



PERBANDINGAN PENANGANAN *IMBALANCED DATA* DALAM MODEL *MULTINOMIAL LOGISTIC REGRESSION* GUNA IDENTIFIKASI FAKTOR STATUS GIZI BALITA DI INDONESIA

MUHAMMAD ABSHOR DZULHIJ RIZKI



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI, SUMBER INFORMASI, PENGGUNAAN AI, DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Penanganan *Imbalanced Data* dalam Model *Multinomial Logistic Regression* guna Identifikasi Faktor Status Gizi Balita di Indonesia” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dalam penyusunan karya ini, penulis menggunakan bantuan kecerdasan *Chat GPT* buatan untuk mereview dan menerjemahkan jurnal berbahasa inggris dan menggali informasi yang ada di internet guna efisiensi waktu dan menemukan celah penelitian.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juni 2026

Muhammad Abshor Dzulhij Rizki
G1401221038



ABSTRAK

MUHAMMAD ABSHOR DZULHIJ RIZKI. Perbandingan Penanganan *Imbalanced Data* dalam Model *Multinomial Logistic Regression* guna Identifikasi Faktor Status Gizi Balita di Indonesia. Dibimbing oleh MUHAMMAD NUR AIDI dan SACHNAZ DESTA OKTARINA.

Imbalanced data merupakan permasalahan yang sering terjadi pada klasifikasi status gizi balita akibat distribusi kategori yang tidak proporsional. Kondisi ini menyebabkan model klasifikasi cenderung bias terhadap kategori mayoritas sehingga kemampuan prediksi pada kategori minoritas menjadi rendah. Penelitian ini bertujuan membandingkan metode penanganan *imbalanced data* pada *Multinomial Logistic Regression* (MLR) untuk klasifikasi status gizi balita. Metode MLR dipilih karena variabel respon status gizi balita terdiri atas lebih dari dua kategori, yaitu normal, *double burden of malnutrition* (DBM), *overweight*, dan *stunting*. Metode penanganan di antaranya tanpa penanganan, *oversampling*, *undersampling*, *median weighted*, dan *inverse square root weighted*. Data yang digunakan merupakan data sekunder Survei Status Gizi Indonesia tahun 2024 dengan sampel 87.428 bayi usia 6-23 bulan. Variabel penjelas adalah praktik pemberian makan bayi dan anak, karakteristik anak, ibu, serta keluarga. Sebelum pemodelan MLR dilakukan analisis *chi-square* sebagai tahap awal seleksi variabel penjelas. Indikator evaluasi model di antaranya uji simultan, jumlah variabel signifikan parsial, *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC), *accuracy*, *balanced accuracy*, dan kemampuan prediksi tiap kategori status gizi balita. Hasil penelitian menunjukkan tidak terdapat metode yang dominan pada seluruh indikator evaluasi sehingga dilakukan pemeringkatan model. Metode *median weighted* memperoleh skor tertinggi dan dipilih sebagai model terbaik. Seleksi *stepwise* menghasilkan 13 variabel signifikan untuk pemodelan logit dan interpretasi faktor yang memengaruhi status gizi balita. Model terbaik menghasilkan nilai AIC 12.980, BIC 12.942, *accuracy* 72,0%, dan *balanced accuracy* 68,5%.

Kata kunci: penanganan ketidakseimbangan data, regresi logistik multinomial, seleksi stepwise, status gizi balita.



ABSTRACT

MUHAMMAD ABSHOR DZULHIJ RIZKI. Comparative Handling Imbalanced Data in the Multinomial Logistic Regression Model to Identify Nutritional Status Factors of Toddlers in Indonesia. Guided by MUHAMMAD NUR AIDI and SACHNAZ DESTA OKTARINA.

