

# Predicting Household Food Insecurity Status in Langsa City Using Double Random Forest and EasyEnsemble Algorithms

Rizqi Ananda<sup>1</sup>, Munirul Ula<sup>2</sup>, Fadlisyah<sup>3\*</sup>

<sup>1,2,3</sup> Teknik Informatika, Fakultas Teknik, Universitas Malikussaleh, Kota Lhokseumawe, 24355, Indonesia

## Informasi Artikel

Diterima : 04 Mei 2026  
Revisi : 10 Mei 2026  
Publikasi : 30 Juni 2026

## Kata Kunci:

Machine Learning  
Kerawanan Pangan  
Double Random Forest  
EasyEnsemble  
SUSENAS

## ABSTRAK

Mendeteksi rumah tangga yang rawan pangan secara akurat merupakan tantangan tersendiri karena kondisi ini bersifat multidimensi. Penelitian ini membandingkan tiga pendekatan machine learning — Random Forest (RF), Double Random Forest (DRF), dan RF yang dikombinasikan dengan teknik penyeimbang kelas EasyEnsemble — dalam memprediksi status kerawanan pangan rumah tangga di Kota Langsa. Data bersumber dari Survei Sosial Ekonomi Nasional (SUSENAS) 2024, mencakup 2.057 rumah tangga dengan 13 variabel prediktor. Variabel target ditetapkan menggunakan ambang berat  $\text{count} \geq 3$  pada indikator FIES, menghasilkan 755 rumah tangga rawan (36,7%) dan 1.302 tidak rawan (63,3%). Eksperimen diulang sebanyak 50 kali dengan skema pembagian data 70:30 untuk memperoleh estimasi performa yang stabil. RF + EasyEnsemble mencapai kinerja terbaik dengan AUC rata-rata 0,8398 dan sensitivitas 79,84%, jauh melampaui DRF yang hanya menghasilkan sensitivitas 2,62%. Uji ANOVA ( $F = 191,899$ ;  $p < 0,001$ ) dan Tukey HSD mengonfirmasi perbedaan antar model signifikan secara statistik. Analisis feature importance mengungkap kepesertaan program bantuan sosial (54,78%) dan kondisi fisik hunian (28,89%) sebagai prediktor paling dominan.

## ABSTRACT

Accurately identifying food-insecure households is challenging because the condition is inherently multidimensional. This study compares three machine-learning approaches — Random Forest (RF), Double Random Forest (DRF), and RF combined with EasyEnsemble class balancing — for predicting household food insecurity in Langsa City. Data from the 2024 SUSENAS survey cover 2,057 households with 13 predictor variables. A  $\text{berat\_count} \geq 3$  threshold on the FIES indicators defines the target variable, yielding 755 food-insecure (36.7%) and 1,302 food-secure (63.3%) households. Fifty repetitions with a 70:30 train-test split yield stable performance estimates. RF + EasyEnsemble achieves the best results with a mean AUC of 0.8398 and sensitivity of 79.84%, far surpassing DRF at 2.62%. ANOVA ( $F = 191.899$ ;  $p < 0.001$ ) and Tukey HSD tests confirm statistically significant differences. Feature importance reveals social-assistance participation (54.78%) and physical housing conditions (28.89%) as the dominant predictors.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



## \*Penulis Koresponden

Email: [iqinanda17@gmail.com](mailto:iqinanda17@gmail.com)

Cara sitasi IEEE::

R. Ananda, M. Ula, F. Fadlisyah, "Predicting Household Food Insecurity Status in Langsa City Using Double Random Forest and EasyEnsemble Algorithms," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 6, no. 2, p. 95-104, Juni 2026. doi: 10.30811/jaise.v6i2.9116

## 1. PENDAHULUAN

Ketahanan pangan merupakan salah satu pilar pembangunan nasional yang terus mendapat perhatian serius dari pemerintah. Undang-Undang Nomor 18 Tahun 2012 tentang Pangan mendefinisikannya sebagai terpenuhinya pangan bagi negara sampai perseorangan, yang tercermin dari ketersediaan pangan yang cukup, aman, beragam, bergizi, merata, dan terjangkau. Data Badan Pusat Statistik (BPS) menunjukkan bahwa pada tahun 2023 prevalensi penduduk dengan kerawanan pangan sedang atau berat di Indonesia mencapai 4,5%, sementara Provinsi Aceh mencatat angka lebih tinggi, yakni 4,70% [1]. Kota Langsa sebagai salah satu daerah otonom di Aceh menghadapi tantangan serupa, sehingga diperlukan cara yang lebih sistematis dan berbasis bukti untuk mengidentifikasi kelompok rumah tangga yang paling rentan.

