

## PENENTUAN FAKTOR-FAKTOR YANG MEMPENGARUHI INTENSITAS CURAH HUJAN DENGAN ANALISIS DISKRIMINAN GANDA DAN REGRESI LOGISTIK MULTINOMIAL

(Studi Kasus: Data Curah Hujan Kota Semarang dari Stasiun Meteorologi  
Maritim Tanjung Emas Periode Oktober 2018 – Maret 2019)

Sheila Faiz Rohmana<sup>1</sup>, Agus Rusgiyono<sup>2</sup>, Sugito<sup>3</sup>

<sup>1,2,3</sup>Departemen Statistika FSM Universitas Diponegoro

[Agus.rusgi@gmail.com](mailto:Agus.rusgi@gmail.com)

### ABSTRACT

Meteorologist develop rainfall forecasting methods to obtain better and more accurate rainfall information. One of them is the research of grid data and the method of grouping rainfall. According to BMKG, rainfall is classified into light, medium, and heavy rain. This study aims to determine the factors that influencing rainfall grouping using multiple discriminant analysis with a stepwise selection method. This study uses the daily climate data of Semarang City for period of October 2018 to March 2019. Based on its partial F value, the wind speed variable is eliminated so the significant variable on rainfall grouping are air temperature, air humidity, and wind direction. This analysis produces discriminant scores obtained from linear combinations between discriminant weights and observation values of significant independent variable. The classification procedure is based on the discriminant score each observations compared to cutting score resulted in classification accuracy of 62.89%. Multinomial logistic regression analysis is used to determine the effect of independent variables on rainfall intensity using the odds ratio. This analysis produces an estimate of the conditional probability of each group using significant independent variables are air temperature, air humidity, wind speed, and wind direction. The classification procedure is based on the largest conditional probability value between rainfall groups resulted in classification accuracy of 69.80%.

**Keywords:** multiple discriminant analysis, multinomial logistic regression, classification accuracy, rainfall

### 1. PENDAHULUAN

Hujan merupakan fenomena alam yang memiliki peranan penting bagi kehidupan manusia dalam berbagai bidang. Mandey (2017) mengemukakan bahwa berbagai upaya terus dilakukan untuk mendapatkan informasi mengenai curah hujan yang lebih baik dan akurat. Para peneliti di bidang meteorologi berusaha mengembangkan metode-metode prakiraan curah hujan. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) berkerjasama dengan para peneliti diluar BMKG untuk melakukan kajian terhadap data grid dan metode pengelompokkan curah hujan. Intensitas curah hujan menurut BMKG dibedakan menjadi tiga kelompok yaitu hujan ringan, hujan sedang, dan hujan lebat. Kota Semarang termasuk daerah iklim tropis yang dipengaruhi oleh angin muson sehingga sesuai dengan pola khas curah hujan di Indonesia. Hujan di Indonesia umumnya berlangsung pada bulan Oktober sampai Maret.

Proses terjadinya hujan berhubungan dengan unsur-unsur cuaca lainnya seperti temperatur udara, kelembaban udara, kecepatan angin, dan arah angin maka dilakukan penelitian untuk menentukan unsur-unsur cuaca apakah yang mempengaruhi pengelompokkan hujan menggunakan analisis diskriminan ganda dan melakukan pemodelan

intensitas curah hujan untuk menentukan besar pengaruh unsur-unsur cuaca terhadap intensitas curah hujan menggunakan analisis regresi logistik multinomial. Kemampuan prosedur klasifikasi pada masing-masing analisis diskriminan ganda dan regresi logistik multinomial dinilai menggunakan tingkat kesalahan klasifikasi atau sebaliknya dengan tingkat ketepatan klasifikasi untuk mengetahui kemampuan prediksi dari model.

Analisis diskriminan ganda adalah suatu teknik statistika yang menggunakan beberapa kombinasi linier dari variabel independen untuk membedakan kelompok-kelompok yang diminati (Dillon dan Goldstein, 1984). Analisis diskriminan biasanya menggunakan metode seleksi bertahap yaitu menambahkan variabel satu per satu, dan pada setiap langkah, variabel diperiksa kembali untuk melihat apakah terdapat variabel yang telah masuk sebelumnya menjadi redundan dengan variabel yang baru ditambahkan (Rencher, 2002). Pengamatan pada  $p$  variabel independen dan suatu variabel dependen dengan  $k$  kelompok akan terbentuk  $\min(p, k - 1)$  fungsi diskriminan yang berasal dari kombinasi linier, dituliskan sebagai berikut:

$$y_r = b_{r1}x_1 + b_{r2}x_2 + b_{r3}x_3 + \dots + b_{rp}x_p, \quad r = 1, 2, \dots, k - 1 \text{ dan } m = 1, 2, \dots, p$$

dengan:  $y_r$  = fungsi diskriminan ke- $r$

$b_{rm}$  = bobot diskriminan variabel independen ke- $m$  fungsi diskriminan ke- $r$

