

# Pemodelan Faktor-Faktor Yang Mempengaruhi Tingkat Buta Huruf Kabupaten/kota di Jawa Timur dengan *Geographically Weighted Ordinal Logistic Regression*

Nur Lailiyah dan Purnadi

Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember

Jl. Arief Rahman Hakim, Surabaya 60111

*E-mail:* purhadi@statistika.its.ac.id

**Abstrak**— Salah satu indikator kualitas hidup suatu negara dapat ditentukan dengan Indeks Pembangunan Manusia (IPM) yang dilihat berdasarkan harapan hidup, melek huruf, pendidikan, dan standar hidup. Tingginya buta huruf terutama di Jawa Timur menjadi hambatan dalam dunia pendidikan sehingga dapat mempengaruhi rendahnya IPM di Indonesia terutama di Jawa Timur. Pada penelitian ini dibahas mengenai pemodelan faktor-faktor yang mempengaruhi tingkat buta huruf kabupaten/kota di Jawa Timur dengan *Geographically Weighted Ordinal Logistic Regression* (GWOLR). GWOLR merupakan pengembangan dari model regresi logistik ordinal yang memperhatikan faktor geografis. Penelitian ini bertujuan untuk memodelkan faktor-faktor yang mempengaruhi tingkat buta huruf kabupaten/kota di Jawa Timur berdasarkan model GWOLR. Hasil penelitian ini adalah pemodelan faktor-faktor yang mempengaruhi tingkat buta huruf dengan GWOLR lebih baik dibandingkan regresi logistik ordinal berdasarkan *Akaike Information Criterion* (AIC). Faktor yang mempengaruhi tingkat buta huruf tiap kabupaten/kota di Jawa Timur berdasarkan model GWOLR adalah persentase daerah berstatus kota dan angka partisipasi murni Sekolah Dasar tiap kabupaten/kota di Jawa Timur. Kedua faktor tersebut berpengaruh positif terhadap model. Model GWOLR terbaik pada penelitian ini yaitu dengan menggunakan pembobot fungsi kernel *Exponential*.

**Kata Kunci**—AIC, Buta Huruf, Fungsi Kernel Exponential, GWOLR, Regresi Logistik Ordinal.

## I. PENDAHULUAN

INDEKS Pembangunan Manusia (IPM) adalah pengukuran perbandingan dari harapan hidup, melek huruf, pendidikan dan standar hidup untuk semua negara seluruh dunia [1]. Tingginya IPM menunjukkan keberhasilan pembangunan kesehatan, kependudukan, pendidikan dan ekonomi di suatu negara. Tingginya pendidikan adalah salah satu penunjang keberhasilan pembangunan manusia, namun banyak hal di Indonesia yang menjadi hambatan dalam mencapai keberhasilan pendidikan. Salah satu hambatan tersebut adalah angka buta huruf yang masih tinggi. Jawa Timur merupakan provinsi yang dikenal memiliki perekonomian yang tinggi tetapi dalam hal pendidikan masih dikatakan rendah, termasuk masih tingginya angka buta huruf. Jawa Timur adalah penyumbang buta aksara paling besar di Indonesia dari pada

provinsi lainya. Sebanyak 1,2 juta jiwa warga Jawa Timur tercatat mengalami buta huruf [2]. Dengan tingginya angka buta huruf di Jawa Timur, maka ingin dilakukan penelitian mengenai faktor-faktor yang mempengaruhi tingkat buta huruf kabupaten/kota di Jawa Timur. Terdapat beberapa penelitian yang pernah dilakukan mengenai buta huruf [3]-[5].

Untuk menganalisis faktor-faktor yang mempengaruhi tingkat buta huruf dapat dilakukan dengan menggunakan analisis regresi logistik ordinal. Namun, analisis regresi logistik ordinal tidak memperhatikan faktor geografis tiap pengamatan. Faktor geografis diperlukan karena tiap lokasi diduga mempunyai karakteristik yang berbeda sehingga diperlukan analisis yang memperhatikan faktor geografis, yaitu *Geographically Weighted Ordinal Logistic Regression* (GWOLR). Atkinson [6] telah melakukan penelitian mengenai *Geographically Weighted Logistic Regression* (GWLR) dan Rifada [7] telah melakukan penelitian mengenai GWOLR. Mengacu pada penelitian sebelumnya maka penelitian ini membahas tentang pemodelan faktor-faktor yang mempengaruhi tingkat buta huruf kabupaten/kota di Jawa Timur berdasarkan model GWOLR.

## II. LANDASAN TEORI

### A. Model Regresi Logistik Ordinal

Regresi logistik ordinal merupakan salah satu metode statistika untuk menganalisis hubungan antara variabel respon yang mempunyai skala ordinal dengan variabel prediktor yang bersifat kategori dan/atau kontinu. Model untuk regresi logistik ordinal adalah *cumulative logit models*. Jika variabel respon terdiri dari 3 kategori ( $J=3$ ), maka model regresi logistik (logit) ordinal atau logit kumulatif adalah:

$$\text{Logit}[P(Y_i \leq j | x_i)] = \beta_{0j} + x_i^T \beta ; j=1, 2, \dots, J-1 \text{ dan } i=1, 2, \dots, n \quad (1)$$