Imbalanced data is a common problem in the classification of child nutritional status due to disproportionate category distributions. This condition causes classification models to be biased toward majority categories, resulting in poor predictive performance for minority categories. This study aimed to compare several imbalanced data handling methods in Multinomial Logistic Regression (MLR) for classifying child nutritional status. MLR was selected because the response variable consisted of more than two categories, namely normal, double burden of malnutrition (DBM), overweight, and stunting. The methods compared included no treatment, oversampling, undersampling, median weighted, and inverse square root weighted. The study used secondary data from the 2024 Indonesian Nutritional Status Survey, consisting of 87,428 infants aged 6–23 months. The explanatory variables included infant feeding, child, maternal, and family characteristics. Before MLR modeling, a chi-square analysis was conducted as an initial variable screening procedure. The evaluation criteria included the simultaneous test, the number of partially significant variables, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), accuracy, balanced accuracy, and predictive performance for each nutritional status category. The results showed that no single method consistently outperformed the others across all evaluation criteria; therefore, a ranking procedure was applied. The median weighted method achieved the highest overall score and was selected as the best-performing model. Stepwise selection identified 13 significant variables for logit modeling and interpretation of factors associated with child nutritional status. The best model produced an AIC of 12,980, a BIC of 12,942, an accuracy of 72.0%, and a balanced accuracy of 68.5%.

Keywords: handling imbalanced data, infant nutritional status, multinomial logistic regression, stepwise.



@Hak cipta milik IPB University

IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



PERBANDINGAN PENANGANAN *IMBALANCED DATA* DALAM MODEL *MULTINOMIAL LOGISTIC REGRESSION* GUNA IDENTIFIKASI FAKTOR STATUS GIZI BALITA DI INDONESIA

MUHAMMAD ABSHOR DZULHIJ RIZKI

Skripsi

sebagai salah satu syarat untuk memperoleh gelar Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
Rahma Anisa, S.Stat., M.Si., M.Act.Sc.



Judul Skripsi : Perbandingan Penanganan *Imbalanced Data* dalam Model
Multinomial Logistic Regression guna Identifikasi Faktor Status
Gizi Balita di Indonesia

Nama : Muhammad Abshor Dzulhij Rizki
NIM : G1401221038

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Muhammad Nur Aidi, M.S.



Pembimbing 2:
Sachnaz Desta Oktarina, S.Stat., M.Agr., Ph.D.



Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, S.Si., M.Si
NIP 197804112005011002

Tanggal Ujian:
(10-Juni-2026)

Tanggal Lulus:
(tanggal penandatanganan oleh Dekan
Fakultas/Sekolah ...)

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Penelitian yang dilaksanakan sejak bulan Desember 2025 sampai bulan Mei 2026 ini judul “Perbandingan Penanganan *Imbalanced Data* dalam Model *Multinomial Logistic Regression* guna Identifikasi Faktor Status Gizi Balita di Indonesia”.

Penulisan karya ilmiah ini tentunya tidak lepas dari dukungan dan bantuan berbagai pihak. Terima kasih penulis ucapkan kepada:

1. Bapak Prof. Dr. Ir. Muhammad Nur Aidi, M.S. dan Ibu Sachnaz Desta Oktarina, S.Stat., M.Agr., Ph.D. selaku dosen pembimbing yang telah membimbing dan banyak memberi saran demi kebaikan penulisan yang berdampak.
2. Ibu Prof. Dr. Fitrah Ernawati, M.Sc. dari Badan Riset Nasional (BRIN) yang telah memberikan banyak masukan dalam hal gizi dan akses berupa data untuk penelitian.
3. Rahma Anisa, S.Stat., M.Si., M.Act.Sc. selaku dosen penguji yang telah memberikan saran, kritik, dan masukan yang membangun bagi penyempurnaan penelitian.
4. Seluruh dosen dan staf administrasi Prodi Statistika dan Sains Data IPB University yang memberikan ilmu dan sarana penunjang lainnya.
5. Ayah saya Muhammad Syafruddin dan Ibu saya Ratna Harini, selaku orang tua yang senantiasa memberikan doa, dukungan dan kasih sayang yang tidak pernah putus kepada penulis. Mas Winno dan mas Diggi selaku kakak yang memberikan dorongan dan motivasi penuh dalam penulisan.
6. Barbang yang di antaranya Dana, Farid, Icad, Hani, Deva dan Adel selaku teman dari PKU yang masih mengikat hingga akhir penulisan. Beta yang di antaranya Bulan, Delia, Cindy, Vito selaku teman di statistika yang berjuang bersama-sama hingga skripsi ini di buat. Nuru yang selalu ada untuk di hubungi setiap saat.
7. Himasurya Plus IPB yang didalamnya ada Aryak, Avril, Syz, Lino, dan kawan kawan lainnya selaku organisasi dan keluarga di Bogor, memberikan wadah dan bertumbuh pesat di dalamnya.
8. Teman-teman *marinestic* yang memberikan pencerahan dan memberikan dukungan serta semangat. Semua pihak yang telah membantu penulis dalam penyusunan karya ilmiah ini yang tidak dapat disebutkan satu persatu.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juni 2026

