

# Analisis Regresi Logistik Multinomial Pada Faktor-Faktor Yang Mempengaruhi Sumber Air Bersih Rumah Tangga di Jawa Timur

Zabrinna Zenitha Zahroh, dan Ismaini Zain

Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data

Institut Teknologi Sepuluh Nopember (ITS)

*e-mail*: ismaini\_z@statistika.its.ac.id

**Abstrak**—Water is an important resource for people's lives to meet their needs such as for drinking, cooking, bathing / laundry and more. It is guaranteed in the Sustainable Development Goals or SDGs program, which becomes a reference for the development of the country, both for developed countries and developing countries in order to answer the country's lack of development by achieving the availability of clean water for the community. In this research, the characteristic of X1 variable is 52% located in rural area and 48% is located in urban area. Variable X2 as much as 70% that is SD, 6% that is junior, 18% is high school, and 6% that is college. The X3 variable is 91% of self-owned households and 9% non-self-owned. Variabel X4 as much as 47% of households use general pipes / hydrants and 53% do not use water delivery. The variable X5 or the number of family members in East Java has an average family member of 3.4558 or as many as 4 family members with a minimum of 1 family member in the family and a maximum of 14 family members in the family. This research yields accuracy value in model analysis equal to 67,3%.

**Kata Kunci**—Consumption Scheme, Multinomial Logistic Regression, Water Supply.

## I. PENDAHULUAN

AIR dikonsumsi untuk memenuhi kebutuhan sehari-hari seperti untuk minum, memasak, mandi/cuci dan lainnya. Air yang dapat dikonsumsi memiliki beberapa jenis antara lain air kemasan bermerek, air isi ulang, leding meteran, leding eceran, sumur bor/pompa, sumur terlindungi, sumur tak terlindungi, mata air terlindungi, mata air tak terlindungi, air permukaan yang terdiri dari sungai, danau, waduk, kolam dan irigasi, lalu ada pula air hujan. Pemenuhan akan konsumsi kebutuhan air bersih yang diperlukan oleh seluruh masyarakat dijamin dalam tujuan MDGs (Millennium Development Goals) yang dibuat oleh organisasi UNDP (United Nations Development Programs), yang menyebutkan bahwa kebutuhan air bersih dan sanitasi yang lebih baik masuk dalam tujuan ke-7, yaitu “ensure environmental sustainability” atau menjamin terciptanya kelestarian lingkungan secara global [1]. Indonesia yang dirasa masih kesulitan untuk mencapai pemenuhan akan kebutuhan air bersih. Maka dari itu dibuatlah kesepakatan baru sebagai pengganti MDGs yang berlaku tahun 2015-2030 yaitu Sustainable Development Goals atau SDGs, dimana menjadi acuan bagi pembangunan negara, baik untuk negara maju ataupun negara berkembang guna menjawab ketertinggalan pembangunan negara salah satunya dengan tercapainya ketersediaan air bersih bagi masyarakat, meskipun hal tersebut masih dianggap tidak

mudah untuk dicapai [2]. Indonesia yang merupakan negara berkembang juga menggunakan program ini untuk memenuhi ketersediaan air bersih, namun di Indonesia masih banyak pulau-pulau yang kekurangan air bersih diantaranya pulau Jawa, Sulawesi, Bali, dan Nusa Tenggara. Dari keempat pulau tersebut Jawa memiliki krisis ketersediaan air yang paling tinggi, berdasarkan Badan Penanggulangan Bencana Daerah (BPBD) Jawa Timur memiliki krisis ketersediaan air paling tinggi dibandingkan Jawa Barat dan Jawa Tengah. Sebanyak 25 daerah di Jawa Timur yang mengalami krisis ketersediaan air bersih.

Survei Sosial Ekonomi Nasional [3] menyebutkan jenis air bersih yang dapat digunakan dalam memenuhi konsumsi kebutuhan air antara lain air kemasan bermerek, air isi ulang, leding meteran, leding eceran, sumur bor/pompa, sumur terlindungi, sumur tak terlindungi, mata air terlindungi, mata air tak terlindungi, air permukaan yang terdiri dari sungai, danau, waduk, kolam dan irigasi, serta air hujan. Masyarakat dunia pasti mengkonsumsi air bersih dalam kebutuhan sehari-harinya dimana dari jenis air yang disebutkan oleh Survei Sosial Ekonomi Nasional (SUSENAS) dapat dikategorikan berdasarkan cara memperolehnya, yaitu baik dengan cara membeli atau tidak membeli. Jika suatu rumah tangga dalam memenuhi konsumsi air bersih dengan cara membeli maka akan ada pengeluaran untuk konsumsi air tersebut sedangkan jika tidak membeli mungkin rumah tangga tersebut memiliki sumber air tersendiri yang tidak perlu adanya pengeluaran. Dalam pemenuhan kebutuhan air bersih tersebut pasti memiliki faktor yang mempengaruhinya. Dari pernyataan tentang cara memperoleh air bersih terdapat suatu bahasan yang menarik untuk diteliti, yaitu tentang pola konsumsi rumah tangga dalam memenuhi kebutuhan air bersih serta ingin mengetahui karakteristik dan model dari pola konsumsi tersebut dengan faktor-faktor yang diduga mempengaruhi.

Menurut Mustikowati [4] dalam penelitiannya menjelaskan bahwa faktor yang mempengaruhi konsumsi air antara lain pendapatan keluarga, jumlah anggota keluarga, kepemilikan tempat tinggal, kepemilikan sumber air lain diluar PDAM dan kualitas air PDAM. Lalu pada penelitian Winarna [5] dalam penelitiannya faktor yang mempengaruhi konsumsi air bersih antara lain jumlah anggota keluarga, serta tingkat pendidikan kepala rumah tangga. Berdasarkan penelitian sebelumnya, faktor yang diduga mempengaruhi pola konsumsi rumah tangga dalam memenuhi kebutuhan air bersih antara lain klasifikasi perkotaan atau pedesaan,

kepemilikan tempat tinggal, pendidikan terakhir kepala rumah tangga, pendapatan keluarga, jenis kegiatan pemakaian air, penyaluran air yang digunakan, jarak ke sanitasi, dan jumlah anggota keluarga.