Parameter  $\beta_{0j}$  merupakan intersep yang tidak diketahui yang memenuhi kondisi  $\beta_{01} \leq \beta_{02} \leq \dots \leq \beta_{0, J-1}$  dan  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$  merupakan vektor koefisien regresi yang tidak diketahui yang bersesuaian dengan  $x_i$ . Peluang masing-masing  $J$  kategori respon adalah:

$$\gamma_j(\mathbf{x}_i) = \frac{\exp(\beta_{0j} + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\beta_{0j} + \mathbf{x}_i^T \boldsymbol{\beta})} - \frac{\exp(\beta_{0j-1} + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\beta_{0j-1} + \mathbf{x}_i^T \boldsymbol{\beta})}, \quad j = 1, 2, 3 \quad (2)$$

Penaksiran parameter regresi logistik ordinal dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Estimasi parameter melalui metode MLE adalah dengan melakukan turunan parsial fungsi *ln-likelihood* terhadap parameter yang akan diestimasi kemudian disamadengankan dengan nol. Turunan parsial pertama dari fungsi *ln-likelihood* terhadap parameter yang akan diestimasi merupakan fungsi yang implisit sehingga digunakan iterasi *Newton Raphson*.

Setelah dilakukan penaksiran parameter model regresi logistik ordinal maka dilakukan pengujian secara serentak dan parsial. Uji serentak dilakukan untuk mengetahui signifikansi parameter  $\beta$  terhadap variabel respon secara bersama-sama [8].

Hipotesis:

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

$H_1$  : paling tidak terdapat satu  $\beta_k \neq 0$ ;  $k = 1, 2, \dots, p$

Statistik uji:

$$G^2 = -2 \ln \left[ \frac{L(\hat{\omega})}{L(\hat{\Omega})} \right] = -2(\ln L(\hat{\omega}) - \ln L(\hat{\Omega})) \quad (3)$$

Dengan  $L(\hat{\omega})$  adalah nilai maksimum *likelihood* di bawah  $H_0$  dan  $L(\hat{\Omega})$  adalah nilai maksimum *likelihood* di bawah populasi. Tolak  $H_0$  jika  $G^2 > \chi^2_{(\alpha, p)}$  di mana  $p$  menunjukkan banyaknya variabel prediktor dalam model atau tolak  $H_0$  jika  $p\text{-value} < \alpha$ . Uji parsial digunakan untuk mengetahui signifikansi parameter  $\beta$  terhadap variabel respon secara parsial dengan hipotesis sebagai berikut [8].

$$H_0 : \beta_k = 0$$

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

Statistik uji yang digunakan adalah:

$$W = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \quad (4)$$

dengan  $\hat{\beta}_k$  adalah penaksir parameter dari  $\beta_k$  dan  $SE(\hat{\beta}_k)$  merupakan standar deviasi dari  $\hat{\beta}_k$ .

### B. Model GWOLR

GWOLR merupakan gabungan dari model GWR (*Geographically Weighted Regression*) dan model regresi logistik ordinal yang digunakan untuk memodelkan hubungan antara variabel respon berskala ordinal dengan variabel prediktor. Pada model ini, masing-masing koefisien regresi bergantung pada lokasi dimana data diamati. Model GWOLR dengan variabel respon dengan  $J$  buah kategori berskala ordinal adalah sebagai berikut [9].

$$\logit[P(Y_i \leq j | \mathbf{x}_i)] = \beta_{0j}(u_i, v_i) + \mathbf{x}_i^T \boldsymbol{\beta}(u_i, v_i); \quad j = 1, \dots, J-1; \quad i = 1, \dots, n \quad (5)$$

Peluang masing-masing  $J$  kategori respon adalah:

$$\gamma_j^*(\mathbf{x}_i) = \frac{\exp(\beta_{0j}(u_i, v_i) + \mathbf{x}_i^T \boldsymbol{\beta}(u_i, v_i))}{1 + \exp(\beta_{0j}(u_i, v_i) + \mathbf{x}_i^T \boldsymbol{\beta}(u_i, v_i))} - \frac{\exp(\beta_{0j-1}(u_i, v_i) + \mathbf{x}_i^T \boldsymbol{\beta}(u_i, v_i))}{1 + \exp(\beta_{0j-1}(u_i, v_i) + \mathbf{x}_i^T \boldsymbol{\beta}(u_i, v_i))} \quad (6)$$

Penaksiran parameter model GWOLR menggunakan metode *Maximum Likelihood Estimation* terboboti  $\boldsymbol{\theta}(u_i, v_i) = [\beta_{01}(u_i, v_i) \dots \beta_{0j-1}(u_i, v_i) \boldsymbol{\beta}(u_i, v_i)]^T$ . Estimasi parameter dilakukan dengan melakukan turunan parsial pertama persamaan terhadap parameter yang akan diestimasi dan kemudian disamakan dengan nol. Turunan parsial pertama berbentuk non linear sehingga digunakan iterasi *Newton-Raphson*. Beberapa pengujian hipotesis pada model GWOLR adalah uji kesamaan model GWOLR dengan model regresi logistik ordinal, uji serentak, dan parsial. Berikut adalah hipotesis dari uji kesamaan model GWOLR dengan model regresi logistik ordinal [9].

$$H_0 : \beta(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_k(u_i, v_i) = \beta_k$$

$H_1$  : paling tidak terdapat satu  $\beta_k(u_i, v_i) \neq \beta_k$ ;  $i = 1, 2, \dots, n$