$x_m$  = variabel independen ke- $m$

Tujuan analisis diskriminan ganda adalah mendapatkan suatu fungsi dengan memaksimalkan perbandingan variabilitas antarkelompok dan dalam-kelompok. Sejalan dengan kriteria Fisher, perbandingan antara varian antarkelompok dan varian dalam-kelompok untuk  $k$  kelompok pada kombinasi linier pertama didefinisikan sebagai:

$$\hat{\lambda} = \frac{\hat{\mathbf{b}}' \mathbf{B} \hat{\mathbf{b}}}{\hat{\mathbf{b}}' \mathbf{W} \hat{\mathbf{b}}}$$

$\hat{\lambda}$  dimaksimumkan dengan melakukan turunan parsial terhadap  $\hat{\mathbf{b}}$  dan disamadengkan 0, kemudian dikalikan dengan  $\mathbf{W}^{-1}$  sehingga diperoleh:

$$(\mathbf{W}^{-1} \mathbf{B} - \hat{\lambda} \mathbf{I}) \hat{\mathbf{b}} = \mathbf{0}$$

Nilai maksimum dari  $\hat{\lambda}$  yaitu nilai akar ciri terbesar dari matriks  $\mathbf{W}^{-1} \mathbf{B}$ . Vektor akar ciri  $\hat{\mathbf{b}}$  memiliki elemen berupa bobot diskriminan yang sesuai dengan kombinasi liniernya.

Ukuran relatif masing-masing nilai akar ciri,  $\hat{\lambda}_r$ , menunjukkan indeks kepentingan masing-masing fungsi diskriminan yang digunakan untuk melihat peranan relatif masing-masing fungsi diskriminan dalam memisahkan anggota kelompok, dapat dinyatakan dalam persentase variansi total sebagai berikut:

$$\% \text{ variansi} = \frac{\hat{\lambda}_r}{\sum_r \hat{\lambda}_r} \times 100\%$$

Pengujian signifikansi fungsi diskriminan dilakukan secara serentak maupun individu. Uji serentak menggunakan uji V-Bartlett untuk menguji signifikansi nilai akar ciri dari  $\mathbf{W}^{-1} \mathbf{B}$  dengan hipotesis nol adalah  $\lambda_1 = \lambda_2 = \dots = \lambda_{k-1} = 0$  atau semua fungsi diskriminan tidak signifikan secara serentak. Statistik uji yang digunakan:

$$V = \left\{ (n - 1) - \frac{1}{2}(p + k) \right\} \sum_{r=1}^{k-1} \ln(1 + \lambda_r)$$

dengan taraf signifikansi  $\alpha$  maka tolak  $H_0$  jika  $V > \chi_{\alpha, p(k-1)}^2$ . Uji individu menggunakan uji V-Bartlett yang menguji masing-masing nilai akar ciri dari  $\mathbf{W}^{-1} \mathbf{B}$  dengan hipotesis nol adalah  $\lambda_r = 0$  atau fungsi diskriminan ke- $r$  tidak signifikan. Statistik uji yang digunakan:

$$V_r = \left\{ (n - 1) - \frac{1}{2}(p + k) \right\} \sum_{s=r}^{k-1} \ln(1 + \lambda_s)$$

dengan taraf signifikansi  $\alpha$  maka tolak  $H_0$  jika  $V_r > \chi_{\alpha, (p-r)(k-r)}^2$ .

Analisis diskriminan ganda sebaiknya memenuhi asumsi normalitas multivariat dan homogenitas matriks kovarian. Pengujian normalitas multivariat secara visual menggunakan plot chi-kuadrat yaitu grafik pasangan  $(q_{c,p}((i - \frac{1}{2})/n_j), d_{(i)}^2)$ . Nilai  $q_{c,p}((i - \frac{1}{2})/n_j)$  adalah kuantil  $\{100(i - \frac{1}{2})/n_j\}$  dari distribusi chi kuadrat berderajat bebas  $p$  dan nilai  $d_{(i)}^2$  adalah jarak mahalanobis  $d_i^2$  yang diurutkan mulai dari yang terkecil. Plot-plot yang mengikuti garis lurus mengindikasikan bahwa data berasal dari sebaran normal multivariat (Johnson dan Wichern, 1998).

$$d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1, 2, \dots, n_j \text{ dan } j = 1, 2, \dots, k$$