Dimana pada penelitian ini, variabel sumber air bersih menjadi variabel respon dengan kategori berdasarkan sumber air bersihnya sedangkan variabel prediktor sebagai faktor-faktor yang mempengaruhinya terdiri dari klasifikasi perkotaan atau pedesaan, kepemilikan tempat tinggal, pendidikan terakhir kepala rumah tangga, biaya pengeluaran keluarga, penyaluran air yang digunakan dan jumlah anggota keluarga, dimana semua data merupakan data kategorik kecuali variabel jumlah anggota keluarga dan biaya pengeluaran keluarga merupakan data rasio. Kemudian dari variabel-variabel diatas dapat didefinisikan bagaimana karakteristik pola konsumsi rumah tangga tersebut dalam memenuhi kebutuhan air bersih dengan menggunakan statistika deskriptif serta sebagai capaian dari penelitian ini adalah terbentuknya model antara sumber air bersih rumah tangga di Jawa Timur dalam memenuhi kebutuhan air bersih dengan faktor-faktor yang mempengaruhinya, karena pada variabel respon dikategorikan menjadi 4 berdasarkan sumber air bersih maka pada penelitian ini menggunakan metode regresi logistik multinomial. Menurut Afidah [6] dalam penelitiannya memberikan hasil bahwa dengan regresi logistik multinomial diperoleh nilai akurasi sebesar 47,7% dalam mempresentasikan model yang dikarenakan prediktor yang kurang mewakili pola model, sedangkan menurut Parlinggoman [7] pada penelitiannya memberikan hasil dengan regresi logistik multinomial diperoleh nilai akurasi sebesar 83,1%.

Berdasarkan uraian tersebut, penelitian ini untuk mengetahui karakteristik dan model analisis pada faktor-faktor yang mempengaruhi sumber air bersih rumah tangga di Jawa Timur dengan menggunakan analisis regresi logistik multinomial.

## II. TINJAUAN PUSTAKA

### A. Statistika Deskriptif

Statistika deskriptif merupakan bagian statistik yang membahas tentang metode untuk menyajikan data sehingga menarik dan informatif. Pada statistika deskriptif terdapat beberapa jenis ukuran data diantaranya ukuran pemusatan data, ukuran penyebaran data, ukuran kemiringan dan ukuran kecondongan [8]. Penyajian data berupa tabel, grafik, diagram, dan besaran-besaran seperti modus dan median.

### B. Uji Independensi

Uji independensi bertujuan untuk mengetahui adanya hubungan antar variabel. Uji independensi dilakukan untuk mengetahui ada atau tidaknya hubungan antar variabel. Pengujian tersebut dilakukan dengan menggunakan uji Chi-Square, berikut hipotesis yang digunakan [9].

$H_0$  : Tidak ada hubungan antar variabel

$H_1$  : Terdapat hubungan antar variabel

Statistik Uji :

$$\chi^2 = \sum_{i=1}^a \sum_{j=1}^b \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \quad (1)$$

Daerah Penolakan : Tolak  $H_0$ , jika  $\chi^2 > \chi_{\alpha, (i-1)(j-1)}^2$

### C. Regresi Logistik Multinomial

Regresi logistik multinomial merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat polikotomus atau multinomial [10]. Variabel respon y terdiri lebih dari 2 kategori yang biasanya y dinotasikan dengan 0,1, atau 2. Hosmer dan Lemeshow [10] menjelaskan bahwa model yang digunakan pada regresi logistik multinomial adalah sebagai berikut.

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_b x_b)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_b x_b)} \quad (2)$$

#### 1. Estimasi Parameter

Estimasi parameter regresi logistik multinomial menggunakan metode Maximum Likelihood Estimation (MLE). Berikut fungsi likelihood.

$$l(\beta) = \prod_{i=1}^n \pi_{1i}^{y_{1i}} \pi_{2i}^{y_{2i}} \pi_{3i}^{y_{3i}} (1 - \pi_{1i} - \pi_{2i} - \pi_{3i})^{1 - y_{1i} - y_{2i} - y_{3i}} \quad (3)$$

Dilakukan transformasi logaritma, sehingga didapat persamaan logaritma likelihood dengan fungsi linier.

$$L(\beta) = \log [l(\beta)] = \sum_{i=1}^n (y_{1i} x_i \beta_1 + y_{2i} x_i \beta_2 + y_{3i} x_i \beta_3 + \ln(1 - \pi_{1i} - \pi_{2i} - \pi_{3i})) \quad (4)$$

Dengan melakukan penurunan dari  $L(\beta)$  terhadap  $\beta_j$ , maka diperoleh nilai  $\beta$  sebagaimana terdapat pada persamaan berikut.

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^n x_i (y_{ji} - \pi_{ji}) \quad (5)$$

#### 2. Pengujian Serentak

Pengujian secara serentak digunakan untuk mengetahui pengaruh variabel prediktor dalam model secara bersama-sama. Berikut hipotesis yang digunakan.

$H_0$ :  $\beta_1 = \beta_2 = \dots = \beta_b = 0$  (Tidak ada pengaruh variabel prediktor terhadap model)

$H_1$  : Minimal terdapat satu  $\beta_j \neq 0, j = 1, 2, \dots, b$

Statistik Uji :

$$G = -2 \ln \left[ \frac{\binom{n_1}{n}^{n_1} \binom{n_2}{n}^{n_2} \binom{n_3}{n}^{n_3}}{\prod_{j=1}^b \pi_1(x)^{y_{1j}} \pi_2(x)^{y_{2j}} \pi_3(x)^{y_{3j}}} \right] \quad (6)$$

Statistik uji G mengikuti distribusi Chi-Square, sehingga untuk memperoleh keputusan dilakukan perbandingan dengan  $\chi_{\alpha, v}^2$  [10]. Kriteria penolakan (Tolak  $H_0$ ) jika nilai  $G > \chi_{\alpha, v}^2$  dimana derajat bebas = v (banyaknya variabel prediktor).