Statistik uji yang digunakan adalah :

$$F_{hit} = \frac{D(\hat{\theta})/df_1}{D(\hat{\theta}^*)/df_2}; \quad df_1 = (J-1)(n-1)-p; \quad df_2 = n - \text{trace}(2\mathbf{S} - \mathbf{S}^T \mathbf{A} \mathbf{S}^{-1}) \quad (7)$$

Dimana  $D(\hat{\theta})$  adalah nilai devians dari model regresi logistik ordinal dengan derajat bebas  $df_1$  dan  $D(\hat{\theta}^*)$  adalah nilai devians dari model GWOLR dengan derajat bebas  $df_2$ . Matrik  $\mathbf{S}$  adalah matrik dengan elemen  $S_{ik} = R_{ik} \frac{z_i(u_i, v_i)}{z_k(u_i, v_i)}$ ,  $\mathbf{A}$  adalah

matriks diagonal bobot varians untuk lokasi ke- $i$  dan  $\mathbf{R}$  adalah matrik dengan elemen sebagai berikut.

$$r_i = \mathbf{x}_i (\mathbf{X}^{*T} \mathbf{W}(u_i, v_i) \mathbf{A}(u_i, v_i) \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{W}(u_i, v_i) \mathbf{A}(u_i, v_i) \quad (8)$$

Dimana  $\mathbf{W}$  adalah matriks diagonal bobot spasial untuk lokasi ke- $i$ ,  $\mathbf{X}^*$  adalah matriks dengan elemen  $\mathbf{X}^* = [\mathbf{y} \mathbf{d} \quad \mathbf{X}]$ . dengan

$$y^d_i = \begin{cases} [1 & 0] & ; \text{jika } y_i \text{ mempunyai kategori 1} \\ [0 & 1] & ; \text{jika } y_i \text{ mempunyai kategori 2} \\ [0 & 0] & ; \text{jika } y_i \text{ mempunyai kategori 3} \end{cases}$$

$$z_k(u_i, v_i) = \frac{1 - \hat{\gamma}_j^*(\mathbf{x}_k)}{\sqrt{\hat{\gamma}_j^*(\mathbf{x}_k) [1 - \hat{\gamma}_j^*(\mathbf{x}_k)]}} \quad (9)$$

Kriteria penolakan yaitu tolak  $H_0$  jika  $F_{hit} > F_{(\alpha, df_1, df_2)}$ . Hipotesis uji serentak model GWOLR adalah sebagai berikut [9].

$$H_0 : \beta(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0$$

$H_1$  : minimal ada satu  $\beta_k(u_i, v_i) \neq 0$ ;  $k = 1, 2, \dots, p$

Statistik uji yang digunakan adalah :

$$G^2 = -2 \sum_{i=1}^n \sum_{j=1}^J y_{ij} \ln \left( \frac{\sum_{k=1}^n y_{kj} w_k(u_i, v_i)}{\sum_{k=1}^n w_k(u_i, v_i)} \right) - 2 \sum_{i=1}^n \sum_{j=1}^J y_{ij} \ln(\hat{\gamma}_j^*(\mathbf{x}_i)) \quad (10)$$

Tolak  $H_0$  jika  $G^2 > \chi^2_{(\alpha, df)}$ , dimana  $df = \text{trace}(\mathbf{S})$ . Hipotesis dari uji parsial model GWOLR adalah sebagai berikut [9].

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

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

Statistik uji yang digunakan adalah :

$$Z_{hit} = \frac{\hat{\beta}_k(u_i, v_i)}{SE(\hat{\beta}_k(u_i, v_i))} \quad (11)$$

Dimana  $SE(\hat{\beta}_k(u_i, v_i)) = \sqrt{\text{var}(\hat{\beta}_k(u_i, v_i))}$ . Tolak  $H_0$  jika  $|Z_{hit}| > Z_{\alpha/2}$

C. Pemilihan Pembobot

Untuk memberikan hasil penaksiran parameter yang berbeda pada lokasi yang berbeda maka digunakan suatu pembobot. Pada analisis spasial, penaksiran parameter di suatu titik  $(u_i, v_i)$  akan lebih dipengaruhi oleh titik-titik yang dekat dengan lokasi  $(u_i, v_i)$  dari pada titik-titik yang lebih jauh. Pemilihan pembobot spasial yang digunakan dalam menaksir parameter sangat penting. Pembobot yang digunakan adalah fungsi kernel yaitu fungsi kernel yang terdiri dari 4 fungsi yaitu:

- Fungsi *Gaussian* :  $w_k(u_i, v_i) = \phi(d_{ia}/\sigma h)$  (12)

- Fungsi *Bisquare* [10]:  

$$w_k(u_i, v_i) = \begin{cases} [1 - (d_{ia}/h)^2]^2, & \text{jika } d_{ia} \leq h \\ 0, & \text{jika } d_{ia} > h \end{cases}$$
 (13)

- Fungsi *Tricube* [10]:  

$$w_k(u_i, v_i) = \begin{cases} [1 - (d_{ia}/h)^3]^3, & \text{jika } d_{ia} \leq h \\ 0, & \text{jika } d_{ia} > h \end{cases}$$
 (14)

- Fungsi *Exponential*:  $w_k(u_i, v_i) = \sqrt{\exp(- (d_{ia}/h))}$  (15)