Pengujian normalitas multivariat secara formal menggunakan prosedur Kolmogorov-Smirnov (Daniel, 1989) dengan hipotesis nol adalah data suatu kelompok berdistribusi normal multivariat ( $d_i^2$  berdistribusi  $\chi_p^2$ ). Statistik uji yang digunakan:

$$D = \sup |S(d_i^2) - F_0(d_i^2)|$$

dengan  $S(d_i^2)$  adalah proporsi jarak mahalanobis yang kurang dari atau sama dengan  $d_i^2$  dan  $F_0(d_i^2)$  adalah fungsi peluang kumulatif dari distribusi chi kuadrat. Pada taraf signifikansi  $\alpha$  dan  $W_{1-\alpha}$  adalah nilai kuantil  $(1 - \alpha)$  dari Kolmogorov-Smirnov, maka tolak  $H_0$  jika  $D < W_{1-\alpha}$ .

Pengujian asumsi homogenitas matriks kovarian menggunakan statistik uji Box' M dengan hipotesis nol adalah  $\Sigma_1 = \Sigma_2 = \dots = \Sigma_k$  atau matriks varian-kovarian dari  $k$  kelompok adalah sama. Statistik uji yang digunakan:

$$u = -2(1 - c_1 \ln M), \text{ dengan:}$$

$$\ln M = \frac{1}{k} \sum_{j=1}^k v_j \ln |S_j| - \frac{1}{k} \left( \sum_{j=1}^k v_j \right) \ln |S_{pool}|; \quad c_1 = \left[ \frac{\sum_{j=1}^k \frac{1}{v_j} - \frac{1}{\sum_{j=1}^k v_j} \right] \left[ \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right]; \quad S_{pool} = \frac{\sum_{j=1}^k v_j S_j}{\sum_{j=1}^k v_j}$$

$v_j = n_j - 1$  dan  $S_j$  adalah matriks kovarian dari sampel ke- $j$  untuk  $j = 1, 2, \dots, k$ . Pada taraf signifikansi  $\alpha$  maka tolak  $H_0$  jika  $u > \chi_{\alpha, [\frac{1}{2}(k-1)p(p+1)]}^2$

Analisis regresi logistik multinomial merupakan analisis regresi logistik yang menggunakan variabel dependen bersifat multinomial yaitu skala nominal dengan lebih dari dua kategori. Bentuk umum peluang bersyarat untuk  $k$  kategori pada variabel dependen  $Y$  dapat ditulis:

$$\pi_j(x) = P(Y = j|x) = \frac{e^{g_j(x)}}{\sum_{t=0}^{k-1} e^{g_t(x)}} = \frac{e^{\beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p}}{\sum_{t=0}^{k-1} e^{\beta_{t0} + \beta_{t1}x_1 + \beta_{t2}x_2 + \dots + \beta_{tp}x_p}}, \quad j = 0, 1, 2, \dots, k - 1$$

dengan vektor  $\beta_0 = \mathbf{0}$  dan  $g_0(x) = 0$ .

Estimasi parameter regresi logistik menggunakan metode maksimum likelihood yang dilanjutkan dengan iterasi Newton Raphson. Metode maksimum likelihood akan memberikan nilai estimasi parameter  $\beta$  yang memaksimalkan fungsi likelihood (Agresti, 2007). Fungsi likelihood bersyarat untuk sebuah sampel dengan  $n$  percobaan secara independen ditunjukkan:

$$\ell(\beta) = \prod_{i=1}^n [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}} \dots \pi_{k-1}(x_i)^{y_{(k-1)i}}]$$

Jika  $\sum y_{ji} = 1$  untuk setiap  $i$  dan  $i = 1, 2, \dots, n$  maka fungsi log-likelihoodnya adalah:

$$L(\beta) = \ln\{\ell(\beta)\} = \sum_{i=1}^n [y_{1i} g_1(x_i) + y_{2i} g_2(x_i) + \dots + y_{(k-1)i} g_{k-1}(x_i) - \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)} + \dots + e^{g_{k-1}(x_i)})]$$

Persamaan likelihood diperoleh dengan melakukan turunan parsial pertama  $L(\beta)$  terhadap  $\beta$ . Bentuk umum persamaan likelihood adalah:

$$\frac{\partial L(\beta)}{\partial \beta_{rm}} = \sum_{i=1}^n \left[ y_{ri} x_{mi} - \frac{x_{mi} e^{g_{k-1}(x_i)}}{1 + e^{g_1(x_i)} + e^{g_2(x_i)} + \dots + e^{g_{k-1}(x_i)}} \right] = \sum_{i=1}^n x_{mi} [y_{ri} - \pi_r(x_i)]$$

dengan masing-masing  $x_{0i} = 1; r = 1, 2, \dots, k - 1$  dan  $m = 0, 1, 2, \dots, p$