#### 3. Pengujian Parsial

Pengujian parsial dilakukan untuk mengetahui apakah variabel prediktor berpengaruh signifikan atau tidak terhadap variabel respon. Uji ini dimaksudkan untuk melihat apakah suatu variabel prediktor layak masuk dalam model [1]. Berikut hipotesisnya

$H_0$ :  $\beta_j = 0$

$H_1$ :  $\beta_j \neq 0, j = 1, 2, \dots, b$

Statistik Uji :

$$W^2 = \frac{\hat{\beta}_j^2}{SE(\hat{\beta}_j^2)} \quad (7)$$

Dengan  $\widehat{SE}(\hat{\beta}_j^2)$  merupakan standar error koefisien dan  $\hat{\beta}_j$  merupakan nilai koefisien dugaan variabel prediktor  $W^2$ . Statistik uji  $W^2$  disebut sebagai statistik uji Wald, mengikuti distribusi *chi-square* sehingga  $H_0$  ditolak apabila  $W^2 > \chi_{(\alpha, v)}^2$  atau P-value  $< \alpha$ , dimana  $v$  adalah banyaknya prediktor.

4. Kesesuaian Model

Statistik uji yang digunakan untuk menguji kesesuaian model regresi logistik yaitu *Goodness of Fit* dengan hipotesis sebagai berikut [10].

$H_0$  : Model sesuai (tidak terdapat perbedaan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

$H_1$  : Model tidak sesuai (terdapat perbedaan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

Statistik uji :

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n_k \hat{\pi}_k)^2}{n_k \hat{\pi}_k (1 - \hat{\pi}_k)} \quad (8)$$

Kriteria penolakan (Tolak  $H_0$ ) adalah jika  $\hat{C} > \chi_{\alpha, v}^2$ , yang menyatakan bahwa model tidak sesuai (terdapat perbedaan antara hasil pengamatan dengan kemungkinan hasil prediksi model).

5. Pengukuran Akurasi Klasifikasi

Akurasi klasifikasi pada data imbalance dapat dihitung menggunakan geometrik mean (G-mean) dan Area Under ROC Curve (AUC). G-mean untuk kasus mengelompokkan dengan jumlah kategori lebih dari dua, dengan menggunakan rumus Expanding G-mean sebagai berikut [11].

$$G_{mean} = \left( \prod_{i=1}^a R_i \right)^{1/a} \quad (9)$$

$$AUC = \frac{1}{c} \sum_{i=1}^a R_i ; R_i = \frac{n_{ii}}{\sum_{l=1}^c n_{il}} ; i = 1, 2, \dots, a$$

Keterangan:

$n_{ii}$  : jumlah prediksi  $\hat{\pi}_i$  yang tepat diklasifikasikan ke  $\pi_i$

$n_{il}$  : jumlah prediksi  $\hat{\pi}_i$  yang salah diklasifikasikan ke  $\pi_l$

6. Odds Ratio

Interpetasi dalam regresi logistik menggunakan nilai odds ratio yang menunjukkan perbandingan berapa kali lipat kenaikan atau penurunan angka kejadian  $Y = i$  terhadap  $Y = 0$  sebagai kategori pembanding jika nilai variabel prediktor (x) berubah sebesar nilai tertentu [10].

$$\psi_1 = \frac{\pi_1(1)/\pi_0(1)}{\pi_1(0)/\pi_0(0)} \quad (10)$$

$$\psi_2 = \frac{\pi_2(1)/\pi_0(1)}{\pi_2(0)/\pi_0(0)}$$

Jika  $1 < \Psi$  menunjukkan bahwa antar kedua variabel terdapat hubungan negatif dan jika  $1 > \Psi$  menunjukkan bahwa antar kedua variabel terdapat hubungan positif.

D. Faktor-faktor yang Mempengaruhi Pemenuhan Kebutuhan Konsumsi Air Bersih

Kebutuhan air adalah banyaknya jumlah air yang dibutuhkan untuk keperluan rumah tangga, industri, penggelontoran kota [12] serta menjadi prioritas kebutuhan air domestik, industri, pelayanan umum dan kebutuhan air untuk mengganti kebocoran [14] dan untuk keperluan

sehari-hari. Konsumsi air bersih dipengaruhi oleh beberapa faktor [15]. Penelitian yang terkait dengan faktor-faktor yang mempengaruhi konsumsi air bersih telah dilakukan oleh Sri Winarna [5] dalam penelitiannya faktor yang mempengaruhi konsumsi air bersih antara lain yaitu jumlah anggota keluarga, serta tingkat pendidikan kepala rumah tangga. Menurut Hestin Mutmainah [16] faktor yang mempengaruhi pola konsumsi air bersih yaitu jumlah anggota keluarga serta kepemilikan sumur, Penyaluran air dan menurut Widayanti Mustikowati [4] dalam penelitiannya menjelaskan bahwa faktor yang mempengaruhi konsumsi air antara lain pendapatan keluarga, jumlah anggota keluarga, kepemilikan tempat tinggal, kepemilikan sumber air lain diluar PDAM dan kualitas air PDAM.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder, diambil dari Survey Ekonomi Sosial (SUSENAS) yang diperoleh di Badan Pusat Statistik (BPS) Propinsi Jawa Timur, dengan data SUSENAS tahun 2016. Unit pengamatan adalah rumah tangga.

B. Variabel Penelitian

Variabel penelitian yang digunakan pada penelitian, adalah sebagai berikut.

Tabel 1.