Dengan  $\phi$  adalah densitas normal standar dan  $\sigma$  merupakan simpangan baku dari vektor jarak  $d_{ia} = \sqrt{(u_i - u_a)^2 + (v_i - v_a)^2}$  yang merupakan jarak *Euclidian* antara lokasi  $(u_i, v_i)$  dan lokasi  $(u_a, v_a)$ , serta  $h$  adalah parameter non negatif yang diketahui dan biasanya disebut parameter penghalus (*bandwidth*). Jika pembobot yang digunakan adalah fungsi kernel maka pemilihan *bandwidth* sangatlah penting karena *bandwidth* merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan data. Nilai *bandwidth* yang cukup besar akan menyebabkan bias yang semakin besar karena model yang dibentuk terlalu halus (*over smoothing*) karena banyaknya pengamatan yang digunakan. Begitupun sebaliknya, jika nilai *bandwidth* terlalu kecil maka model yang dibentuk terlalu kasar (*under smoothing*). Metode yang digunakan untuk menentukan *bandwidth* optimum adalah metode *Cross Validation* (CV) [9] dengan rumus sebagai berikut.

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{zi}(h))^2 \quad (16)$$

Dengan  $\hat{y}_{zi}(h)$  adalah nilai penaksir  $y_i$  dimana pengamatan di lokasi  $i$  dihilangkan dari proses penaksiran,  $\hat{y}_i(h)$  adalah nilai penaksir  $y_i$  dimana pengamatan di lokasi  $i$  dimasukkan dari proses penaksiran.

D. Pemilihan Model Terbaik

Dalam pemilihan model terbaik digunakan *Akaike Information Criterion* (AIC) sebagai acuan. AIC merupakan kriteria kesesuaian model dalam mengestimasi model secara statistik. AIC bertujuan untuk mendapatkan faktor-faktor yang berpengaruh terhadap model yang dirumuskan sebagai berikut.

$$AIC = D(h) + 2K \quad (17)$$

$D(h)$  merupakan nilai *deviance* model dengan *bandwidth*  $h$  dan  $K$  adalah jumlah parameter tanpa interaksi dalam model. Model terbaik adalah model dengan nilai AIC terkecil [11].

E. Buta Huruf

Buta huruf atau biasa disebut buta aksara adalah ketidakmampuan membaca dan menulis baik bahasa Indonesia maupun bahasa lainnya. Buta huruf juga dapat diartikan sebagai ketidakmampuan untuk menggunakan bahasa dan menggunakannya untuk mengerti sebuah bacaan, mendengar perkataan, mengungkapkannya dalam bentuk tulisan, dan berbicara. Dalam perkembangan saat ini kata buta huruf diartikan sebagai ketidakmampuan untuk membaca dan menulis pada tingkat yang baik untuk berkomunikasi dengan orang lain, atau dalam taraf bahwa seseorang dapat menyampaikan idenya dalam masyarakat yang mampu baca-tulis, sehingga dapat menjadi bagian dari masyarakat tersebut [12]. Angka Buta huruf merupakan persentase penduduk yang buta huruf terhadap penduduk seluruhnya di suatu daerah. Angka buta huruf dapat dihitung melalui rumus sebagai berikut.

$$\text{Angka Buta Huruf (ABH)} = \frac{\text{Jumlah penduduk buta huruf}}{\text{Jumlah penduduk seluruhnya}} \times 100\% \quad (18)$$

III. METODOLOGI PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik yaitu data Provinsi Jawa Timur dalam Angka tahun 2011, data publikasi Indikator Ekonomi dan Sosial Jawa Timur tahun 2010, dan data SUSENAS tahun 2010 Provinsi Jawa Timur. Dalam penelitian ini juga memperhatikan letak astronomis dari masing-masing titik pengamatan yaitu letak lintang dan letak bujur sebagai faktor pembobot geografisnya. Pada penelitian ini yang dijadikan unit observasi adalah 38 kabupaten/kota di Provinsi Jawa Timur. Variabel respon pada penelitian ini adalah tingkat buta huruf pada 38 kabupaten/kota di Provinsi Jawa Timur. Pada awalnya, data tingkat buta huruf ini berasal dari data angka buta huruf yang bersifat kontinu dan berdistribusi normal kemudian dibagi menjadi 3 kategori berdasarkan selang kepercayaan 95% dengan kategori I yaitu angka buta huruf di bawah batas bawah selang kepercayaan (<8,41), kategori II yaitu angka buta huruf di antara batas selang kepercayaan (8,41-13,22), dan kategori III yaitu angka buta huruf di atas batas atas selang kepercayaan (>13,22). Variabel prediktor pada penelitian ini yaitu persentase penduduk miskin ( $x_1$ ), persentase daerah berstatus kota ( $x_2$ ), tingkat pengangguran terbuka ( $x_3$ ), persentase pekerja Sosial Masyarakat ( $x_4$ ), dan angka Partisipasi murni SD ( $x_5$ ).

B. Langkah Analisis

Berikut adalah langkah-langkah analisis yang dilakukan untuk mencapai tujuan penelitian.

1. Melakukan statistika deskriptif untuk mendeskripsikan data tingkat buta huruf kabupaten/kota di Jawa Timur dan faktor-faktor yang diduga mempengaruhinya
2. Memeriksa multikolinieritas antara variabel-variabel prediktor dengan menggunakan nilai VIF dan koefisien korelasi *Pearson*.