Matriks turunan parsial kedua digunakan untuk memperoleh matriks informasi dan estimator matriks kovarian dari estimator maksimum likelihood (Hosmer dan Lemesho, 2000).  $\hat{I}(\hat{\beta})$  adalah matriks informasi yang memiliki elemen-elemen  $\hat{I}(\hat{\beta})_{rr} = (X'V_r X)$  dan  $\hat{I}(\hat{\beta})_{r_1 r_2} = \hat{I}(\hat{\beta})_{r_2 r_1} = -(X'V_{r_1 r_2} X)$ . Matriks  $V_r$  adalah matriks diagonal berukuran  $n \times n$  dengan elemen umum:  $\hat{\pi}_r(x_i)[1 - \hat{\pi}_r(x_i)]$  dan matriks  $V_{r_1 r_2}$  adalah matriks diagonal berukuran  $n \times n$  dengan elemen umum:  $\hat{\pi}_{r_1}(x_i) \hat{\pi}_{r_2}(x_i)$ . Estimator matriks kovariannya ditunjukkan dengan  $\widehat{\text{var}}(\hat{\beta}) = \hat{I}(\hat{\beta})^{-1}$ . Nilai  $\hat{\beta}$  diperoleh menggunakan iterasi Newton-Raphson dengan memilih taksiran awal untuk  $\beta$  misal  $\hat{\beta} = \mathbf{0}$ , selanjutnya menghitung taksiran baru pada setiap  $t + 1$  yaitu:

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + \hat{I}(\hat{\beta})^{-1} \{M(Y - \pi(x))\}$$

dengan  $M = I_{(k-1) \times (k-1)} \otimes X'$  dan iterasi berakhir jika diperoleh:  $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)}$

Pengujian signifikansi parameter  $\hat{\beta}$  dilakukan secara serentak maupun individu. Uji serentak menggunakan uji rasio likelihood dengan hipotesis nol adalah  $\beta_{1p} = \beta_{2p} = \dots = \beta_{(k-1)p} = 0$  atau semua variabel independen tidak signifikan secara bersama-sama. Statistik uji yang digunakan:

$$G = -2 \ln \left( \frac{\text{likelihood tanpa variabel independen}}{\text{likelihood dengan variabel independen}} \right)$$

dengan taraf signifikansi  $\alpha$  dan  $v = (k - 1) \sum_{m=1}^p (J_m - 1)$  dimana  $k$  dan  $J_m$  adalah banyaknya kategori pada variabel dependen dan variabel independen, maka tolak  $H_0$  jika  $G > \chi^2_{(\alpha, v)}$ . Uji individu menggunakan uji Wald dengan hipotesis nol adalah  $\beta_{rm} = 0$  atau variabel independen ke- $m$  tidak signifikan. Statistik uji yang digunakan:

$$W_{rm} = \left( \frac{\hat{\beta}_{rm}}{\text{se}(\hat{\beta}_{rm})} \right)^2$$

dengan taraf signifikansi  $\alpha$  maka tolak  $H_0$  jika  $G > \chi^2_{(\alpha, 1)}$ .

Interpretasi dari model regresi logistik menggunakan *odds* ( $\Omega$ ) dan *odds ratio* ( $\theta$ ). *Odds ratio* merupakan perbandingan dua *odds*. Apabila kategori 0 sebagai kategori dasar, maka *odds ratio* untuk kategori  $j$  dibandingkan kategori 0 pada variabel dependen  $Y$  yaitu:

$$\theta_{(j,0)} = \frac{\Omega_j(x)}{\Omega_0(x)} = \frac{P(Y = j|x) / (1 - P(Y = j|x))}{P(Y = 0|x) / (1 - P(Y = 0|x))}$$

Kemampuan prosedur klasifikasi digunakan untuk memprediksi keanggotaan kelompok yang dapat dinilai menggunakan probabilitas kesalahan klasifikasi disebut tingkat kesalahan klasifikasi (*error rate*) atau menggunakan tingkat ketepatan klasifikasi (*correct classification rate*). Hasil prosedur klasifikasi dapat ditunjukkan pada tabel klasifikasi. Nilai ketepatan klasifikasi dari pengamatan  $k$  kelompok adalah:

$$\text{apparent correct classification rate} = \frac{n_{11} + n_{22} + n_{33} + \dots + n_{kk}}{n_1 + n_2 + n_3 + \dots + n_k}$$

dengan  $n_j = n_{j1} + n_{j2} + \dots + n_{jk}$  untuk  $j = 1, 2, \dots, k$

## 2. METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah data iklim harian di Kota Semarang. Data tersebut merupakan data sekunder yang diperoleh dari [website www.dataonline.bmkg.go.id](http://www.dataonline.bmkg.go.id) yaitu basis data Badan Meteorologi, Klimatologi, dan

Geofisika (BMKG) yang diamati oleh Stasiun Meteorologi Maritim Tanjung Emas. Penelitian menggunakan data periode Oktober 2018 – Maret 2019.