Variabel Prediktor Penelitian		
Variabel	Uraian	Skala
X <sub>1</sub>	Klasifikasi Desa/Kota	Kategorik
X <sub>2</sub>	Pend. Terakhir Kepala Rumah Tangga	Kategorik
X <sub>3</sub>	Status Kepemilikan Bangunan	Kategorik
X <sub>4</sub>	Penyalura Air yang Digunakan	Kategorik
X <sub>5</sub>	Jumlah Anggota Keluarga	Rasio

Berikut ini penjelasan mengenai masing-masing variabel yang digunakan dalam penelitian ini Variabel X<sub>1</sub> dengan kategori 1 adalah Desa dan kategori 2 adalah Kota. Variabel X<sub>2</sub> dengan kategori 1 adalah SD, kategori 2 adalah SMP, kategori 3 adalah SMA, dan kategori 4 adalah Perguruan Tinggi. Variabel X<sub>3</sub> dengan kategori 1 adalah Bukan Milik Sendiri dan kategori 2 adalah Milik Sendiri. Variabel X<sub>4</sub> dengan kategori 1 adalah Perpipaan/Hidran Umum/Terminal Air, kategori 2 adalah Tidak Semua. Variabel respon dengan 4 kategori disajikan pada Tabel 2.

Tabel 2.

Variabel Respon Penelitian		
Kategori	Keterangan	Jumlah (RT)
1	Sumber air bersih seluruhnya membeli	4.630
2	Sebagian besar sumber air bersih membeli	1.821
3	Sebagian kecil sumber air bersih membeli	4.358
4	Sumber air bersih seluruhnya tidak membeli	18.668

C. Langkah Analisis

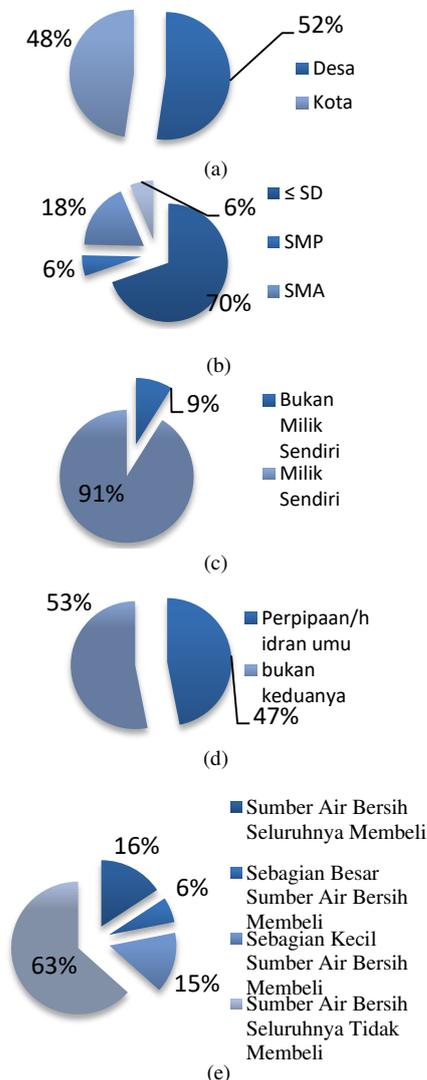
Langkah-langkah yang dilakukan untuk menganalisis dalam penelitian ini adalah sebagai berikut.

1. Mendeskripsikan karakteristik pola konsumsi rumah tangga di Jawa Timur dalam memenuhi kebutuhan air bersih maka digunakan diagram-diagram untuk variabel kategorik dan variabel rasio.
2. Melakukan analisis model antara pola konsumsi rumah tangga di Jawa Timur dalam memenuhi kebutuhan air bersih dengan faktor-faktor yang mempengaruhi dengan langkah sebagai berikut.
  - a. Menguji independensi semua variabel prediktor terhadap variabel respon.
  - b. Melakukan analisis regresi logistik multinomial.
  - c. Melakukan uji serentak pada variabel prediktor yang mempunyai hubungan dengan variabel respon dan uji parsial pada variabel prediktor yang mempunyai hubungan dengan variabel respon.
  - d. Melakukan uji kesesuaian model
  - e. Menghitung ketepatan klasifikasi model
  - f. Membentuk fungsi logit pada masing-masing kategori respon di setiap model dan Menginterpretasikan model regresi logistik biner dan odds ratio yang diperoleh.

IV. PEMBAHASAN

A. Analisis Statistik Deskriptif

Pada poin ini dilakukan analisis deskriptif terhadap variabel kategorik dan variabel numerik. Statistika deskriptif untuk variabel kategorik disajikan pada Gambar 1.



Gambar 1. Statistika Deskriptif Variabel Kategorik.

Berdasarkan Gambar 1 ,(a) merupakan variabel klasifikasi desa/kota dengan sebanyak 52% rumah tangga berada pada kategori desa, sedangkan pada kategori kota sebanyak 48% rumah tangga berada pada kategori kota. Gambar (b) merupakan variabel pendidikan terakhir kepala rumah tangga dengan sebanyak 70% kepala rumah tangga pendidikan terakhirnya adalah sekolah dasar (SD), kepala rumah tangga dengan pendidikan terakhir SMP sebanyak 6%, kepala rumah tangga dengan pendidikan terakhir SMA sebanyak 18%, dan pendidikan terakhir kepala rumah tangga yaitu perguruan tinggi sebesar 6%. Pada gambar (c) yang merupakan variabel status kepemilikan bangunan dengan sebanyak 91% rumah tangga merupakan bangunan yang berstatus milik sendiri, sedangkan sisanya sebanyak 9% bangunan rumah tangga memiliki status bukan milik sendiri. Lalu gambar (d) merupakan variabel penyaluran air yang digunakan dengan sebanyak 47% rumah tangga menggunakan perpipaan/hidran umum dalam menyalurkan sumber air bersih menuju rumah masing-masing dan sebanyak 53% rumah tangga tidak menggunakan perpipaan/hidran umum atau tidak menggunakan penyaluran air dalam menyalurkan sumber air bersih menuju rumah masing-masing. Gambar (e) merupakan variabel respon dimana kategori sumber air bersih seluruhnya membeli sebesar 16%, kategori sebagian besar sumber air bersih membeli sebesar 6%, kategori sebagian kecil sumber air bersih membeli sebesar 15%, dan kategori sumber air bersih seluruhnya tidak membeli sebesar 63%.