3. Mendapatkan model regresi logistik ordinal dengan melakukan uji parameter secara serentak dan parsial serta uji kesesuaian model
4. Menyusun model GWOLR dengan langkah berikut.
  - a. Menentukan  $u_i$  dan  $v_i$  berdasarkan garis Bujur Timur dan garis Lintang Selatan untuk setiap kantor pemerintahan (kantor Bupati atau Walikota) masing-masing kabupaten/kota di Provinsi Jawa Timur.
  - b. Menghitung jarak *Euclidian* antara lokasi  $i$  yang terletak pada koordinat  $(u_i, v_i)$  terhadap lokasi  $a$  yang terletak pada koordinat  $(u_a, v_a)$ . Perhitungan ini dilakukan untuk seluruh lokasi pengamatan.
  - c. Mengurutkan jarak *euclidian* dari seluruh lokasi terhadap suatu lokasi  $i$ , sehingga diperoleh urutan tetangga terdekat dari lokasi  $i$ . Pengurutan jarak *euclidian* dilakukan untuk  $i=1,2,\dots,38$  sampai diperoleh urutan tetangga terdekat untuk seluruh lokasi.
  - d. Menentukan *bandwidth* optimum menggunakan metode *Cross Validation* (CV).
  - e. Menghitung matriks pembobot  $w_k(u_i, v_i)$  dengan memasukkan jarak *Euclidian* dan nilai *bandwidth* optimum dimana  $k=1,2,\dots,38$ , sehingga setiap lokasi  $k-i$  akan mempunyai pembobot sejumlah 38 buah.
  - f. Mendapatkan penaksir parameter model GWOLR.
  - g. Melakukan pengujian kesamaan model regresi logistik ordinal dan GWOLR
  - h. Melakukan pengujian parameter model GWOLR secara serentak dan parsial.
5. Mendapatkan model regresi terbaik untuk pemodelan tingkat buta huruf kabupaten/kota di Jawa Timur dengan membandingkan model regresi logistik ordinal dan GWOLR.
6. Menginterpretasikan hasil pengolahan atau analisis data.

#### IV. HASIL DAN PEMBAHASAN

##### A. Pemodelan Tingkat Buta Huruf Menggunakan Regresi Logistik Ordinal

Sebelum dilakukan pemodelan menggunakan regresi logistik ordinal maupun GWOLR, antar variabel prediktor dilakukan uji multikolinieritas terlebih dahulu. Berdasarkan pengujian multikolinieritas, diperoleh kesimpulan bahwa tidak ada multikolinieritas antar variabel prediktor, sehingga data dapat digunakan untuk pemodelan regresi logistik ordinal maupun GWOLR. Pemodelan tingkat buta huruf menggunakan regresi logistik ordinal dilakukan secara univariabel terlebih dahulu. Hal ini bertujuan untuk mendapatkan variabel prediktor yang berpengaruh secara nyata terhadap variabel respon. Pemodelan regresi logistik ordinal univariabel dilakukan dengan cara memodelkan variabel prediktor satu-persatu terhadap variabel respon. Didapatkan bahwa *p-value* pada semua variabel kurang dari nilai  $\alpha=0,05$ , kecuali pada variabel persentase pekerja Sosial Masyarakat ( $x_4$ ), sehingga dapat disimpulkan bahwa persentase penduduk miskin ( $x_1$ ), persentase daerah berstatus kota ( $x_2$ ), tingkat pengangguran terbuka ( $x_3$ ), dan angka Partisipasi murni SD ( $x_5$ ) berpengaruh signifikan terhadap tingkat buta huruf tiap kabupaten/kota di Jawa Timur. Setelah

dilakukan analisis regresi logistik ordinal secara univariabel, selanjutnya keempat variabel yang signifikan dimasukkan ke dalam permodelan regresi logistik ordinal secara multivariabel dan diuji secara serentak serta parsial. Pengujian secara serentak parameter regresi logistik ordinal adalah sebagai berikut.

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

$$H_1 : \text{paling tidak terdapat satu } \beta_k \neq 0 ; k = 1, 2, 3, \text{ dan } 5$$

Dengan menggunakan statistik uji  $G^2$ , didapatkan nilai  $G^2$  sebesar 42,991 pada permodelan regresi logistik secara multivariabel. Nilai  $G^2$  kemudian dibandingkan dengan nilai  $\chi^2_{(0,05;4)}$  sebesar 2,132. Dapat diketahui bahwa nilai  $G^2$  lebih besar dari pada nilai  $\chi^2_{(0,05;4)}$ , sehingga dapat disimpulkan bahwa paling tidak terdapat satu parameter yang berpengaruh secara signifikan terhadap model. Setelah dilakukan pengujian secara serentak, maka dilakukan pengujian secara parsial dengan hipotesis sebagai berikut.

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0 ; k = 1, 2, 3, \text{ dan } 5$$

Dengan menggunakan tingkat signifikansi sebesar 5%, maka didapatkan nilai  $Z_{\alpha/2}$  sebesar 1,96. Berikut adalah nilai statistik uji tiap variabel.