Persoalan kerawanan pangan tidak sesederhana kurangnya ketersediaan bahan makanan. Kondisi ini umumnya muncul dari perpaduan berbagai faktor sekaligus, mulai dari kemiskinan struktural, rendahnya tingkat pendidikan, keterbatasan akses layanan kesehatan, hingga lemahnya infrastruktur dasar permukiman [2]. Karena sifatnya yang multidimensi inilah, pendekatan statistik konvensional kerap kurang mampu menangkap kerumitan hubungan antar variabel secara memadai. Di sinilah algoritma machine learning berbasis pohon keputusan ensemble seperti Random Forest menjadi relevan, karena terbukti andal dalam menganalisis data tabular sosial ekonomi dengan banyak variabel kategorik dan hubungan non-linear [3].

Salah satu hambatan teknis yang umum dijumpai dalam prediksi kerawanan pangan adalah ketidakseimbangan kelas: proporsi rumah tangga rawan jauh lebih kecil daripada yang tidak rawan, sehingga model cenderung bias ke kelas mayoritas dan gagal mendeteksi kasus minoritas yang justru paling memerlukan intervensi [4]. Untuk menjawab tantangan ini, Double Random Forest (DRF) — yang dikembangkan Han et al. (2020) — memperkenalkan lapisan randomisasi tambahan pada setiap simpul pohon guna memperbesar keragaman antar pohon [5]. Di sisi lain, EasyEnsemble bekerja dengan membangun beberapa subset pelatihan yang seimbang secara proporsional tanpa harus membuang informasi dari kelas mayoritas [6].

Rais et al. (2024) telah menggabungkan DRF dan EasyEnsemble pada data SUSENAS di Sulawesi Selatan dan melaporkan AUC 0,692 [7]. Sebelumnya, Ramadhani et al. (2022) melaporkan Random Forest menghasilkan AUC 0,718 untuk prediksi kerawanan pangan di Provinsi Aceh [2]. Meski demikian, belum ada penelitian yang secara spesifik menggunakan data SUSENAS 2024 Kota Langsa.

Bertolak dari kesenjangan tersebut, penelitian ini menetapkan tiga tujuan: (1) menerapkan dan membandingkan RF, DRF, serta RF + EasyEnsemble; (2) menguji signifikansi statistik perbedaan antar model melalui ANOVA dan Tukey HSD; dan (3) mengidentifikasi variabel-variabel paling berkontribusi menggunakan analisis feature importance berbasis Mean Decrease Impurity (Gini).

## 2. METODE

### 2.1 Random Forest (RF)

Random Forest diperkenalkan Breiman (2001) sebagai metode ensemble yang membangun banyak pohon keputusan dari subset bootstrap berbeda, lalu menggabungkan hasilnya via voting mayoritas [9]. Pada setiap simpul, hanya sejumlah  $m \approx \sqrt{p}$  fitur yang dipilih secara acak guna mereduksi korelasi antar pohon. Prediksi akhir RF dirumuskan sebagai:

$$\hat{C}_B(x) = \underset{j}{\operatorname{argmax}} \sum_{b=1}^B I[\hat{C}_b(x) = j] \quad (1)$$

di mana  $\hat{C}_B(x)$  adalah keputusan akhir ensemble,  $B$  jumlah pohon,  $\hat{C}_b(x)$  prediksi pohon ke- $b$ , dan  $I(\cdot)$  fungsi indikator.