Tabel 3.  
Statistika Deskriptif Variabel Numerik

Variabel	Mean	Minimum	Maximum
X5	3,4558	1	14

Tabel 3, menunjukkan bahwa variabel X5 atau jumlah anggota keluarga di Jawa Timur rata-rata memiliki anggota keluarga sebanyak 3,4558 atau sebanyak 4 anggota keluarga dengan minimum 1 anggota keluarga dalam keluarga dan maximum 14 anggota keluarga dalam keluarga.

B. Uji Independensi

Pada subbab ini dijelaskan mengenai analisis yang dilakukan untuk menentukan model regresi logistik multinomial pada analisis sumber air bersih rumah tangga di Jawa Timur. Sebelum melakukan analisis pemodelan, terlebih dahulu dilakukan uji independensi variabel yang disajikan pada Tabel 4 dan Tabel 5.

Tabel 4.  
Chi-Square Test Variabel Kategorik

		Y				Total	Chi-Square Test
		1	2	3	4		
X1	1	3501	1472	2868	7583	15424	0,00
	2	1129	349	1490	11085	14053	
	3	1237	555	1168	2438	5398	
	4	720	212	474	464	1870	
X2	1	2474	937	2487	14645	20543	0,00
	2	199	117	229	121	1666	
	3	1237	555	1168	2438	5398	
	4	720	212	474	464	1870	
X3	1	710	387	587	935	2619	0,00
	2	3920	1434	3771	17733	26858	
X4	1	4438	940	1570	6876	13824	0,00
	2	192	881	2788	11792	15653	

Hasil analisis hubungan pada variabel kategorik terhadap variabel respon (Y).

Klasifikasi desa/kota ( $X_1$ ) – Y,

$H_0$  : Tidak ada hubungan antar variabel

$H_1$  : Ada hubungan antar variabel

Berdasarkan Tabel 4 nilai chi-square test bernilai 0,000, dimana nilai tersebut lebih kecil dari alfa 5%. Maka, keputusannya adalah tolak  $H_0$ . Artinya terdapat hubungan antara variabel  $X_1$  dengan Y.

Hal yang sama dengan variabel pendidikan terakhir kepala rumah tangga ( $X_2$ ), Status kepemilikan bangunan ( $X_3$ ), dan Penyaluran air ( $X_4$ ) yang mana nilai chi-square test bernilai 0,000. Nilai tersebut lebih kecil dari alfa 5%. Maka, keputusan tolak  $H_0$ . Maka, variabel  $X_2$ ,  $X_3$ , dan  $X_4$  terdapat hubungan dengan Y.

Tabel 5.  
Korelasi Variabel Numerik

	$X_5$	Y
$X_5$	1	-0,019 (0,001)
Y	-0,019 (0,001)	1

Hasil analisis hubungan pada variabel kategorik terhadap variabel respon (Y).

Jumlah anggota keluarga ( $X_5$ ) – Y,

$H_0$  : Tidak ada hubungan antar variabel

$H_1$  : Ada hubungan antar variabel

Berdasarkan Tabel 5 nilai signifikansi dari korelasi bernilai 0,001, dimana nilai tersebut lebih kecil dari alfa 5%. Maka, keputusannya adalah tolak  $H_0$ . Artinya terdapat hubungan antara variabel  $X_5$  dengan Y.

### C. Uji Kesesuaian Model

Sebelum melakukan analisis regresi logistik multinomial lebih lanjut, diuji dahulu apakah model yang terbentuk sudah sesuai dengan data (fit).

Tabel 6.  
Uji Kesesuaian Model

	Chi-Square	df	Sig.
Pearson	1150,301	765	0,000
Deviance	1132,358	765	0,000

Berdasarkan hasil pada Tabel 6 didapat nilai p-value (sig.) di baris hasil uji kelayakan metode Pearson yaitu sebesar 0,000. Nilai tersebut menyatakan bahwa uji kelayakan metode Pearson kurang dari alfa (5% = 0,05), maka tolak  $H_0$  yang menjelaskan bahwa model tidak fit atau model tidak sesuai. Artinya model tidak dapat menjelaskan data.

Tabel 7.  
Nilai Pseudo R-Square

Cox and Snell	0,309
Nagelkerke	0,354
McFadden	0,179

Selain uji fit model, maka perlu memperhitungkan besarnya ukuran kebaikan model pada variabel bebas kualitatif. Berdasarkan Tabel 7 nilai tersebut diambil dari nilai Nagelkerke yaitu sebesar 0,354. Nilai tersebut mengindikasikan bahwa keragaman data variabel bebas mampu menjelaskan keragaman data variabel respon sebesar 35,4% sedangkan sisanya dijelaskan oleh variabel bebas lain yang ada di luar model. Nilai 0,354 tersebut sudah dianggap baik dalam menjelaskan nilai ukuran kebaikan model.

### D. Pengujian Signifikansi Parameter

#### 1. Uji Serentak

Pengujian ini dilakukan untuk memeriksa koefisien  $\beta$  secara serentak atau bersamaan terhadap variabel respon.

Tabel 8.  
Model Fitting Information

Model	Model Fitting Criteria		Likelihood Ratio Test	
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept				
Only	13795,265			
Final	2891,955	10903,310	21	0,000

Berdasarkan hasil uji serentak di Tabel 8 menunjukkan nilai pada baris final dengan nilai chi-square sebesar 10903,310 dengan nilai chi-square tabel sebesar 11,070 yang artinya minimal terdapat satu variabel bebas yang secara statistik signifikan, selain itu nilai p-value (sig.) sebesar 0,000. Dimana nilai tersebut lebih kecil dari alfa 5% atau 0,05, maka tolak  $H_0$  yang artinya minimal terdapat satu variabel bebas yang secara statistik signifikan memengaruhi variabel respon. Karena, pada uji serentak menyatakan signifikan, maka dilanjutkan ke uji parsial.

