

## SEGMENTASI KESEJAHTERAAN KELUARGA DAN IDENTIFIKASI FAKTOR YANG BERPENGARUH (STUDI KASUS : KOTA BOGOR)

Rizka Annisa Mingka<sup>1\*</sup>, Syella Zignora Limba<sup>2</sup>,  
Yeky Abil Nizar<sup>3</sup>, Ahmad Qeis Tenridapi<sup>4</sup>

<sup>1,2,3,4</sup>IPB University

Jl. Raya Dramaga, Kab. Bogor, 16680, Jawa Barat, Indonesia

### Abstrak

Penelitian ini bertujuan untuk mengidentifikasi segmentasi kesejahteraan keluarga di Kota Bogor serta faktor-faktor yang memengaruhi tingkat kesejahteraan tersebut. Pendekatan kuantitatif digunakan dengan metode K-Means Clustering untuk mengelompokkan keluarga berdasarkan kemiripan karakteristik ekonomi dan demografis, serta Regresi Logistik Multinomial untuk menganalisis pengaruh variabel independen terhadap tingkat kesejahteraan keluarga. Data dikumpulkan menggunakan teknik *two-stage cluster sampling* dari enam kecamatan di Kota Bogor, dengan total sampel sebanyak 120 rumah tangga. Hasil K-Means menunjukkan adanya tiga segmen kesejahteraan, yakni kesejahteraan rendah, kesejahteraan sedang, dan sejahtera, dengan mayoritas keluarga berada di kategori kesejahteraan sedang dan rendah. Regresi logistik mengungkapkan bahwa pengeluaran bulanan dan jumlah tanggungan berpengaruh signifikan terhadap peluang keluarga masuk dalam kategori kesejahteraan sedang. Sedangkan untuk kategori sejahtera, jumlah tanggungan dan usia kepala keluarga menjadi faktor yang signifikan. Temuan ini memberikan dasar bagi perumusan kebijakan sosial yang lebih tepat sasaran di Kota Bogor.

**Kata Kunci:** Kesejahteraan Keluarga, K-Means Clustering, Regresi Logistik Multinomial

### Abstract

*This study aims to identify household welfare segmentation in Bogor City and the factors influencing welfare levels. A quantitative approach was applied, employing K-Means Clustering to segment households based on economic and demographic characteristics, and Multinomial Logistic Regression to analyze the impact of independent variables on household welfare levels. Data were collected using a two-stage cluster sampling technique from six sub-districts in Bogor City, with a total sample of 120 households. The K-Means results revealed three welfare segments—low, medium, and high—with the majority of households falling into the medium and low welfare categories. The logistic regression analysis indicated that monthly expenditure and number of dependents significantly influence the likelihood of households belonging to the medium welfare category. For the high welfare category, number of dependents and the age of the household head were significant factors. These findings provide a basis for more targeted social policy formulation in Bogor City.*

**Keywords:** Household Welfare, K-Means Clustering, Multinomial Logistic Regression.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

## 1. PENDAHULUAN

Kesejahteraan keluarga merupakan indikator utama dalam menilai keberhasilan pembangunan dan kualitas hidup masyarakat secara menyeluruh. Kesejahteraan tidak hanya mencakup aspek ekonomi, tetapi juga meliputi dimensi pendidikan, kesehatan, pemenuhan kebutuhan dasar, serta kondisi tempat tinggal [1]. Pendapatan sering digunakan sebagai tolok ukur utama, namun studi empiris menunjukkan bahwa kesejahteraan tidak semata ditentukan oleh besaran pendapatan. Pendapatan keluarga memang memainkan peran penting dalam menentukan kemampuan rumah tangga memenuhi kebutuhan tersebut, namun tidak dapat berdiri sendiri sebagai penentu kesejahteraan [2].

Beberapa penelitian menunjukkan bahwa hubungan antara pendapatan dan kesejahteraan bersifat kompleks dan tidak selalu linier. Rumah tangga dengan jumlah tanggungan yang besar memiliki beban konsumsi yang lebih tinggi, sehingga meskipun pendapatan tergolong cukup, tingkat kesejahteraannya bisa tetap rendah [1]. Nata et al. (2020) menyoroti bahwa dalam konteks rumah tangga petani pisang, meskipun pendapatan tergolong sedang, banyak rumah tangga belum mencapai status sejahtera karena beban tanggungan yang tinggi [3]. Oleh karena itu, dalam menilai kesejahteraan, penting untuk mempertimbangkan variabel jumlah tanggungan sebagai faktor penentu yang signifikan [4]. Terdapat pula temuan yang menyatakan bahwa jumlah anggota keluarga maupun pendapatan tidak secara signifikan mempengaruhi kesejahteraan [5]. Perbedaan hasil ini menunjukkan pentingnya analisis yang mempertimbangkan interaksi antar faktor, terutama dalam konteks lokal seperti Kota Bogor.