Tabel 1.  
Estimasi Parameter Model Regresi Logistik Ordinal Multivariabel

Variabel	Koefisien	Standar Error Koefisien	Wald	P-Value
Intersept(1)	-202,669	83,710	-2,42	0,015*
Intersept(2)	-200,228	83,422	-2,40	0,016*
$x_1$	0,007	0,189	0,04	0,970
$x_2$	0,184	0,086	2,14	0,032*
$x_3$	0,141	0,399	0,35	0,723
$x_5$	1,983	0,828	2,40	0,017*

\*)signifikan dengan  $\alpha=0,05$

Berdasarkan Tabel 1, didapatkan kesimpulan bahwa parameter-parameter yang berpengaruh secara signifikan terhadap model adalah  $\beta_{01}, \beta_{02}, \beta_2, \beta_5$ . Variabel yang berpengaruh terhadap tingkat buta huruf kabupaten/kota di Jawa Timur adalah persentase daerah berstatus kota dan angka partisipasi murni SD tiap kabupaten/kota di Jawa Timur. Jadi, model terbaik regresi logistik ordinal secara multivariabel adalah sebagai berikut.

$$\text{Logit}[\hat{P}(Y_i \leq 1|x_i)] = -202,669 + 0,007x_1 + 0,184x_2 + 0,141x_3 + 1,983x_5$$

$$\text{Logit}[\hat{P}(Y_i \leq 2|x_i)] = -200,228 + 0,007x_1 + 0,184x_2 + 0,141x_3 + 1,983x_5$$

Variabel yang berpengaruh nyata pada model regresi logistik ordinal adalah variabel  $x_2$  dan  $x_5$ . Koefisien variabel pada model menunjukkan tanda yang positif. Hal ini berarti bahwa pengaruh variabel-variabel tersebut berbanding lurus dengan model. Setiap peningkatan 1% daerah berstatus kota tiap kabupaten/kota di Jawa Timur, maka resiko tingkat buta huruf kabupaten/kota di Jawa Timur masuk kategori rendah meningkat sebesar 18,4% dibandingkan tingkat buta huruf kategori sedang atau tinggi dan tingkat buta huruf kategori rendah atau sedang meningkat 18,4% dibandingkan tingkat buta huruf kategori tinggi di tiap kabupaten/kota di Jawa Timur.

**B. Pemodelan Tingkat Buta Huruf Menggunakan GWOLR**

Langkah awal yang dilakukan untuk memodelkan tingkat buta huruf adalah dengan menentukan letak garis Lintang Selatan dan Bujur Timur yang ditentukan berdasarkan letak kantor tiap kabupaten/kota di Jawa Timur. Setelah itu menghitung jarak *Euclidian* antara lokasi ke-*i* terhadap lokasi ke-*j*, menentukan nilai *bandwidth* optimum dengan metode *Cross Validation*. Selanjutnya yaitu mendapatkan nilai pembobot dengan cara memasukkan jarak *Euclidian* dan *bandwidth* optimum ke dalam fungsi pembobot. Fungsi pembobot yang digunakan pada penelitian ini adalah fungsi kernel yang terdiri dari fungsi kernel *Gaussian*, fungsi kernel *Exponential*, fungsi kernel *Bisquare* dan fungsi kernel *Tricube*. Pembobot yang sudah didapatkan kemudian digunakan untuk menaksir parameter pada masing-masing kabupaten/kota di Jawa Timur. Untuk mendapatkan model GWOLR yang terbaik, maka digunakan kriteria *Akaike Information Criterion* (AIC). Berikut adalah perbandingan nilai *bandwidth* optimum dan nilai AIC pada masing-masing fungsi kernel.

Tabel 2.  
Perbandingan Nilai *Bandwidth* Optimum dan AIC tiap Fungsi Kernel

	Fungsi Pembobot Kernel			
	<i>Gaussian</i>	<i>Exponential</i>	<i>Bisquare</i>	<i>Tricube</i>
<i>Bandwidth</i> Optimum	2,776	2,2	4,483	4,775
AIC	44,759	44,378	46,120	47,346

Tabel 2 menunjukkan bahwa model GWOLR terbaik adalah dengan menggunakan pembobot fungsi kernel *Exponential* karena memiliki nilai AIC paling kecil. Setelah mendapatkan estimasi parameter model GWOLR, maka dilakukan pengujian kesamaan model GWOLR dengan Model Regresi Logistik Ordinal, pengujian serentak dan parsial. Berikut adalah hipotesis dari pengujian kesamaan model GWOLR dengan Model Regresi Logistik Ordinal.

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

$$H_1 : \text{minimal ada satu } \beta_k(u_i, v_i) \neq \beta_k ; i = 1, \dots, 38; k = 1, 2, 3, 5$$

Statistik uji yang digunakan adalah statistik uji F dengan daerah penolakannya adalah tolak  $H_0$  jika  $F_{hit} > F_{(\alpha, df1, df2)}$ . Tingkat signifikansi ( $\alpha$ ) yang digunakan adalah sebesar 0,05. Berikut adalah nilai statistik uji yang digunakan.

Tabel 3.

Nilai F Hitung pada uji kesamaan model GWOLR dengan model regresi logistik ordinal

Model	Devians	df	F hitung
Regresi Logistik Ordinal	37,176	70	24,04
GWOLR	29,464	1334	

Tabel 3 menunjukkan bahwa nilai F Hitung pada pengujian kesamaan model GWOLR dengan model regresi logistik ordinal adalah sebesar 24,04. Jika membandingkan nilai F hitung dengan  $F_{(0,05;70;1334)}=1,3042$  maka dapat diputuskan bahwa tolak  $H_0$ . Dengan demikian dapat disimpulkan bahwa ada perbedaan yang signifikan antara model GWOLR dan model regresi logistik ordinal pada data tingkat buta huruf tiap kabupaten/kota di Jawa Timur.