#### 2. Uji Parsial

Pengujian ini dilakukan untuk memeriksa kemaknaan koefisien  $\beta$  secara parsial. Dimana sebelumnya telah terlihat hasil uji serentak yang menjelaskan bahwa seluruh variabel prediktor berpengaruh terhadap variabel respon. Pada uji parsial akan dilihat pengaruh tiap variabel prediktor terhadap variabel respon.

Tabel 9.  
Uji Parsial

Effect	Model Fitting Criteria		Likelihood Ratio Tests	
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
$X_1$	3113,854	3060,166	3	0,000
$X_2$	2640,951	2545,250	9	0,000
$X_3$	1013,137	960,878	3	0,000
$X_4$	6416,345	6363,723	3	0,000
$X_5$	298,701	26,846	3	0,000

Berdasarkan Tabel 9, uji parsial menunjukkan nilai chi-square pada kelima variabel bebas lebih besar dari nilai chi-square tabel yaitu 11,070 yang artinya kelima variabel bebas yang digunakan secara statistik memiliki pengaruh yang signifikan terhadap variabel respon, selain itu pada kolom signifikan untuk semua variabel bernilai sebesar 0,000. Nilai-nilai tersebut lebih kecil dari alfa 5% atau lebih kecil dari 0,05, maka tolak  $H_0$  yang artinya variabel bebas secara statistik signifikan memengaruhi variabel respon. Dengan demikian, kelima variabel bebas yang digunakan secara statistik memiliki pengaruh yang signifikan terhadap variabel respon.

### E. Interpretasi Analisis Regresi Logistik Multinomial

Parameter pembentuk fungsi logit disajikan pada Tabel 10 yang didapat dari nilai Beta, sedangkan nilai odds ratio disajikan pada kolom  $\exp(\text{Beta})$  pada Tabel 11.

Tabel 10.  
Parameter Pembentuk Fungsi Logit

Ya	Beta		
	1	2	3
Konstanta	-3,605	-2,633	-0,811
$X_5$	0,004	0,021	0,039
$[X_1=1]$	1,328	1,550	0,798

[X <sub>1</sub> =2]			
[X <sub>2</sub> =1]	-1,637	-1,423	-1,506
[X <sub>2</sub> =2]	-1,731	-1,003	-1,335
[X <sub>2</sub> =3]	-0,920	-0,564	-0,694
[X <sub>2</sub> =4]			
[X <sub>3</sub> =1]	0,724	1,054	0,665
[X <sub>3</sub> =2]			
[X <sub>4</sub> =1]	3,703	0,632	-0,026
[X <sub>4</sub> =2]			

Berdasarkan Tabel 10, menghasilkan fungsi logit, sehingga didapatkan pula model regresi logistik.

$$\pi_1(x) = \frac{\exp P_1(x)}{1 + \exp P_1(x) + \exp P_2(x) + \exp P_3(x)}$$

$$\pi_2(x) = \frac{\exp P_2(x)}{1 + \exp P_1(x) + \exp P_2(x) + \exp P_3(x)}$$

$$\pi_3(x) = \frac{\exp P_3(x)}{1 + \exp P_1(x) + \exp P_2(x) + \exp P_3(x)}$$

Dengan nilai P1, P2, dan P3 sebagai berikut.

$$P_1(x) = \ln\left(\frac{P(Y = 1)1 | x}{P(Y = 1)0 | x}\right)$$

$$= -3,605 + 1,328X_{1.1} - 1,637X_{2.1} - 1,731X_{2.2} + (-0,920X_{2.3}) + 0,724X_{3.1} + 3,703X_{4.1} + 0,004X_5$$

$$P_2(x) = \ln\left(\frac{P(Y = 1)2 | x}{P(Y = 1)0 | x}\right)$$

$$= -2,633 + 1,550X_{1.1} - 1,423X_{2.1} - 1,003X_{2.2} + (-0,564X_{2.3}) + 1,054X_{3.1} + 0,632X_{4.1} + 0,021X_5$$

$$P_3(x) = \ln\left(\frac{P(Y = 1)3 | x}{P(Y = 1)0 | x}\right)$$

$$= -0,811 + 0,798X_{1.1} - 1,506X_{2.1} - 1,335X_{2.2} + (-0,694X_{2.3}) + 0,665X_{3.1} - 0,026X_{4.1} + 0,039X_5$$

Tabel 11.

Interpretasi Analisis Regresi Logistik Multinomial

Variabel	Y		
	Sumber Air Bersih Seluruhnya Membeli - Sumber Air Bersih Seluruhnya Tidak Membeli	Sebagian Besar Sumber Air Bersih Membeli - Sebagian Besar Sumber Air Bersih Tidak Membeli	Sebagian Kecil Sumber Air Bersih Membeli - Sebagian Kecil Sumber Air Bersih Tidak Membeli
	Exp(B)	Exp(B)	Exp(B)
X <sub>5</sub>	1,004	1,021	1,039
X <sub>1.1</sub>	3,772	4,701	2,221
X <sub>1.2</sub>			
X <sub>2.1</sub>	0,195	0,241	0,222
X <sub>2.2</sub>	0,177	0,367	0,263
X <sub>2.3</sub>	0,398	0,569	0,499
X <sub>2.4</sub>			
X <sub>3.1</sub>	2,063	2,869	1,944
X <sub>3.2</sub>			
X <sub>4.1</sub>	40,555	1,882	0,975
X <sub>4.2</sub>			

Berdasarkan Tabel 11, pada variabel X<sub>1</sub> atau klasifikasi desa/kota dengan kategori kota sebagai kontrol, menunjukkan bahwa rumah tangga di daerah pedesaan memiliki kecenderungan resiko sumber air bersih seluruhnya membeli 1,004 lebih besar dibandingkan sumber air bersih seluruhnya tidak membeli, lalu untuk sebagian besar sumber air bersih membeli 1,021 lebih besar dibanding tidak membeli dan sebagian kecil sumber air bersih membeli 1,039 lebih besar dibandingkan sumber air bersih seluruhnya tidak membeli.