Muhammad Abshor Dzulhij Rizki



DAFTAR ISI

DAFTAR GAMBAR	ix
DAFTAR TABEL	ix
DAFTAR LAMPIRAN	ix
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
TINJAUAN PUSTAKA	3
2.1 Status Gizi pada Balita	3
2.2 Distribusi Multinomial	4
2.3 Uji <i>Chi-Square</i>	4
2.4 <i>Generalized Variance Inflation Factor (GVIF)</i>	5
2.5 Regresi Logistik Multinomial	6
2.6 Pendugaan Parameter	7
2.7 Penanganan <i>Imbalanced Data</i>	10
2.8 Ketepatan Klasifikasi	12
2.9 Keباikan Model	13
2.10 Interpretasi Model	13
III METODE	15
3.1 Data Penelitian	15
3.2 Prosedur Analisis Data	15
IV HASIL DAN PEMBAHASAN	18
4.1 Persiapan Data	18
4.2 Analisis <i>Bivariate</i>	19
4.3 Model MLR tanpa penanganan	20
4.4 Model MLR dengan <i>Oversampling</i>	22
4.5 Model MLR dengan <i>Undersampling</i>	24
4.6 Model MLR dengan Pembobotan <i>Median Weighted</i>	26
4.7 Model MLR dengan Pembobotan <i>Inverse Square Root</i>	28
4.8 Perbandingan Kinerja MLR	30
4.9 Pemodelan dengan Model Terbaik	33
4.10 Identifikasi Faktor dengan Model Terpilih	35
4.11 Interpretasi dengan Model Terpilih	37
V SIMPULAN DAN SARAN	42
5.1 Simpulan	42
5.2 Saran	42
DAFTAR PUSTAKA	44
LAMPIRAN	48
RIWAYAT HIDUP	58

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR GAMBAR

1	Diagram alir penelitian	15
2	Distribusi status gizi balita	18
3	Distribusi hasil dari <i>filtering</i> data	21
4	Distribusi hasil dari <i>oversampling</i> data	23
5	Distribusi hasil dari <i>undersampling</i> data	25

DAFTAR TABEL

1	Kategori dan ambang batas status gizi balita	3
2	Ketepatan klasifikasi	12
3	Hasil chi-square variabel penjelas terhadap variabel respon	19
4	Pemeriksaan multikolinearitas antar variabel penjelas	20
5	Hasil uji <i>wald</i> variabel yang signifikan model <i>baseline</i>	21
6	Ketepatan klasifikasi	22
7	Uji <i>wald</i> variabel signifikan model MLR <i>oversampling</i>	23
8	Ketepatan klasifikasi model <i>oversampling</i>	24
9	Hasil uji <i>wald</i> variabel yang signifikan model MLR <i>Undersampling</i>	25
10	Ketepatan klasifikasi model <i>undersampling</i>	26
11	Nilai <i>median weighted</i> setiap kategori variabel respon	27
12	Hasil uji <i>wald</i> model <i>median weighted</i>	27
13	Ketepatan klasifikasi model <i>median weighted</i>	28
14	Nilai <i>inverse sqrt weighted</i> setiap kategori variabel respon	29
15	Hasil uji <i>wald</i> model <i>inverse sqrt</i>	29
16	Ketepatan klasifikasi model <i>invers sqrt</i>	30
17	Perbandingan kinerja MLR	31
18	Pemilihan model terbaik dengan pemeringkatan	32
19	Hasil pengujian parameter parsial dengan uji <i>wald</i>	33
20	Perbandingan variabel penjelas yang memengaruhi	36
21	Dugaan nilai <i>relative risk ratio</i>	37

DAFTAR LAMPIRAN

1	Distribusi setiap variabel penjelas terhadap variabel respon	49
2	Variabel Penjelas Beserta Definisi Operator	53