Dalam menilai hubungan tersebut, penggunaan metode statistik seperti Regresi Logistik sangat relevan karena memungkinkan analisis hubungan variabel-variabel prediktor (seperti pendapatan, jumlah tanggungan, dan usia kepala keluarga) terhadap variabel target kategorikal, yaitu status kesejahteraan. Regresi logistik tidak hanya memungkinkan estimasi peluang (probabilitas) status kesejahteraan, tetapi juga mengukur kekuatan pengaruh tiap faktor independen melalui interpretasi koefisien dan *odds ratio* [6]. Penelitian yang dilakukan oleh Permatasari dan Yuliana tahun 2021 berhasil menerapkan model ini pada data Susenas di Bali dan menemukan bahwa variabel seperti jumlah anggota rumah tangga, pendidikan, dan sektor pekerjaan kepala keluarga memiliki pengaruh signifikan terhadap kesejahteraan [7].

Lebih jauh, untuk menggali segmentasi rumah tangga berdasarkan karakteristik ekonomi, pendekatan data mining seperti *K-Means Clustering* dapat digunakan. Dalam konteks penelitian sosial ekonomi, metode statistik dan data mining sangat dibutuhkan untuk menggali pola-pola tersembunyi dalam data keluarga [8]. *K-Means Clustering* dapat mengelompokkan rumah tangga berdasarkan karakteristik tertentu seperti pendapatan dan jumlah tanggungan. *K-Means* memberikan pendekatan yang eksploratif untuk menemukan kelompok rumah tangga dengan kondisi sosial ekonomi yang serupa [8].

Kota Bogor sendiri memiliki enam kecamatan dengan karakteristik sosial dan ekonomi yang beragam. Pengambilan sampel dilakukan dengan metode *two-stage cluster sampling* yang merupakan teknik berjenjang di mana pada tahap pertama dipilih unit-unit *cluster* besar (dalam hal ini kecamatan), dan pada tahap kedua dilakukan pemilihan elemen-elemen dalam klaster tersebut (rumah tangga) dengan menggunakan *convenience sampling*. Teknik ini efisien untuk populasi yang tersebar luas secara geografis dan membantu menurunkan biaya serta waktu survei, tanpa mengorbankan keterwakilan data secara statistik [9]. Pendekatan ini sangat sesuai untuk studi sosial-ekonomi di wilayah perkotaan yang kompleks seperti Kota Bogor.

Berdasarkan hal yang telah disebutkan, dengan mempertimbangkan kompleksitas hubungan antar variabel ekonomi dan demografis dalam menentukan kesejahteraan keluarga, serta adanya variasi karakteristik antar wilayah di Kota Bogor, dibutuhkan pendekatan analitis yang menyeluruh dan berbasis data. Penelitian ini mengintegrasikan analisis regresi logistik untuk melihat hubungan langsung antar variabel, *K-Means Clustering* untuk mengidentifikasi segmentasi rumah tangga berdasarkan pendapatan dan tanggungan, serta metode *two-stage cluster sampling* yang digunakan untuk memastikan keterwakilan data antar kecamatan. Diharapkan, hasil penelitian ini dapat memberikan pemahaman yang lebih komprehensif terhadap dinamika kesejahteraan rumah tangga, sekaligus menjadi dasar dalam penyusunan kebijakan sosial-ekonomi yang lebih tepat sasaran di tingkat lokal.

## 2. METODE PENELITIAN

### 2.1. Data

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif dan inferensial. Analisis deskriptif dilakukan untuk menggambarkan kesejahteraan rumah tangga berdasarkan variabel ekonomi dan demografis. Analisis inferensial dilakukan untuk mengelompokkan rumah tangga ke dalam segmen-segmen kesejahteraan menggunakan *K-Means Clustering* [8], serta menguji pengaruh variabel-variabel independen terhadap status kesejahteraan keluarga menggunakan Regresi Logistik Multinomial [6].

Data dikumpulkan melalui penelitian lapangan yang dilaksanakan di Kota Bogor, Provinsi Jawa Barat, yang terdiri atas enam kecamatan. Pelaksanaan penelitian berlangsung pada bulan Mei 2025, mulai dari penyusunan instrumen, pengumpulan data lapangan, analisis data, hingga pelaporan hasil.