### 2.2 Double Random Forest (DRF)

DRF merupakan perluasan RF yang memperkenalkan bootstrap di dua tingkatan — pada level dataset dan pada setiap simpul non-terminal selama pohon tumbuh [5]. Apabila jumlah data di suatu simpul melampaui ambang 10% dari total data pelatihan, DRF melakukan resampling sebelum memilih fitur pemisah:

$$\text{DRF-node } t: \text{ jika } |D_t| > 0,1N \rightarrow \text{resample } D_t^* \quad (2)$$

Mekanisme ini berpotensi meningkatkan keragaman pohon secara signifikan pada data dengan fitur kontinu yang banyak dan tingkat kebisingan tinggi [5]. Namun pada dataset yang hampir seluruhnya terdiri dari variabel kategorik — seperti SUSENAS Kota Langsa — resampling tambahan tidak menghasilkan perbedaan pemisahan yang bermakna karena fitur-fitur paling informatif tetap mendominasi pilihan [7].

### 2.3 EasyEnsemble

EasyEnsemble dirancang untuk menangani ketidakseimbangan kelas melalui boosting-based undersampling [6]. Kelas mayoritas dibagi menjadi T subset acak, masing-masing digabungkan dengan seluruh data kelas minoritas sehingga terbentuk T dataset seimbang. Pada tiap dataset dilatih satu model AdaBoost; prediksi akhir diperoleh dari gabungan seluruh model:

$$H(x) = \text{sign} \left[ \sum_{t=1}^T \alpha_t \cdot h_t(x) \right] \quad (3)$$

di mana  $h_t(x)$  adalah model AdaBoost ke-t dan  $\alpha_t$  bobotnya. Berbeda dari random undersampling biasa, EasyEnsemble memastikan setiap sampel mayoritas dimanfaatkan dalam setidaknya satu subset sehingga tidak ada informasi yang benar-benar terbuang [6].

### 2.4 Feature Importance Berbasis Mean Decrease Impurity (Gini)

Interpretabilitas model pada penelitian ini menggunakan metode feature importance bawaan Random Forest, yang mengukur kontribusi tiap variabel prediktor berdasarkan rata-rata penurunan impuritas Gini di seluruh pohon dalam ensemble [9]. Setiap kali sebuah variabel dipilih sebagai pemisah pada suatu simpul, penurunan impuritas yang dihasilkan dicatat. Rata-rata penurunan di seluruh pohon itulah yang menjadi skor kepentingan variabel tersebut:

$$FI_j = (1/B) \cdot \sum_{b=1}^B \sum_{t \in T_b: \text{split on } j} \Delta \text{Gini}(t) \quad (4)$$

di mana B adalah jumlah pohon,  $T_b$  himpunan simpul pohon ke-b, dan  $\Delta \text{Gini}(t)$  penurunan impuritas Gini pada simpul t ketika variabel j digunakan sebagai pemisah. Nilai  $FI_j$  kemudian dinormalisasi sehingga jumlah seluruh skor sama dengan 1. Variabel dengan  $FI_j$  tinggi berarti lebih sering dipilih sebagai pemisah dan menghasilkan partisi kelas yang lebih bersih — sehingga lebih informatif dalam membedakan rumah tangga rawan dari yang tidak rawan [3].

### 2.5 Metrik Evaluasi

Evaluasi dilakukan menggunakan confusion matrix sebagai dasar (Tabel 1), dari mana empat metrik diturunkan.

Tabel 1. Confusion Matrix

Prediksi	Aktual: Rawan	Aktual: Tdk Rawan
Rawan	TP	FP
Tidak Rawan	FN	TN

Dari tabel tersebut diturunkan empat metrik:

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN) \quad (5)$$

$$\text{Sensitivitas} = TP / (TP + FN) \quad (6)$$

$$\text{Spesifisitas} = TN / (TN + FP) \quad (7)$$

$$\text{AUC} = P(\text{score}_{\text{pos}} > \text{score}_{\text{neg}}) \quad (8)$$

Pada data tidak seimbang, sensitivitas dan AUC lebih informatif daripada akurasi semata, karena model yang hanya memprediksi kelas mayoritas dapat menghasilkan akurasi tinggi namun sensitivitas mendekati nol [11].