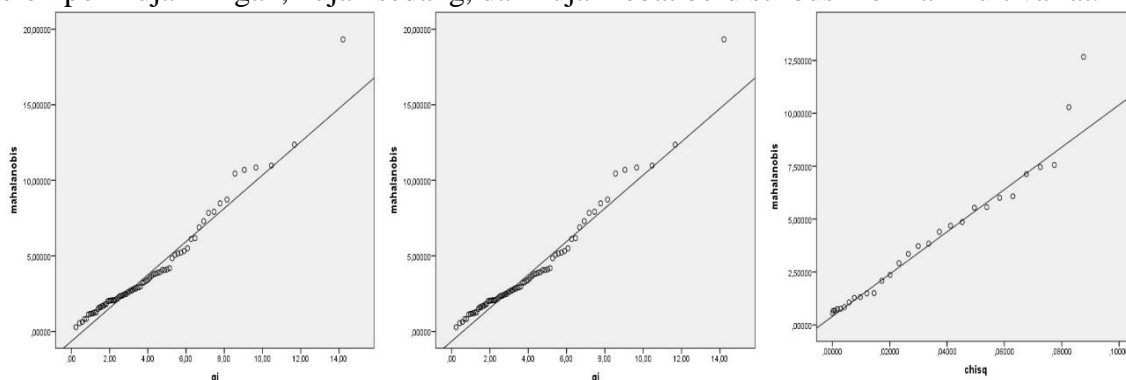
Penelitian menggunakan suatu variabel dependen dan beberapa variabel independen. Variabel dependen yang digunakan adalah intensitas curah hujan yang diklasifikasikan menjadi 3 kelompok yaitu kelompok 0 = hujan ringan, 1 = hujan sedang, dan 2 = hujan lebat. Variabel independen menggunakan peubah kontinu yaitu variabel temperatur udara, kelembaban udara, kecepatan angin, dan arah angin.

Tahapan dalam analisis diskriminan ganda adalah pemeriksaan asumsi normalitas multivariat dan homogenitas matriks kovarian, selanjutnya membentuk fungsi diskriminan dan pengujian signifikansi fungsi diskriminan secara serentak maupun individu, mengevaluasi fungsi diskriminan dan menilai ketepatan klasifikasi dari analisis diskriminan ganda tersebut.

Tahapan dalam analisis regresi logistik multinomial adalah membentuk model awal regresi logistik multinomial dan pengujian signifikansi parameter secara serentak maupun individu, menentukan model akhir, menghitung nilai *odds ratio* untuk interpretasi model dan menilai ketepatan klasifikasi dari analisis regresi logistik multinomial tersebut.

### 3. HASIL DAN PEMBAHASAN

Pemeriksaan asumsi normalitas multivariat dilakukan pada masing-masing kelompok hujan untuk mengetahui apakah  $p$  variabel independen berdistribusi normal multivariat. Pengujian normalitas multivariat secara visual menggunakan plot chi kuadrat pada Gambar 1 menunjukkan plot-plot setiap grafik mengikuti garis lurus yang berarti bahwa jarak mahalnobis mengikuti sebaran chi kuadrat, sehingga data masing-masing kelompok hujan ringan, hujan sedang, dan hujan lebat berdistribusi normal multivariat.



Gambar 1. Plot Chi Kuadrat Kelompok Hujan Ringan, Hujan Sedang, Hujan Lebat  
 Pengujian normalitas multivariat secara formal menggunakan prosedur Kolmogorov-Smirnov dengan hipotesis nol adalah data kelompok hujan ke- $j$  mengikuti sebaran normal multivariat untuk  $j = 1,2,3$ .

Tabel 1. Hasil Prosedur Kolmogorov-Smirnov

$j$	Hujan	$n_j$	$D$	$W_{0,95;n_j}$	Keputusan
1	Ringan	55	0,1791	0,183	$H_0$ diterima
2	Sedang	75	0,1063	0,157	$H_0$ diterima
3	Lebat	29	0,1788	0,246	$H_0$ diterima

Pada taraf signifikansi  $\alpha = 5\%$   $H_0$  diterima untuk masing-masing kelompok karena  $D < W_{0,95}$  sehingga masing-masing kelompok hujan ringan, hujan sedang, dan hujan lebat mengikuti sebaran normal multivariat.