Variabel yang digunakan terdiri atas variabel prediktor dan variabel target. Variabel prediktor meliputi pendapatan bulanan keluarga, pengeluaran bulanan, jumlah tanggungan, pendidikan terakhir kepala keluarga, usia kepala keluarga, jumlah kendaraan, pekerjaan kepala keluarga, kepemilikan kendaraan, dan kepemilikan aset. Variabel target adalah tingkat kesejahteraan keluarga yang diklasifikasikan ke dalam tiga kategori: Kesejahteraan Rendah, Kesejahteraan Sedang, dan Sejahtera [10].

Nama akhir pada penulisan pertama et, al. | 3-5 kata pada awal judul artikel ...

**Tabel 1.** Variabel Penelitian

Variabel	Label	Kategori	Skala
X1	Pendapatan Bulanan Keluarga	-	Rasio
X2	Pengeluaran Bulanan	-	Rasio
X3	Jumlah Tanggungan	-	Rasio
X4	Pendidikan Terakhir Kepala Keluarga	0 = Tidak Sekolah 1 = SD 2 = SMP 3 = SMA 4 = S1/D4 5 = S2	Ordinal
X5	Usia Kepala Keluarga	-	Rasio
X6	Jumlah Kendaraan	-	Rasio
X7	Pekerjaan Kepala Keluarga	1 = Karyawan 2 = Wiraswasta 3 = Wirausaha 4 = PNS 5 = Lainnya	Multinomial
X8	Kepemilikan Kendaraan	1 = Motor Saja 2 = Mobil Saja 3 = Mobil dan Motor 4 = Tidak Punya	Multinomial
X9	Kepemilikan Aset	1 = Milik Sediri 2 = Menumpang 3 = Kontrak/Sewa 4 = Lainnya	Multinomial

## 2.2. Teknik Pengumpulan Data

Populasi dalam penelitian ini adalah seluruh rumah tangga yang berdomisili di Kota Bogor. Sampel dipilih menggunakan metode two-stage cluster sampling [9], dengan tahap pertama berupa pemilihan kecamatan sebagai klaster, dan tahap kedua berupa pemilihan rumah tangga secara convenience sampling di masing-masing kecamatan. Jumlah sampel sebanyak 120 rumah tangga, dengan alokasi 20 rumah tangga per kecamatan.

Data dikumpulkan melalui kuesioner terstruktur yang disusun berdasarkan variabel penelitian, mencakup pendapatan, pengeluaran, jumlah tanggungan, tingkat pendidikan, usia kepala keluarga, pekerjaan kepala keluarga, kepemilikan kendaraan,

dan kepemilikan aset. Kuesioner disebarikan secara langsung kepada responden melalui wawancara tatap muka atau pengisian mandiri.

### 2.3. Teknik Analisis Data

Analisis data dilakukan dalam dua tahap. Tahap pertama menggunakan K-Means Clustering untuk mengidentifikasi segmentasi kesejahteraan keluarga berdasarkan karakteristik ekonomi dan demografis [8]. Jumlah kluster ditentukan secara eksploratif menggunakan metode Elbow Method. Tahap kedua menggunakan Regresi Logistik Multinomial untuk mengidentifikasi pengaruh masing-masing variabel prediktor terhadap tingkat kesejahteraan keluarga [6][7]. Model diuji menggunakan evaluasi performa seperti confusion matrix, precision, recall, F1-score, serta akurasi keseluruhan model.

#### 2.3.1 K-Means Clustering

K-Means merupakan algoritma unsupervised learning yang digunakan untuk mengelompokkan data ke dalam sejumlah kluster berdasarkan kemiripan karakteristik [8]. Pada penelitian ini, K-Means digunakan untuk mengelompokkan rumah tangga berdasarkan variabel ekonomi dan demografis. Jumlah kluster ditentukan secara eksploratif menggunakan Elbow Method.

Proses K-Means diawali dengan inisialisasi centroid secara acak, diikuti dengan proses iteratif pengelompokan data ke dalam kluster terdekat dan perhitungan ulang posisi centroid hingga mencapai konvergensi. Rumus jarak Euclidean yang digunakan untuk menentukan kedekatan data terhadap centroid:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

**Keterangan:**

- $d(x, y)$  = jarak antara titik data  $x$  dan centroid  $y$
- $x_i, y_i$  = nilai atribut ke  $i$
- $n$  = jumlah atribut

Rumus perhitungan centroid pada iterasi ke- $j$ :

$$C_j = \frac{1}{|K_j|} \sum_{xi \in K_j} x_i$$

**Keterangan:**

Nama akhir pada penulisan pertama et, al. | 3-5 kata pada awal judul artikel ...