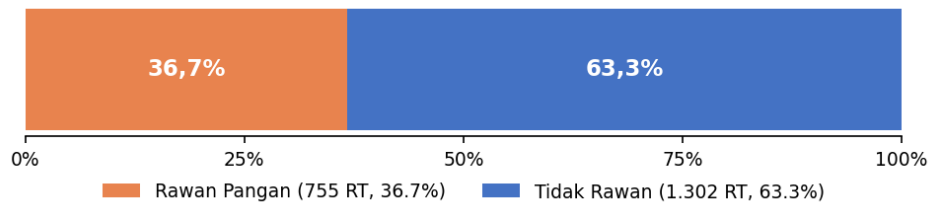
## 3. HASIL DAN PEMBAHASAN

### 3.1 Data dan Variabel Penelitian

Data yang digunakan berasal dari SUSENAS 2024 BPS Kota Langsa, dengan 2.057 rumah tangga sampel. Variabel target (Y) ditetapkan berdasarkan indikator berat FIES: sebuah rumah tangga dikategorikan rawan pangan apabila menjawab 'Ya' pada seluruh tiga indikator berat secara bersamaan — Y1704 (makan lebih sedikit dari seharusnya), Y1705 (kehabisan persediaan makanan), dan Y1706 (ada anggota yang tidak makan sehari-hari).

Definisi berat  $\text{count} \geq 3$  dipilih karena ambang yang lebih longgar ( $\geq 1$  indikator) menghasilkan proporsi kelas rawan 99,9%, yang secara statistik tidak layak dimodelkan. Dengan ambang ini, distribusi kelas

menjadi proporsional: 755 rawan (36,7%) dan 1.302 tidak rawan (63,3%), dengan rasio 1:1,72 yang masih dapat ditangani metode standar. Gambar 1 memperlihatkan distribusi tersebut secara visual.



Gambar 1. Distribusi Kelas Target — Rawan vs Tidak Rawan Pangan

Tabel 2 merinci seluruh variabel prediktor yang digunakan beserta tipe datanya.

Kode	Tipe	Keterangan
X1804	Kontinu	Luas lantai rumah (m <sup>2</sup> )
X1807	Kategorik	Jenis dinding dominan
X1808	Kategorik	Jenis lantai dominan
X1809A	Kategorik	Fasilitas buang air besar
X1810A	Kategorik	Sumber air minum utama
X1816	Kategorik	Sumber penerangan listrik
X1817	Kategorik	Bahan bakar memasak
X701	Kategorik	Kepemilikan rekening bank
X2101A	Kategorik	Lapangan usaha penghasilan utama
X2202	Kategorik	Status kepemilikan lahan
X2203	Kategorik	Penerima program PKH
X2211A	Kategorik	Penerima bantuan sosial rutin
X2211B	Kategorik	Penerima bantuan sosial tidak rutin

### 3.2 Preprocessing Data

Sebelum model dilatih, beberapa langkah preprocessing diterapkan secara berurutan sebagaimana dirangkum pada Tabel 3. Gambar 2 menyajikan alur preprocessing dari data mentah hingga siap dilatih dalam bentuk diagram alir.

Table 3 Tahapan Preprocessing Data

Tahap	Target	Perlakuan
Missing values	Variabel kategorik	Imputasi nilai modus
Encoding	Semua variabel kategorik	Label encoding (bilangan bulat)
Normalisasi	X1804 (luas lantai)	Scaling min-max ke rentang [0, 1]
Resampling	Data training saja	EasyEnsemble (T = 10 subset)
Pembagian	Seluruh dataset	Training 70% / Testing 30%



Gambar 2. Diagram Alir Tahapan Preprocessing Data

Imputasi dengan modus dipilih karena hampir semua variabel bersifat kategorik, sehingga nilai modus paling mewakili kondisi mayoritas. Label encoding mengubah kategori menjadi bilangan bulat agar dapat diproses algoritma berbasis pohon. Normalisasi min-max hanya diterapkan pada satu variabel kontinu (X1804) agar tidak mendominasi perhitungan impuritas Gini. Resampling EasyEnsemble hanya diterapkan pada data training — data testing dibiarkan mencerminkan distribusi aslinya untuk menghindari kebocoran data yang dapat menggelembungkan estimasi performa [4].