Uji serentak dilakukan untuk mengetahui signifikansi parameter  $\beta(u_i, v_i)$  terhadap variabel respon secara bersama-sama pada model GWOLR. Berikut hipotesisnya :

$$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \beta_3(u_i, v_i) = \beta_5(u_i, v_i) = 0$$

$$H_1 : \text{minimal ada satu } \beta_k(u_i, v_i) \neq 0 ; i = 1, 2, \dots, 38; k = 1, 2, 3, 5$$

Dengan menggunakan statistik uji  $G^2$ , diperoleh nilai  $G^2$  sebesar 50,969 dengan derajat bebas sebesar 7. Jika nilai  $G^2$  dengan nilai  $\chi^2_{(0,05;7)} = 14,067$  dibandingkan, maka tolak  $H_0$ . Dengan demikian, dapat disimpulkan bahwa minimal ada satu variabel prediktor yang berpengaruh signifikan terhadap tingkat buta huruf kabupaten/kota di Jawa Timur.

Uji parsial digunakan untuk mengetahui signifikansi parameter  $\beta(u_i, v_i)$  terhadap variabel respon secara parsial pada model GWOLR tiap kabupaten/kota di Jawa Timur. Hipotesis dari pengujian ini adalah sebagai berikut.

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

$$H_1 : \beta_k(u_i, v_i) \neq 0 ; i = 1, 2, \dots, 38; k = 1, 2, 3, 5$$

Pada penelitian ini, setiap kabupaten mempunyai model yang berbeda-beda sehingga variabel yang berpengaruh juga berbeda tiap kabupaten/kota. Misalkan untuk mengetahui parameter yang diduga berpengaruh di kabupaten Pacitan maka dapat dilihat nilai Z hitung pada masing-masing parameter di bawah ini :

Tabel 4.  
Nilai Z Hitung Model GWOLR di Kabupaten Pacitan

Parameter	Estimasi	Standard Error	Z-Hitung
$\beta_{01}$	-209,6807	105,1685	-1,9938*
$\beta_{02}$	-206,4667	104,5522	-1,9748*
$\beta_1$	-0,0587	0,2643	-0,2221
$\beta_2$	0,1976	0,1093	1,8079
$\beta_3$	-0,0635	0,4282	-0,1482
$\beta_5$	2,0695	1,0421	1,9860*

\*)signifikan dengan  $\alpha=0,05$

Berdasarkan Tabel 4, dapat disimpulkan bahwa parameter yang signifikan adalah  $\beta_{01}$ ,  $\beta_{02}$ , dan  $\beta_5$ . Jadi, faktor yang mempengaruhi tingkat buta huruf di Kabupaten Pacitan berdasarkan model GWOLR adalah angka partisipasi murni SD. Pemodelan tingkat buta huruf di Kabupaten Pacitan dapat ditulis sebagai berikut.

$$\text{Logit}[\hat{P}(Y_1 \leq 1 | x_i)] = -209,6087 - 0,0587x_1 + 0,198x_2 - 0,063x_3 + 2,07x_5$$

$$\text{Logit}[\hat{P}(Y_1 \leq 2 | x_i)] = -206,4667 - 0,0587x_1 + 0,198x_2 - 0,063x_3 + 2,07x_5$$

Berdasarkan model di atas, variabel yang berpengaruh nyata adalah variabel  $x_5$ . Hal ini berarti bahwa setiap peningkatan 1% angka partisipasi murni SD di Kabupaten Pacitan, maka resiko Kabupaten Pacitan masuk dalam tingkat buta huruf kategori rendah meningkat sebesar 1,986% dibandingkan tingkat buta huruf kategori sedang atau tinggi dan tingkat buta huruf kategori rendah atau sedang meningkat 1,986% dibandingkan tingkat buta huruf kategori tinggi di Kabupaten Pacitan. Pengelompokan kabupaten/kota berdasarkan variabel yang berpengaruh di Jawa Timur disajikan pada Tabel 5 berikut.

Tabel 5.

Pengelompokkan Kabupaten/Kota Berdasarkan Variabel yang Signifikan pada Model GWOLR dengan Pembobot *Kernel Exponential*

Kabupaten/Kota	Variabel yang Signifikan
Jember, Banyuwangi, Bondowoso, Pamekasan, Situbondo, Sumenep	Tidak ada variabel yang signifikan
Pacitan, Trenggalek, Tulungagung, Lumajang, Probolinggo, Pasuruan, Sidoarjo, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Kota Probolinggo, Kota Pasuruan, Kota Madiun, Kota Surabaya.	Angka partisipasi murni SD ( $X_5$ )
Ponorogo, Blitar, Kediri, Malang, Mojokerto, Jombang, Nganjuk, Kota Kediri, Kota Blitar, Kota Malang, Kota Mojokerto, Kota Batu.	Persentase daerah berstatus kota ( $X_2$ ) Angka partisipasi murni SD ( $X_5$ )

Tabel 5 menunjukkan bahwa variabel yang berpengaruh dengan menggunakan model GWOLR pembobot *Exponential* pada umumnya adalah persentase daerah berstatus kota ( $X_2$ ) dan angka partisipasi murni SD ( $X_5$ ) di wilayah tersebut. Kedua variabel tersebut memiliki pengaruh yang positif terhadap model. Namun, ada 6 kabupaten yang tidak mempunyai variabel maupun *intercept* yang signifikan terhadap tingkat buta huruf di daerah tersebut yaitu Jember, Banyuwangi, Bondowoso, Situbondo, Pamekasan, dan Sumenep.

Dengan adanya pemodelan menggunakan GWOLR, maka dapat diketahui variabel-variabel yang berpengaruh terhadap angka buta huruf masing-masing kabupaten/kota di Jawa Timur berdasarkan letak geografis tiap kabupaten/kota.