Pengujian asumsi homogenitas matriks kovarian antarkelompok hujan ringan, hujan sedang, dan hujan lebat menggunakan uji Box' M dengan hipotesis nol adalah  $\Sigma_1 = \Sigma_2 =$

$\Sigma_3$ . Pada taraf signifikansi  $\alpha = 5\%$   $H_0$  diterima karena  $u = 14,235 < \chi_{0,05;20}^2 = 31,4104$  sehingga matriks kovarian antarkelompok hujan sama.

Asumsi normalitas multivariat dan homogenitas matriks kovarian terpenuhi sehingga dapat dilakukan analisis diskriminan ganda. Prosedur analisis diskriminan bertahap menunjukkan bahwa variabel kecepatan angin memiliki nilai F terkecil dan  $F_{kec\_angin} = 3,604 < F_{in} = 3,84$  sehingga variabel kecepatan angin dieliminasi. Variabel yang signifikan membedakan antarkelompok hujan sedang, hujan ringan, dan hujan lebat adalah variabel temperatur udara, kelembaban udara, dan arah angin sehingga dua fungsi diskriminan yang terbentuk adalah:

$$y_1 = 1,009 \text{ temperatur} - 0,521 \text{ kelembaban} + 0,079 \text{ arah angin}$$

$$y_2 = 0,553 \text{ temperatur} + 0,905 \text{ kelembaban} - 0,734 \text{ arah angin}$$

Pengujian signifikansi fungsi diskriminan secara serentak menggunakan uji V-Bartlett dengan hipotesis nol  $\lambda_1 = \lambda_2 = 0$  atau kedua fungsi diskriminan tidak signifikan secara serentak. Pada taraf signifikansi  $\alpha = 5\%$   $H_0$  ditolak karena  $V = 96,347 > \chi_{0,05;8}^2 = 15,507$  sehingga kedua fungsi diskriminan signifikan secara serentak.

Uji serentak terpenuhi maka dapat dilakukan pengujian signifikansi fungsi diskriminan secara individu menggunakan uji V-Bartlett masing-masing nilai akar ciri dengan hipotesis nol  $\lambda_r = 0$  atau fungsi diskriminan ke- $r$  tidak signifikan untuk  $r = 1,2$ .

Tabel 2. Hasil Uji V-Bartlett Masing-Masing Fungsi Diskriminan

Fungsi	$V_r$	df	$\chi_{0,05;df}^2$	Keputusan
1	96,347	6	12,592	$H_0$ ditolak
2	15,767	2	5,991	$H_0$ ditolak

Pada taraf signifikansi  $\alpha = 5\%$   $H_0$  ditolak karena  $V_r > \chi_{0,05;df}^2$  sehingga masing-masing fungsi diskriminan pertama dan fungsi diskriminan kedua signifikan.

Fungsi diskriminan pertama memiliki 86,4% keragaman berarti fungsi diskriminan pertama mampu menjelaskan variasi pada kelompok hujan sebesar 86,4%.  $R^2 = 0,4058$  berarti semua variabel independen pada fungsi diskriminan pertama dapat menjelaskan curah hujan sebesar 40,58% dan sisanya sebesar 59,42% dijelaskan oleh faktor-faktor lainnya.

Fungsi diskriminan kedua memiliki 13,6% keragaman berarti bahwa fungsi diskriminan kedua mampu menjelaskan variasi pada kelompok hujan sebesar 13,6%.  $R^2 = 0,0967$  berarti semua variabel independen pada fungsi diskriminan kedua dapat menjelaskan curah hujan sebesar 9,67% dan sisanya dijelaskan oleh faktor-faktor lainnya.

Tabel 3. Hasil Klasifikasi Analisis Diskriminan Ganda

Pengamatan	Prediksi			Benar (%)
	Hujan Ringan	Hujan Sedang	Hujan Lebat	
Hujan Ringan	34	17	4	61,82
Hujan Sedang	19	43	13	57,33
Hujan Lebat	1	5	23	79,31
Keseluruhan (%)				62,89

Prosedur klasifikasi pada analisis diskriminan ganda dilakukan dengan membandingkan masing-masing skor diskriminan dan skor pembeda (*cutting score*), dimana skor diskriminan tiap pengamatan diperoleh dengan memasukkan nilai-nilai pengamatan ke dalam fungsi diskriminan. Pengklasifikasian curah hujan menunjukkan bahwa curah hujan dapat diklasifikasikan dengan tepat sebesar 62,9% dan mengalami kesalahan klasifikasi sebesar 37,1%.