Pada variabel X<sub>2</sub> atau pendidikan terakhir kepala rumah tangga dengan kategori perguruan tinggi sebagai kontrol menunjukkan bahwa rumah tangga dengan pendidikan terakhir kepala rumah tangga yaitu ≤ SD memiliki kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,195 lebih kecil dibanding sumber air bersih seluruhnya membeli, lalu kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,241 lebih kecil dibandingkan sebagian besar sumber air bersih membeli dan kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,222 lebih kecil dibandingkan sebagian kecil sumber air bersih membeli. Rumah tangga dengan pendidikan terakhir kepala rumah tangga yaitu SMP memiliki kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,177 lebih kecil dibanding sumber air bersih seluruhnya membeli, lalu kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,367 lebih kecil dibandingkan sebagian besar sumber air bersih membeli dan kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,263 lebih kecil dibandingkan sebagian kecil sumber air bersih membeli. Rumah tangga dengan pendidikan terakhir kepala rumah tangga yaitu SMA memiliki kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,398 lebih kecil dibanding sumber air bersih seluruhnya membeli, lalu kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,569 lebih kecil dibandingkan sebagian besar sumber air bersih membeli dan kecenderungan resiko sumber air bersih seluruhnya tidak membeli 1/0,499 lebih kecil dibandingkan sebagian kecil sumber air bersih membeli.

Pada variabel X<sub>3</sub> atau status kepemilikan bangunan dengan kategori bangunan milik sendiri sebagai kontrol menunjukkan bahwa rumah tangga dengan status bangunan bukan milik sendiri memiliki kecenderungan resiko sumber air bersih seluruhnya membeli 2,063 lebih besar dibandingkan sumber air bersih seluruhnya tidak membeli, lalu kecenderungan resiko sebagian besar sumber air bersih membeli 2,869 lebih besar dibandingkan sumber air bersih seluruhnya tidak membeli dan kecenderungan resiko sebagian kecil sumber air bersih membeli 1,944 lebih besar dibandingkan sumber air bersih seluruhnya tidak membeli.

Pada variabel X<sub>4</sub> atau penyaluran air yang digunakan dengan kategori tidak menggunakan penyaluran air sebagai kontrol menunjukkan bahwa rumah tangga dengan penyaluran air menggunakan perpipaan/hidran memiliki kecenderungan resiko sumber air bersih seluruhnya membeli 40,555 lebih besar dibandingkan sumber air bersih seluruhnya tidak membeli dan kecenderungan resiko

sebagian besar sumber air bersih membeli 1,882 lebih besar dibandingkan sumber air bersih seluruhnya membeli.

Pada variabel  $X_5$  atau jumlah anggota keluarga dengan jumlah sedikit atau banyak anggota keluarga memiliki kecenderungan resiko sumber air bersih seluruhnya membeli 1,039 lebih besar dibandingkan sebagian kecil sumber air bersih membeli.

F. Ketepatan Klasifikasi Model

Ketepatan klasifikasi untuk mengetahui atau mengevaluasi model yang telah terbentuk.

Tabel 12. Ketepatan Klasifikasi Model

Observed	Predicted				Percent Correct
	1	2	3	4	
1	1975	0	32	2623	42,7%
2	459	0	139	1223	0,0%
3	544	0	337	3477	7,7%
4	841	0	308	17519	93,8%
Overall Percentage	13,0%	0,0%	2,8%	84,3%	67,3%

Berdasarkan Tabel 12, ketepatan klasifikasi pada tiap kategori yaitu kategori 1 sebanyak 1975 observasi, kategori 2 sebanyak 0 observasi, katetgori 3 sebanyak 337 observasi, dan kategori 4 sebanyak 17519 observasi. Akurasi model yang didapatkan adalah sebesar 67,3%.

IV. KESIMPULAN

Berdasarkan analisis yang telah dilakukan diperoleh kesimpulan bahwa Proporsi rumah tangga dari variabel bebas yang dilihat dari Pie Chart, dimana pada variabel klasifikasi desa/kota sebanyak 52% terletak didaerah pedesaan dan 48% terletak didaerah perkotaan. Variabel pendidikan terakhir kepala rumah tangga sebanyak 70% yaitu SD, 6% yaitu SMP, 18% yaitu SMA, dan 6% yaitu perguruan tinggi. Variabel status kepemilikan bangunan, sebanyak 91% bangunan rumah tangga berstatus milik sendiri dan 9% berstatus bukan milik sendiri. Variabel penyaluran air yang digunakan, sebanyak 47% rumah tangga menggunakan perpipaan/hidran umum dan sebanyak 53% rumah tangga tidak menggunakan penyaluran air dengan perpipaan atau hidran umum. Variabel jumlah anggota keluarga di Jawa Timur rata-rata memiliki anggota keluarga sebanyak 3,4558 atau sebanyak 4 anggota keluarga dengan minimum 1 anggota keluarga dalam keluarga dan maximum 14 anggota keluarga dalam keluarga. Variabel respon yang terbagi menjadi empat kategori, sebanyak 16% rumah tangga sumber air bersih seluruhnya membeli, sebanyak 6% rumah tangga sebagian besar sumber air bersih membeli, sebanyak 15% rumah tangga sebagian kecil sumber air bersih membeli, dan sebanyak 63% rumah tangga sumber air bersih seluruhnya tidak membeli.