- $C_j = \text{centroid klaster ke } j$
- $K_j = \text{kumpulan data dalam klaster ke } j$
- $x_i = \text{data ke } i \text{ dalam klaster}$

### 2.3.2 Regresi Logistik

Regresi logistik digunakan untuk memodelkan hubungan antara satu atau lebih variabel prediktor dengan probabilitas kejadian suatu kategori pada variabel dependen yang bersifat kategorik [6][7]. Dalam penelitian ini, regresi logistik multinomial digunakan untuk memprediksi probabilitas keluarga berada pada salah satu kategori tingkat kesejahteraan (rendah, sedang, tinggi), berdasarkan variabel-variabel ekonomi dan demografis.

Persamaan model regresi logistik yang digunakan:

$$P = \frac{1}{1 + e^{-z}}, \text{ dengan } z = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Transformasi logit (log odds):

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Keterangan:

- $P = \text{probabilitas terjadinya kejadian } (Y = 1)$
- $\beta_0 = \text{intercept}$
- $\beta_i = \text{koefisien regresi}$
- $X_i = \text{variabel independen}$
- $e = \text{bilangan eksponensial } (\approx 2.718)$

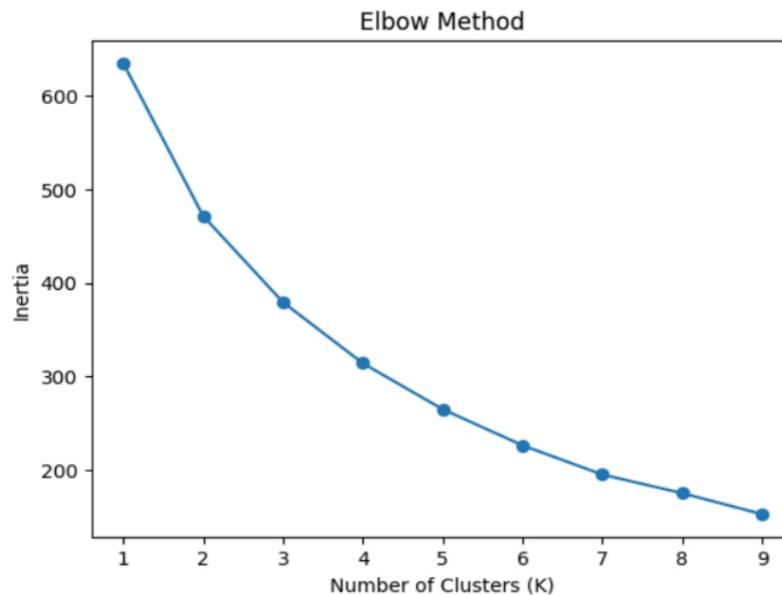
Evaluasi performa model dilakukan menggunakan *confusion matrix*, precision, recall, F1-score, dan akurasi.

## 3. HASIL DAN PEMBAHASAN

### 3.1. K-Means Clustering

Proses segmentasi untuk mengelompokkan rumah tangga berdasarkan tingkat kesejahteraan yang dimiliki bertujuan untuk mengidentifikasi pola-pola tersembunyi dalam data dan melihat bagaimana karakteristik sosial ekonomi tertentu membentuk kelompok-kelompok yang serupa. Dalam hal ini, metode K-Means Clustering dipilih karena kemampuannya dalam mengelompokkan data numerik secara eksploratif tanpa memerlukan variabel target.

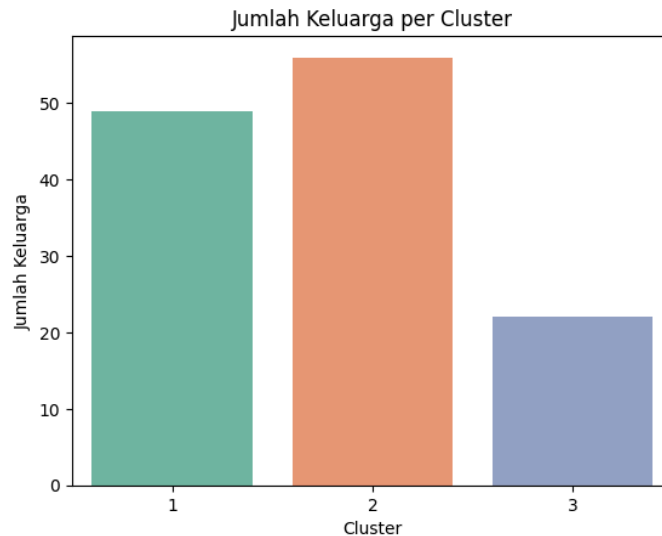
K-Means dinilai tepat untuk menganalisis kesejahteraan keluarga karena mampu membentuk kluster berdasarkan variabel-variabel penting seperti usia, pendapatan bulanan, pengeluaran bulanan, jumlah tanggungan, serta jumlah kendaraan. Dengan meminimalkan variasi dalam setiap kluster dan memaksimalkan perbedaan antar kluster, metode ini memungkinkan pemetaan kelompok rumah tangga yang memiliki kemiripan kondisi kesejahteraan.