Analisis regresi logistik multinomial menggunakan variabel curah hujan sebagai variabel dependen  $Y$  yang dikategorikan menjadi hujan ringan, hujan sedang, dan hujan lebat. Kategori dasar yang digunakan pada analisis ini adalah hujan ringan. Model peluang bersyarat untuk masing-masing kategori curah hujan ringan, hujan sedang, dan hujan lebat adalah:

$$\pi_0(x) = \frac{1}{1 + \exp(0,709 - 0,697X_1 + 1,037X_2 + 0,747X_3 - 0,592X_4) + \exp(-1,697 - 2,91X_1 + 1,409X_2 + 0,434X_3 - 0,026X_4)}$$

$$\pi_1(x) = \frac{\exp(0,709 - 0,697X_1 + 1,037X_2 + 0,747X_3 - 0,592X_4)}{1 + \exp(0,709 - 0,697X_1 + 1,037X_2 + 0,747X_3 - 0,592X_4) + \exp(-1,697 - 2,91X_1 + 1,409X_2 + 0,434X_3 - 0,026X_4)}$$

$$\pi_2(x) = \frac{\exp(-1,697 - 2,91X_1 + 1,409X_2 + 0,434X_3 - 0,026X_4)}{1 + \exp(0,709 - 0,697X_1 + 1,037X_2 + 0,747X_3 - 0,592X_4) + \exp(-1,697 - 2,91X_1 + 1,409X_2 + 0,434X_3 - 0,026X_4)}$$

Pengujian signifikansi parameter  $\hat{\beta}$  dilakukan secara serentak menggunakan uji rasio likelihood dengan hipotesis nol adalah  $\beta_{11} = \beta_{12} = \dots = \beta_{24} = 0$ . Pada taraf signifikansi  $\alpha = 5\%$   $H_0$  ditolak karena  $G = 110,214 > \chi_{0,05;8}^2 = 15,507$  maka variabel independen secara bersama-sama signifikan mempengaruhi intensitas curah hujan.

Uji serentak terpenuhi maka dapat dilakukan pengujian signifikansi parameter  $\hat{\beta}$  secara individu menggunakan uji Wald dengan hipotesis nol adalah  $\beta_{rm} = 0$  untuk  $r = 1,2$  dan  $m = 1,2,3,4$ .

Tabel 4. Hasil Uji Wald

Variabel	Nilai $\beta_{rm}$	$se(\beta_{rm})$	$W_{rm}$	$\chi_{1;0,05}^2$	Keputusan
Temperatur	$\beta_{11} = -0,697$	0,279	6,240	3,84	$H_0$ ditolak
	$\beta_{21} = -0,291$	0,537	29,449	3,84	$H_0$ ditolak
Kelembaban	$\beta_{12} = 1,037$	0,268	14,941	3,84	$H_0$ ditolak
	$\beta_{22} = 1,409$	0,450	9,790	3,84	$H_0$ ditolak
Kecepatan angin	$\beta_{14} = 0,747$	0,268	7,752	3,84	$H_0$ ditolak
	$\beta_{24} = 0,434$	0,357	1,476	3,84	$H_0$ diterima
Arah angin	$\beta_{15} = -0,592$	0,228	6,747	3,84	$H_0$ ditolak
	$\beta_{25} = -0,062$	0,373	0,028	3,84	$H_0$ diterima

Pada taraf signifikansi  $\alpha = 5\%$   $H_0$  ditolak untuk  $W_{rm} > \chi_{0,05;1}^2 = 3,84$  sehingga masing-masing variabel temperatur udara, kelembaban udara, kecepatan angin, dan arah angin signifikan mempengaruhi intensitas curah hujan.

Tabel 5. Interpretasi Model Regresi Logistik Multinomial

Model	Variabel	$\beta$	$e^\beta$	$1 - e^\beta$
Hujan Sedang	Temperatur ( $X_{11}$ )	-0,697	0,498	-0,502
	Kelembaban ( $X_{12}$ )	1,037	2,820	1,820
	Kecepatan angin ( $X_{13}$ )	0,747	2,110	1,110
	Arah angin ( $X_{14}$ )	-0,592	0,553	-0,447
Hujan Lebat	Temperatur ( $X_{21}$ )	-2,914	0,054	-0,946
	Kelembaban ( $X_{22}$ )	1,409	4,094	3,094

Interpretasi model regresi logistik multinomial menggunakan nilai *odds ratio* yang ditunjukkan pada Tabel 5. Setiap kenaikan 1 unit pada variabel temperatur akan menurunkan peluang terjadinya hujan sedang sebesar 0,502 kali dibandingkan peluang terjadinya hujan ringan, sedangkan setiap kenaikan 1 unit pada variabel temperatur akan menurunkan peluang terjadinya hujan lebat sebesar 0,946 kali dibandingkan peluang terjadinya hujan ringan.