### 3.3 Prosedur Eksperimen

Data dibagi dengan rasio 70:30 (training:testing) secara acak. Pengulangan sebanyak 50 kali dengan seed berbeda dilakukan untuk mendapatkan estimasi performa yang stabil dan tidak bergantung pada satu konfigurasi pembagian tertentu. Parameter model: RF dan DRF menggunakan  $n\_estimators = 100$ ,  $max\_features = 'sqrt'$ ,  $criterion = 'gini'$ ; DRF menerapkan node-bootstrap threshold 10% sesuai Han et al. (2020) [5]; EasyEnsemble menggunakan  $n\_subsets = 10$  dengan Random Forest sebagai base learner pada tiap subset yang seimbang.

### 3.4 Ilustrasi Perhitungan RF

Berikut ilustrasi satu pohon pada iterasi pertama. Misalkan bootstrap menghasilkan 200 sampel dari data training. Pada node akar, dipilih  $m = \sqrt{13} \approx 4$  fitur acak:

{X2211A, X1804, X1808, X701}. Untuk fitur X2211A (ya/tidak):

Kiri (X2211A = 'Ya'): 80 rawan, 20 tidak rawan

$$Gini\_kiri = 1 - [(80/100)^2 + (20/100)^2] = 0,32 \tag{9}$$

Kanan (X2211A = 'Tidak'): 10 rawan, 90 tidak rawan

$$\text{Gini\_kanan} = 1 - \frac{[(10/100)^2 + (90/100)^2]}{0,18} = \quad (10)$$

$$\text{Gini\_split} = (100/200) \times 0,32 + (100/200) \times 0,18 = 0,25 \quad (11)$$

X2211A dipilih sebagai pemisah terbaik karena menghasilkan Gini\_split terendah. Proses ini berulang di setiap simpul hingga pohon mencapai batas yang ditetapkan. Keputusan akhir diperoleh dari voting mayoritas B = 100 pohon (persamaan 1).

### 3.5 Ilustrasi Perhitungan DRF

DRF menambahkan resampling di setiap simpul. Setelah pemisahan pertama menghasilkan 100 sampel di partisi kiri, DRF melakukan bootstrap ulang atas 100 sampel tersebut. Misalkan hasilnya bergeser menjadi 85 rawan dan 15 tidak rawan. Dengan fitur baru {X1804, X1807, X1809A, X2203}, jika X1804 memberikan pemisahan terbaik:

$$\text{Gini}(X1804 \leq 36m^2) = (60/85) \times 0,24 + (25/85) \times 0,12 = 0,206 \quad (12)$$

Pada data kategorik dominan seperti SUSENAS Kota Langsa, resampling tambahan ini tidak banyak mengubah struktur pohon karena fitur yang paling informatif (X2211A, X2211B) tetap mendominasi pilihan — hal inilah yang menjelaskan rendahnya sensitivitas DRF.

### 3.6 Ilustrasi Perhitungan EasyEnsemble

Dari 755 sampel rawan dan 1.302 tidak rawan, EasyEnsemble membentuk  $T = 10$  subset. Tiap subset berisi 755 sampel acak dari kelas mayoritas digabung dengan 755 sampel minoritas  $\rightarrow$  1.510 sampel seimbang. Pada tiap subset dilatih satu model Random Forest; prediksi akhir diperoleh dari rata-rata probabilitas seluruh model. Kunci keunggulannya adalah setiap sampel mayoritas terwakili dalam minimal satu subset sehingga tidak ada informasi yang benar-benar terbuang.

$$\text{Prediksi\_ensemble}(x) = (1/T) \cdot \sum_{t=1}^T P_t(\text{rawan} | x) \quad (13)$$

di mana  $P_t$  adalah probabilitas prediksi model ke- $t$ , dan  $T = 10$  subset. Probabilitas rata-rata di atas ambang 0,5 diklasifikasikan sebagai rawan. Kunci keunggulannya: setiap sampel mayoritas terwakili dalam minimal satu subset, sehingga tidak ada informasi yang benar-benar terbuang, dan sensitivitas terhadap kelas minoritas meningkat signifikan. variabel prediktor hanya 1 yang kontinu sehingga bootstrap berulang di setiap simpul yang menjadi keunggulan DRF tidak menghasilkan diversitas pemisahan yang bermakna [7].