### C. Perbandingan Model Regresi Logistik Ordinal dan Model GWOLR

Setelah dilakukan pemodelan tingkat buta huruf di Jawa Timur dengan regresi logistik ordinal dan dengan GWOLR, maka dilakukan perbandingan antara kedua model tersebut berdasarkan nilai AIC yang didapatkan pada kedua model. Perbandingan ini bertujuan untuk mengetahui model yang terbaik antara model regresi logistik ordinal dan model GWOLR. Model terbaik adalah mempunyai nilai AIC yang lebih kecil. Berikut adalah nilai AIC kedua model.

Tabel 6.

Nilai AIC pada Model Regresi Logistik Ordinal dan Model GWOLR

Model	Nilai AIC
Regresi Logistik Ordinal	45,176
GWOLR (Exponential)	44,378

Pada Tabel 6 dapat dilihat bahwa pemodelan tingkat buta huruf di Jawa Timur yang terbaik adalah dengan menggunakan GWOLR. Hal ini dikarenakan nilai AIC pada model GWOLR lebih kecil daripada pada model regresi logistik ordinal, tetapi antara regresi logistik ordinal dengan GWOLR memiliki perbedaan yang tidak cukup signifikan.

## V. KESIMPULAN

Pemodelan faktor-faktor yang mempengaruhi tingkat buta huruf tiap kabupaten/kota di Jawa Timur dengan menggunakan GWOLR lebih baik daripada menggunakan model regresi logistik ordinal berdasarkan kriteria AIC. Faktor

yang mempengaruhi tingkat buta huruf tiap kabupaten/kota di Jawa Timur tahun 2010 berdasarkan model GWOLR secara umum adalah persentase daerah berstatus kota dan angka partisipasi murni Sekolah Dasar tiap kabupaten di Jawa Timur. Kedua faktor tersebut berpengaruh secara positif terhadap model. Namun, terdapat 6 kabupaten yang tidak mempunyai variabel yang signifikan yaitu Jember, Banyuwangi, Bondowoso, Pamekasan, Situbondo, dan Sumenep. Model GWOLR yang terbaik pada penelitian ini adalah model GWOLR dengan pembobot fungsi *Kernel Exponential*.

## UCAPAN TERIMA KASIH

Penulis N. L. mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) Provinsi Jawa Timur yang telah menyediakan data Provinsi Jawa Timur dalam Angka tahun 2011, data publikasi Indikator Ekonomi dan Sosial Jawa Timur tahun 2010, dan data SUSENAS tahun 2010 Provinsi Jawa Timur.

## DAFTAR PUSTAKA

- [1] Wikipedia. (2012). *Indeks Pembangunan Manusia*. [http://id.wikipedia.org/wiki/Indeks\\_Pembangunan\\_Manusia](http://id.wikipedia.org/wiki/Indeks_Pembangunan_Manusia).
- [2] Rasiyo. (2010). *1,2 Juta Warga Jawa Timur Masih Buta Aksara*. <http://www.tempo.co/read/news/2010/08/01/180267911/12-Juta-Warga-Jawa-Timur-Masih-Buta-Aksara>.
- [3] Umami, D.R. (2010). *Analisis Indikator Pembangunan Berkelanjutan Di Jawa Timur Menggunakan Metode Struktural Equation Modelling-Partial Least Square*. Tugas Akhir. Surabaya: Institut Teknologi Sepuluh Nopember.
- [4] Eka, B. (2011). *Pemodelan dan Pemetaan Angka Buta Huruf Provinsi Jawa Timur dengan Pendekatan Regresi Spasial*. Tugas Akhir. Surabaya: Institut Teknologi Sepuluh Nopember.
- [5] Astuti, A. D. (2011). *Transkeho Sebagai Pembelajaran Keaksaraan Fungsional*. <http://imadiklus.com>.
- [6] Atkinson, P. M., S. E. German, D. A. Sear, dan M. J. Clark. (2003). *Exploring The Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression*. Ohio: Ohio State University. Vol. 35, Issue: 1, Pages: 58-82.
- [7] Rifada, M. (2011). *Model Geographically Weighted Ordinal Logistic Regression Studi Kasus: Tingkat Kerawanan Desa atau Kelurahan terhadap Penyakit Demam Berdarah Dengue di Kabupaten Lamongan Tahun 2009*. Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.
- [8] Hosmer, D.W. dan S. Lemeshow. (2000). *Applied Logistic Regression*. USA: John Wiley and Sons.
- [9] Purhadi, M.Rifada, dan S.P.Wulandari,(2012). *Geographically Weighted Ordinal Logistic Regression Model*. International Journal of Mathematics and Computation. Vol.16, Issue: 3.
- [10] Chasco, C., I. Garcia, dan J. Vicens.(2008). *Modeling Spatial Variations in Household Disposable Income with Geographically Weighted Regression*. Journal of Spanish Statistics. Vol.60, Issue: 168, Pages: 321-360.
- [11] Nakaya, T., A.S. Fotheringham, C. Brunson, and M. Charlton. (2005). *Geographically Weighted Poisson Regression for Disease Association Mapping*. Statistics in Medicine. Vol. 24, Issue: 17.
- [12] Permana, H. H. (2011). *Buta Huruf*. Cimahi. <http://herhaiper.blogspot.com/2011/06/buta-huruf.html>