**Gambar 1.** Grafik Elbow Method

Untuk menentukan jumlah cluster yang optimal dalam metode K-Means, digunakan pendekatan Elbow Method [8]. Grafik Elbow menunjukkan bahwa penurunan nilai inertia mulai melandai pada nilai  $K = 3$ , sehingga dipilih 3 cluster sebagai jumlah yang optimal. Pemilihan ini mempertimbangkan bahwa setelah titik ini penurunan inertia tidak signifikan, yang menandakan efisiensi segmentasi sudah tercapai.

Berdasarkan hasil segmentasi, Cluster pertama menginformasikan bahwa profil rumah tangga dengan kondisi ekonomi tinggi. Cluster kedua mewakili kelompok menengah dengan tanggungan lebih banyak, sementara cluster ketiga mencerminkan rumah tangga dengan kondisi ekonomi rendah.



**Gambar 2.** Tabel Jumlah Keluarga per Cluster

Grafik menunjukkan bahwa sebagian besar keluarga berada dalam cluster kesejahteraan sedang (Cluster 2), dengan jumlah sekitar 56 keluarga. Kemudian disusul oleh cluster kesejahteraan tinggi (Cluster 1) dengan sekitar 49 keluarga. Sementara itu, keluarga dengan kesejahteraan rendah (Cluster 3) merupakan kelompok paling sedikit, hanya sekitar 23 keluarga. Hal ini mengindikasikan bahwa mayoritas keluarga dalam data ini berada pada tingkat kesejahteraan menengah dan tinggi, sedangkan hanya sebagian kecil keluarga yang tergolong memiliki kesejahteraan rendah. Ini bisa mencerminkan distribusi kondisi sosial ekonomi yang belum merata dalam populasi tersebut.

### 3.2. Regresi Logistik Multinomial

Setelah diperoleh hasil cluster dari segmentasi tingkat kesejahteraan keluarga, langkah selanjutnya dalam penelitian ini adalah membangun model klasifikasi untuk memprediksi tingkat kesejahteraan rumah tangga berdasarkan sejumlah variabel independen. Pemodelan dilakukan menggunakan metode Regresi Logistik Multinomial [6], yang memungkinkan analisis terhadap bagaimana masing-masing variabel seperti pendapatan, pengeluaran, pendidikan, hingga kepemilikan aset berkontribusi terhadap kemungkinan suatu rumah tangga berada pada kategori kesejahteraan tertentu [7].

Metode ini sangat sesuai untuk memodelkan hubungan antara variabel prediktor dan variabel target kategorik, serta dapat menghasilkan probabilitas prediksi yang dapat diinterpretasikan sebagai kecenderungan rumah tangga masuk ke dalam masing-masing kategori kesejahteraan [6]. Pada subbab ini, akan dibahas secara mendalam hasil estimasi parameter model, signifikansi variabel prediktor, serta interpretasi dari pengaruh masing-masing variabel terhadap klasifikasi tingkat kesejahteraan.



**Tabel 2.** Rincian Hasil Regresi Logistik Multinomial pada Kesejahteraan Sedang

Variabel	Koefisien	Standart Error	Z-Value	P-Value
Intercept	0.7185	0.209966	3.422152	0.000621
Pendapatan Bulanan	0.0000	3.65E-08	-0.02419	0.980699
Pengeluaran Bulanan	0.0000	3.83E-08	2.054619	0.039916
Jumlah Tanggungan	-0.8408	0.336583	-2.49793	0.012492
Pendidikan Terakhir Kepala Keluarga	0.7386	0.167905	-0.07844	0.407019
Usia	-0.0013	0.018038	-0.07003	0.944169
Jumlah Kendaraan	0.2222	0.343405	0.64696	0.517658
Pekerjaan Kepala Keluarga	-0.1313	0.121069	-0.84405	0.364409
Kepemilikan Kendaraan	-0.2675	0.123549	-0.51076	0.302846
Kepemilikan Aset	1.1709	0.29889	-0.61352	0.119644

Dari hasil regresi logistik multinomial yang ditampilkan pada Tabel 2, terlihat bahwa dari **uji parsial** yang dihasilkan, terdapat beberapa variabel independen yang berpengaruh signifikan secara statistik terhadap peluang suatu keluarga berada dalam kategori kesejahteraan sedang dibandingkan dengan kategori sejahtera.