Peningkatan sensitivitas RF + EasyEnsemble dari 64,70% (RF biasa) menjadi 79,84% sangat krusial secara kebijakan: model mampu mendeteksi hampir 8 dari 10 rumah tangga yang benar-benar rawan pangan. Dalam konteks intervensi pemerintah, false negative (melewatkan rumah tangga rawan) memiliki konsekuensi sosial yang jauh lebih besar dibandingkan false positive (over-prediksi) [12].

### 3.7 Perbandingan Performa Model

Tabel 4 merangkum rata-rata performa ketiga model dari 50 repetisi. Sel berwarna hijau menandai model terbaik pada tiap metrik.

Table 4 Rata-rata Performa Model (50 repetisi)

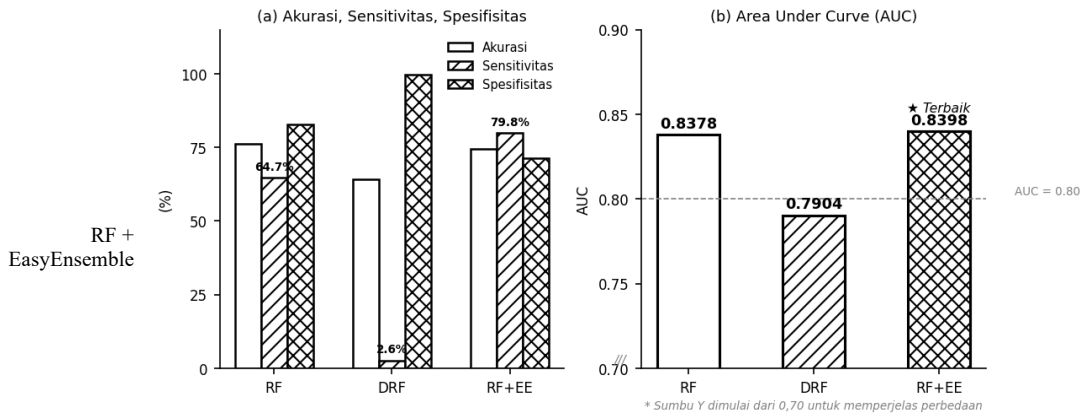
Model	AUC	Akurasi	Sensit.	Spesif.
Random Forest (RF)	0,8378	76,11%	64,70%	82,73%
Double Random Forest	0,7904	64,10%	2,62%	99,79%
RF + EasyEnsemble	0,8398	74,49%	79,84%	71,39%

RF + EasyEnsemble unggul pada dua metrik yang paling relevan: AUC (0,8398) dan sensitivitas (79,84%). DRF justru menghasilkan sensitivitas yang sangat rendah hanya 2,62% meski spesifisitasnya mendekati sempurna (99,79%), yang artinya model hampir selalu memprediksi semua rumah tangga sebagai tidak rawan dan secara praktis tidak berguna untuk deteksi dini. Ini konsisten dengan temuan Rais et al. (2024) yang juga melaporkan kelemahan DRF pada data kategorik dominan [7].

Peningkatan sensitivitas dari RF biasa (64,70%) ke RF + EasyEnsemble (79,84%) cukup substansial. Dalam konteks pengambilan kebijakan, melewatkan rumah tangga rawan pangan (false negative) berdampak sosial jauh lebih serius daripada memprediksi secara berlebihan (false positive) [12].

### 3.8 Visualisasi Performa

Gambar 3 dan 4 menyajikan perbandingan sensitivitas dan AUC ketiga model secara visual untuk memudahkan pembacaan tren antar model.



Gambar 3. Perbandingan Performa Ketiga Model: (a) Akurasi, Sensitivitas, Spesifisitas dan (b) AUC

Gambar 3 panel (a) menegaskan kesenjangan sensitivitas yang sangat mencolok: DRF hanya 2,62% berbanding 79,84% pada RF + EasyEnsemble. Panel (b) menunjukkan bahwa selisih AUC antara RF (0,8378) dan RF + EasyEnsemble (0,8398) sebenarnya sangat kecil, namun keduanya sama-sama jauh di atas DRF (0,7904). Dipadukan dengan Gambar 4 (uji Tukey) dan Gambar 5 (feature importance), ketiga visualisasi ini membentuk narasi lengkap: performa terbaik diraih RF + EasyEnsemble, perbedaannya signifikan secara statistik, dan prediktor dominannya adalah variabel bantuan sosial.

