik dibandingkan

pemodelan dengan regresi linier. Hal ini dikarenakan nilai AIC yang didapatkan dengan pemodelan GWR lebih kecil yaitu sebesar 32,2707 dibandingkan model regresi linier yaitu sebesar 41,6471.

Pemodelan angka buta huruf dengan GWR diperoleh hubungan variabel prediktor dan respon yang tidak sesuai dengan kenyataan. Oleh karena itu, disarankan untuk peneliti selanjutnya, agar dilakukan kajian yang lebih komprehensif berkaitan dengan faktor-faktor yang mempengaruhi angka buta huruf di kabupaten/kota di Provinsi Sumatera Barat serta menggunakan pembobot fungsi kernel *adaptif* untuk pemodelan GWR. Selain itu, untuk memodelkan angka buta huruf dengan metode lain untuk data yang tidak memenuhi dependensi spasial dan heterogenitas spasial karena hasil analisis belum cukup hanya menggunakan metode tersebut sehingga dapat digunakan sebagai pembandingan dalam menentukan model yang sesuai.

#### DAFTAR PUSTAKA

- [1] BPS RI. (2015). *Persentase Penduduk Buta Huruf di Indonesia Tahun 2014*. Badan Pusat Statistika.
- [2] Draper, N. R., & Smith, H. (1992). *Applied Regression Analysis Second Edition*. New York: John Wiley & Sons, Inc.
- [3] Hocking, R. R. (1996). *Method and Applications of Linier Models*. New York: John Wiley and Sons Inc.
- [4] Gujarati, D.N. (2003). *Basic Econometric Fourth Edition*. New York: Mc Graw Hill Companies.
- [5] Anselin, L. (1998). *Spatial Econometric: Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- [6] Fotheringham, A.S., Brundson C., & Charlton, M.E. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*. England: John Wiley and Sons Ltd.