Model yang terbentuk untuk analisis regresi logistik multinomial pada sumber air bersih rumah tangga di Jawa Timur adalah sebagai berikut.

$$P_1(x) = \ln \left( \frac{P(Y = 1)1 | x}{P(Y = 1)0 | x} \right) = -3,605 + 1,328 \text{desa} - 1,637 \text{SD} - 1,731 \text{SMP} + (-0,920 \text{SMA}) + 0,724 \text{bukan milik sendiri} + 3,703 \text{ perpipaan / hidran umum} + 0,004 \text{ jumlah anggota keluarga}$$

$$P_2(x) = \ln \left( \frac{P(Y = 1)2 | x}{P(Y = 1)0 | x} \right) = -2,633 + 1,550 \text{desa} - 1,423 \text{SD} - 1,003 \text{SMP} + (-0,564 \text{SMA}) + 1,054 \text{bukan milik sendiri} + 0,632 \text{ perpipaan / hidran umum} + 0,021 \text{ jumlah anggota keluarga}$$

$$P_3(x) = \ln \left( \frac{P(Y = 1)3 | x}{P(Y = 1)0 | x} \right) = -0,811 + 0,798 \text{desa} - 1,506 \text{SD} - 1,335 \text{SMP} + (-0,694 \text{SMA}) + 0,665 \text{bukan milik sendiri} + (-0,026 \text{ perpipaan / hidran umum}) + 0,039 \text{ jumlah anggota keluarga}$$

Berdasarkan analisis dan pembahasan serta kesimpulan yang diperoleh dapat dirumuskan saran, bahwa dari pembahasan diketahui bahwa rumah tangga lebih cenderung untuk membeli sumber air bersih, maka dari pemerintah dapat mengalokasikan dan mendistribusikan air bersih yang cukup untuk masyarakat diluar air bersih yang dapat diperoleh dengan tidak membeli.

Lalu pertimbangan untuk penelitian selanjutnya adalah pemodelan dapat menggunakan metode yang sesuai serta dapat mengatasi imbalance data sehingga pada saat pengolahan data menghasilkan hasil akhir yang lebih baik.

DAFTAR PUSTAKA

- [1] S. P. Hasibuan, "Manajemen Sumber Daya Manusia," 2007.
- [2] ESCAP-UNDP. (2002). ESCAP-UNDP Initiative for The Achievement of Millenium Development Goals in Asia and The Pacific. Bangkok: UN ESCAP.
- [3] Giddings, B. (2002). Environment, Economy, and Society: Fitting Them together Into Sustainable Development. Sust. Dev. Sustainable Cities Research Institute.
- [4] Survei Sosial Ekonomi Nasional. (2016).
- [5] Mustikowati, W. (2014). Analisis Faktor-Faktor yang Mempengaruhi Konsumsi Air Bersih Golongan Pelanggan Rumah Tangga III Wilayah Pelayanan Cabang Timur PDAM Kota Semarang. Semarang: Universitas Diponegoro.
- [6] Winarna, S. (2003). Analisis Konsumsi Air Bersih Pelanggan Rumah Tangga Berdasarkan Faktor-Faktor Yang Mempengaruhinya (Studi Kasus pada PDAM Kabupaten Karanganyar). Semarang: Universitas Diponegoro.
- [7] Afidah, L. N. (2011). Pola Tingkat Keparahan Korban Kecelakaan Lalu Lintas Dengan Menggunakan Regresi Logistik Multinomial (Studi Kasus Kecelakaan Lalu Lintas Di Surabaya). Surabaya: Institut Teknologi Sepuluh Nopember.
- [8] Parlingoman, C. (2016). Pemodelan Faktor-Faktor Pemilihan Fasilitas Kesehatan Untuk Berobat Di Sulawesi Tenggara Tahun 2012 Menggunakan Regresi Logistik Multinomial. Surabaya: Institut Teknologi Sepuluh Nopember.

- [9] Walpole, R. E. (1995). Pengantar Statistika Edisi ke-3. Jakarta: Gramedia Pustaka Utama.
- [10] Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley and Sons, Inc.
- [11] Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley and Sons, Inc.
- [12] Bekkar, M., Djeemaa, H., & Alitouch, T. (2013). Evaluation Measures for Models Assesment over Imbalanced Data Sets. *Journal of Information Engineering and Application*, 3, 27-38.
- [13] PERPAMSI. (1994). *Diklat Tenaga Teknik Penyediaan Air Minum*. Bandung: PERPAMSI dan ITB.
- [14] Agresti, A. (2002). *Categorical Data Analysis*. New York: John Wiley and Sons, Inc.
- [15] Moegijantoro. (1996). *Kebutuhan Air*. Surabaya: PT Empat Sekawan.
- [16] Hall, & James, A. (2001). *Sistem Informasi Akutansi* (3 ed.). Jakarta: Salemba Empat.
- [17] Mutmainah, H. (2011). *Faktor-Faktor Yang Mempengaruhi Permintaan Air Bersih dari Masyarakat Terhadap Perusahaan Daerah Air Minum Kabupaten Karanganyar*.
- [18] Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2 ed.). New York: John Wiley and Sons, Inc.
- [19] Bappenas. (2004). *Laporan Perkembangan Pencapaian Tujuan Pembangunan Milenium (Millenium Development Goals)*. Jakarta: Bappenas.
- [20] Batista, G. E., Bazzan , A. L., & Monard, M. C. (2003). Balancing Training Data for Automated Annotation of Keywords : a Case Study.
- [21] Hall, & James, A. (2001). *Sistem Informasi Akutansi* (3 ed.). Jakarta: Salemba Empat.
- [22] Johnson, R. A., & Winchern, D. W. (2007). *Applied Mulivariate Statistical Analysis* (6 ed.). New Jersey: Pearson Prentice Hall.
- [23] Solberg, A., & Solberg , R. (1996). A Large-Scale Evaluation of Features for Automatic Detection of Oil Spills in ERS SAR Images. In *International Geoscience and Remote Sensing Sypmosium*, pp. 1484-1486.
- [24] Widarjono, A. (2010). *Analisis Statistika Multivariat Terapan* (1 ed.). Yogyakarta: UPP STIM YKPN.
- [25] Yu, X., Zhou, M., Chen, X., Deng, L., & Wang, L. (2017). Using Class Imbalance Learning for Cross-Company Defect Prediction.