Variabel pengeluaran bulanan berpengaruh signifikan terhadap peluang seseorang berada pada tingkat kesejahteraan sedang, dengan nilai *p-value* sebesar 0,039. Hal ini menunjukkan bahwa semakin tinggi pengeluaran bulanan, maka semakin besar pula peluang individu untuk berada pada tingkat kesejahteraan sedang. Variabel jumlah tanggungan berpengaruh negatif dan signifikan terhadap kesejahteraan sedang dengan nilai *p-value* = 0,012. Artinya, semakin banyak jumlah tanggungan dalam rumah tangga, maka semakin kecil kemungkinan keluarga tersebut termasuk dalam kategori kesejahteraan sedang.

Sementara itu, variabel-variabel lain seperti pendapatan bulanan, pendidikan terakhir kepala keluarga, usia, jumlah kendaraan, pekerjaan kepala keluarga, kepemilikan kendaraan, dan kepemilikan aset tidak menunjukkan pengaruh yang signifikan terhadap kategori kesejahteraan sedang (semua dengan *p-value* > 0,05). Hal ini menunjukkan bahwa secara statistik, pengaruh dari variabel-variabel tersebut belum cukup kuat untuk dijadikan indikator utama dalam menentukan status kesejahteraan sedang pada model yang dibangun.

Z-Value positif mengindikasikan peningkatan variabel prediktor akan meningkatkan log odds untuk kategori “sedang”. Sedangkan Z-Value negatif mengindikasikan peningkatan variabel prediktor akan menurunkan peluang untuk kategori “sedang”. Hanya variabel dengan *p-value* < 0.05 yang berpengaruh signifikan secara statistik.

**Tabel 3. Rincian Hasil Regresi Logistik Multinomial pada Kategori Kesejahteraan Rendah**

Variabel	Koefisie n	Standart Error	Z-Score	P-Value
Intercept	-0.3537	0.227398	-1.555614	0.119800
Pendapatan Bulanan	-0.0000	4.69E-08	0.180148	0.857037
Pengeluaran Bulanan	0.0000	1.72E-07	-0.96244	0.335826
Jumlah Tanggungan	-1.2660	0.432529	-2.92697	0.003423
Pendidikan Terakhir	-0.4931	0.200552	-0.40436	0.444904
Usia	0.0788	0.02135	3.690673	0.000224
Jumlah Kendaraan	-0.1236	0.466443	-0.26501	0.791002
Pekerjaan Kepala Keluarga	-0.0020	0.17014	0.113153	0.57404
Kepemilikan Kendaraan	0.1968.	0.156934	0.539722	0.613102
Kepemilikan Aset	-0.5951	0.362424	-0.14609	0.526219

Dari hasil regresi logistik multinomial yang ditampilkan pada Tabel 3, terlihat bahwa dari **uji parsial** yang dihasilkan, terdapat beberapa variabel independen yang berpengaruh signifikan secara statistik terhadap peluang suatu keluarga berada dalam kategori kesejahteraan rendah dibandingkan dengan kategori sejahtera.

Variabel jumlah tanggungan memiliki pengaruh negatif dan signifikan terhadap peluang individu masuk dalam kategori sejahtera, dengan nilai *p-value* sebesar 0,003. Artinya, semakin banyak jumlah tanggungan dalam keluarga, maka semakin kecil kemungkinan keluarga tersebut termasuk dalam kategori sejahtera.

Variabel usia menunjukkan pengaruh positif yang signifikan (*p-value* = 0,0002). Hal ini mengindikasikan bahwa semakin bertambah usia kepala keluarga, maka semakin besar peluang untuk berada pada kategori sejahtera meningkat menjadi kesejahteraan rendah.

Sementara itu, variabel-variabel lain seperti pendapatan bulanan, pengeluaran bulanan, pendidikan terakhir kepala keluarga, jumlah kendaraan, pekerjaan kepala keluarga, kepemilikan kendaraan, dan kepemilikan aset tidak memberikan pengaruh yang signifikan terhadap kesejahteraan sejahtera (*p-value* > 0,05). Hal ini menunjukkan bahwa secara statistik, variabel-variabel tersebut belum menjadi penentu utama dalam mengklasifikasikan kesejahteraan pada kategori kesejahteraan rendah.

Z-Score membantu menunjukkan variabel mana yang secara statistik benar-benar berkontribusi dalam membedakan keluarga yang “kurang sejahtera” dari yang “sejahtera”. Dalam model ini, hanya jumlah tanggungan (negatif) dan usia (positif) yang menunjukkan pengaruh yang bermakna secara statistik terhadap probabilitas masuk dalam kategori kurang sejahtera.