Setiap kenaikan 1 unit pada variabel kelembaban akan meningkatkan peluang terjadinya hujan sedang sebesar 1,82 kali dibandingkan peluang terjadinya hujan ringan, sedangkan kenaikan 1 unit pada variabel kelembaban akan meningkatkan peluang terjadinya hujan lebat sebesar 3,094 kali dibandingkan peluang terjadinya hujan ringan. Setiap kenaikan 1 unit pada variabel kecepatan angin akan meningkatkan peluang terjadinya hujan sedang sebesar 1,11 kali dibandingkan peluang terjadinya hujan ringan. Setiap kenaikan 1 unit pada variabel arah angin akan menurunkan peluang terjadinya hujan sedang sebesar 0,447 kali dibandingkan peluang terjadinya hujan ringan.

Tabel 6. Hasil Klasifikasi Analisis Regresi Logistik Multinomial

Pengamatan	Prediksi			Benar (%)
	Hujan Ringan	Hujan Sedang	Hujan Lebat	
Hujan Ringan	37	17	1	67,3
Hujan Sedang	14	55	6	73,3
Hujan Lebat	1	9	19	65,5
Keseluruhan (%)				69,8

Prosedur klasifikasi pada analisis regresi logistik multinomial dilakukan dengan menentukan nilai peluang terbesar dari 3 kelompok curah hujan pada masing-masing pengamatan, estimasi peluang tiap pengamatan diperoleh dengan memasukkan nilai-nilai pengamatan ke dalam model peluang bersyarat masing-masing kelompok hujan ringan, hujan sedang, dan hujan lebat. Pengklasifikasian kelompok curah hujan menunjukkan bahwa curah hujan dapat diklasifikasikan dengan tepat sebesar 69,8% dan mengalami kesalahan klasifikasi sebesar 30,2%.

#### 4. KESIMPULAN

Berdasarkan hasil dan pembahasan penentuan faktor-faktor yang mempengaruhi intensitas curah hujan di Kota Semarang, diperoleh kesimpulan bahwa faktor-faktor yang signifikan terhadap pengelompokan hujan berdasarkan analisis diskriminan ganda dengan metode seleksi bertahap adalah temperatur udara, kelembaban udara, dan arah angin. Fungsi diskriminan yang terbentuk adalah:

$$y_1 = 1,009 \text{ temperatur} - 0,521 \text{ kelembaban} + 0,079 \text{ arah angin}$$

$$y_2 = 0,553 \text{ temperatur} + 0,905 \text{ kelembaban} - 0,734 \text{ arah angin}$$

Pemodelan intensitas curah hujan berdasarkan regresi logistik multinomial menggunakan variabel independen yang signifikan mempengaruhi intensitas curah hujan yaitu temperatur udara (X1), kelembaban udara (X2), kecepatan angin (X3), dan arah angin (X4). Fungsi logit yang terbentuk adalah:

$$g_1(x) = 0,709 - 0,697X_1 + 1,037X_2 + 0,747X_3 - 0,592X_4$$

$$g_2(x) = -1,697 - 2,914X_1 + 1,409X_2 + 0,434X_3 - 0,026X_4$$

Penelitian ini menghasilkan ketepatan klasifikasi curah hujan berdasarkan analisis diskriminan ganda sebesar 62,9% dan ketepatan klasifikasi intensitas curah hujan berdasarkan analisis regresi logistik multinomial sebesar 69,8%.

#### DAFTAR PUSTAKA

- Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. USA: John Wileys & Sons.
- Daniel, W.W. 1989. *Statistika Nonparametrik Terapan*. Diterjemahkan oleh: Alex Tri Kantjono W. Jakarta: PT Gramedia. Terjemahan dari: *Applied Nonparametric Statistics*.
- Dillon, W.R., Goldstein, M. 1984. *Multivariate Analysis: Methods and Applications*. New York: John Wiley & Sons.



- Hosmer, D.W., Lemeshow, S. 2000. *Applied Logistic Regression*. USA: John Wiley & Sons.
- Johnson, R.A., Wichern, D.W. 1998. *Applied Multivariate Statistical Analysis*. USA: Pearson Education International.
- Mandey, F.N. 2017. *Pemodelan Sistem Prediksi Intensitas Curah Hujan di Kota Manado dengan Menggunakan Kontrol Logika Fuzzy*. Jurnal MIPA UNSRAT Online Vol. 2, No. 6: Hal. 19-23.
- Rencher, A.C. 2002. *Methods of Multivariate Analysis*. USA: John Wiley & Sons.