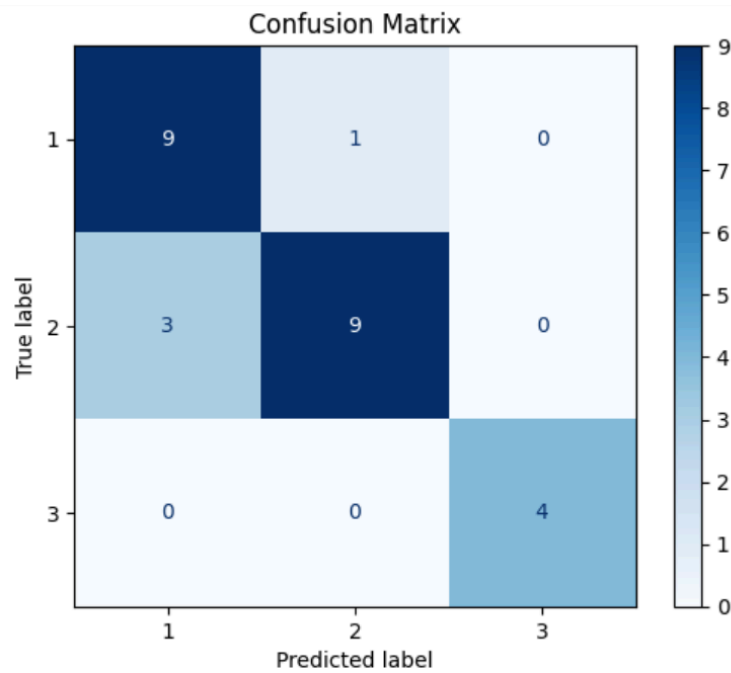
Berdasarkan nilai konstanta yang dihasilkan dari tabel, persamaan logit yang terbentuk adalah sebagai berikut. Proses training menghasilkan dua buah fungsi logit, yaitu fungsi logit untuk kategori kesejahteraan sedang dan kategori kesejahteraan

rendah, sementara kategori yang digunakan sebagai pembanding adalah kategori sejahtera.

$$g_2(x) = \ln\left(\frac{P(Y=2|x)}{P(Y=1|x)}\right) = 0.7185 - 0.8408x_3 + 0.7386x_4 - 0.0013x_5 + 0.222x_6 \\ - 0.1313x_7 - 0.2675x_8 + 1.1709x_9$$

$$g_3(x) = \ln\left(\frac{P(Y=3|x)}{P(Y=1|x)}\right) = -0.3537 - 0.8408x_3 + 0.7386x_4 - 0.0013x_5 + 0.222x_6 \\ - 0.1313x_7 - 0.2675x_8 + 1.1709x_9$$

Selanjutnya, *confusion matrix* ditampilkan untuk menyajikan kinerja model secara lebih mendalam.



**Gambar 3.** *Confussion Matrix*

Dari pemetaan *confusion matrix* yang ditampilkan pada Gambar 1, terlihat bahwa kelas sejahtera (kelas 1) mampu terprediksi dengan cukup baik, dengan prediksi yang tepat sebanyak 9 data. Namun demikian, masih terdapat 1 data dari kelas sejahtera yang salah diprediksi sebagai kelas kesejahteraan sedang (kelas 2). Sementara itu, kelas kesejahteraan sedang tidak mampu terprediksi sebaik kelas kelas sejahtera, masih terdapat 3 data yang salah diklasifikasikan ke dalam kelas sejahtera. Hal ini menunjukkan bahwa model memiliki kecenderungan untuk mengklasifikasikan sebagian data kelas 2 ke kelas 1. Di sisi lain, kelas kesejahteraan rendah (kelas 3) merupakan bagian terbaik dari performa model ini, di mana seluruh data kelas kesejahteraan rendah (sebanyak 4 data) berhasil diprediksi dengan tepat tanpa kesalahan. Temuan ini mengindikasikan bahwa model cukup andal dalam mengenali pola pada kelas 3, namun masih memerlukan perbaikan dalam membedakan antara kelas 1 dan kelas 2.

**Tabel 4. Classification Report**

	Precision	Recall	F1-score	support
1	0.75	0.90	0.3422	10
2	0.90	0.75	0.0241	12
3	1	1.00	1.00	4
Akurasi = 0.84				

Kesimpulan yang dapat ditarik dari kinerja model ini adalah bahwa akurasi model sebesar 84% menunjukkan bahwa model cukup akurat secara keseluruhan dalam memprediksi kelas kesejahteraan.

Pada kelas sejahtera (**kelas 1**), model menunjukkan kemampuan yang cukup baik dalam mengenali sebagian besar sampel aktual (recall = 0,90), meskipun terdapat beberapa prediksi yang salah (precision = 0,75). Nilai *F1-score* sebesar 0,34 menunjukkan bahwa keseimbangan antara precision dan recall masih belum optimal, sehingga akurasi prediksi untuk kelas ini masih perlu ditingkatkan.

Untuk kelas kesejahteraan sedang (**kelas 2**), model memiliki presisi yang sangat baik (precision = 0,90), menandakan bahwa sebagian besar prediksi yang dibuat untuk kelas ini memang benar. Namun, recall-nya hanya sebesar 0,75, artinya masih ada sebagian data aktual yang tidak berhasil diklasifikasikan sebagai kelas 2. Nilai *F1-score* yang rendah (0,02) menunjukkan ketidakseimbangan yang cukup signifikan antara precision dan recall, sehingga performa model pada kelas ini masih tergolong lemah.

Sementara itu, kelas kesejahteraan rendah (**kelas 3**) menunjukkan hasil yang sangat baik, di mana model mampu memprediksi semua data kelas ini secara tepat (precision = 1,00; recall = 1,00; *F1-score* = 1,00), tanpa adanya kesalahan klasifikasi.

Secara keseluruhan, meskipun akurasi model tinggi, kinerja model masih kurang stabil pada kelas 1 dan 2. Hal ini tercermin dari nilai *F1-score* yang rendah dan tidak seimbang, sehingga diperlukan evaluasi lanjutan atau perbaikan model agar performa terhadap seluruh kelas dapat lebih merata dan representatif.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian, diperoleh kesimpulan sebagai berikut:

Segmentasi kesejahteraan keluarga di Kota Bogor terbagi menjadi tiga kelompok, yaitu kategori sejahtera, kesejahteraan sedang, dan kesejahteraan rendah, dengan proporsi terbesar berada pada kelompok kesejahteraan sedang dan sejahtera.

Variabel pengeluaran bulanan dan jumlah tanggungan terbukti berpengaruh signifikan terhadap peluang keluarga untuk berada pada kategori kesejahteraan sedang. Pada kategori kesejahteraan rendah, jumlah tanggungan dan usia kepala keluarga memiliki pengaruh yang signifikan terhadap tingkat kesejahteraan.

## DAFTAR PUSTAKA

- [1] Purwanto, E. (2018). *Konsep Kesejahteraan dan Faktor Penentunya dalam Kajian Sosial Ekonomi*. Jurnal Ekonomi dan Pembangunan, 26(1), 25–34.
- [2] Utaminingsih, A., & Suwendra, I. W. (2022). *Relevansi Pendapatan Terhadap Kesejahteraan Rumah Tangga di Indonesia*. Jurnal Ekonomi Pembangunan, 20(1), 78–85.
- [3] Nata, R., Siregar, H., & Susilowati, S. H. (2020). *Analisis Kesejahteraan Rumah Tangga Petani Pisang di Sumatera Utara*. Jurnal Sosial Ekonomi Pertanian dan Agribisnis, 8(1), 45–54.
- [4] Sari, R. P. (2023). *Pengaruh Tanggungan Rumah Tangga terhadap Kesejahteraan di Daerah Perdesaan*. Jurnal Ekonomi Syariah dan Sosial, 5(1), 19–28.
- [5] Agustriyani, A. (2022). *Analisis Faktor-faktor yang Mempengaruhi Tingkat Kesejahteraan Rumah Tangga*. Jurnal Ekonomi dan Sosial, 13(2), 55–67.
- [6] Ariefianto, D. (2012). *Ekonometrika: Esensi dan Aplikasi dengan Menggunakan Eviews*. Jakarta: Erlangga.
- [7] Permatasari, N. P., & Yuliana, D. (2021). *Analisis Faktor Penentu Kesejahteraan Rumah Tangga Menggunakan Model Regresi Logistik*. Jurnal Statistika dan Komputasi, 6(2), 33–42.
- [8] Wulandari, D. (2022). *Penerapan K-Means Clustering untuk Segmentasi Sosial Ekonomi Rumah Tangga*. Jurnal Data Mining dan Sistem Informasi, 3(1), 15–23.
- [9] Ugwu, C. N., & Madukaife, N. I. (2022). *Sampling Techniques in Social Research: A Review of Cluster Sampling Methods*. International Journal of Social Science Research, 10(4), 104–118.
- [10] Badan Pusat Statistik. (2023). *Indikator Kesejahteraan Rakyat Indonesia 2023*. Jakarta: BPS RI.