### 3.9 Uji Signifikansi Statistik

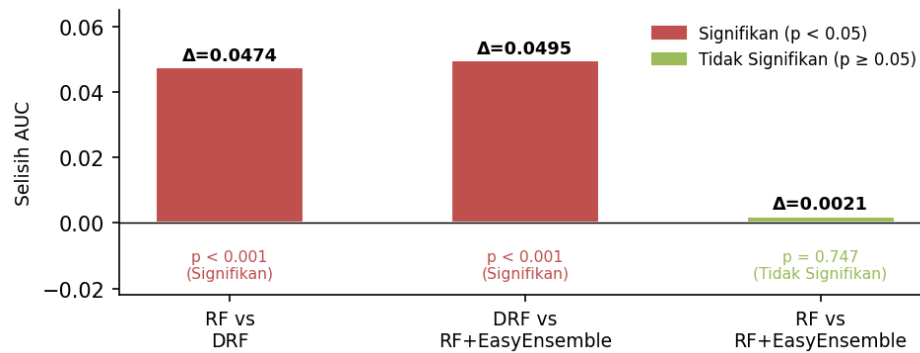
Uji ANOVA satu arah terhadap distribusi nilai AUC dari 50 repetisi menghasilkan:

$$F = \text{MSB}/\text{MSW} = 0,039175/0,000204 = 191,899 \quad (p < 0,001) \quad (14)$$

F = 191,899 jauh melampaui F kritis = 3,06 pada df = (2, 147), sehingga H<sub>0</sub> ditolak — setidaknya satu pasangan model berbeda secara signifikan. Uji lanjutan Tukey HSD (q kritis = 3,36; HSD = 0,0068) disajikan pada Tabel 5. Gambar 5 memvisualisasikan hasil signifikansi antar pasangan model.

Tabel 5. Hasil Uji Post-hoc Tukey HSD

Pasangan Model	Selisih AUC	p-value	Simpulan
RF vs DRF	0,0474	0,001	Signifikan
DRF vs RF+EasyEnsemble	0,0495	0,001	Signifikan
RF vs RF+EasyEnsemble	0,0021	0,747	Tidak Signifikan



Gambar 4. Visualisasi Signifikansi Statistik Uji Tukey HSD

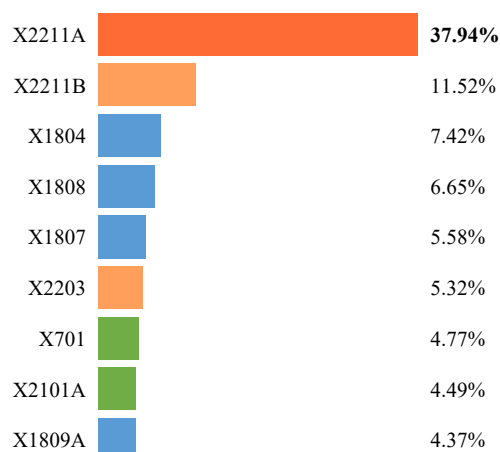
Tiga poin kunci dari hasil ini: Pertama, RF secara statistik lebih unggul dari DRF ( $p = 0,001$ ), menunjukkan mekanisme DRF tidak memberikan nilai tambah pada data kategorik ini. Kedua, RF + EasyEnsemble lebih baik dari DRF ( $p = 0,001$ ), membuktikan pentingnya penanganan ketidakseimbangan kelas. Ketiga, selisih AUC antara RF dan RF + EasyEnsemble tidak signifikan secara statistik ( $p = 0,747$ ) — namun perbedaan sensitivitasnya sangat nyata secara praktis (79,84% vs 64,70%), menegaskan bahwa AUC saja tidak cukup sebagai satu-satunya tolok ukur pemilihan model [11].

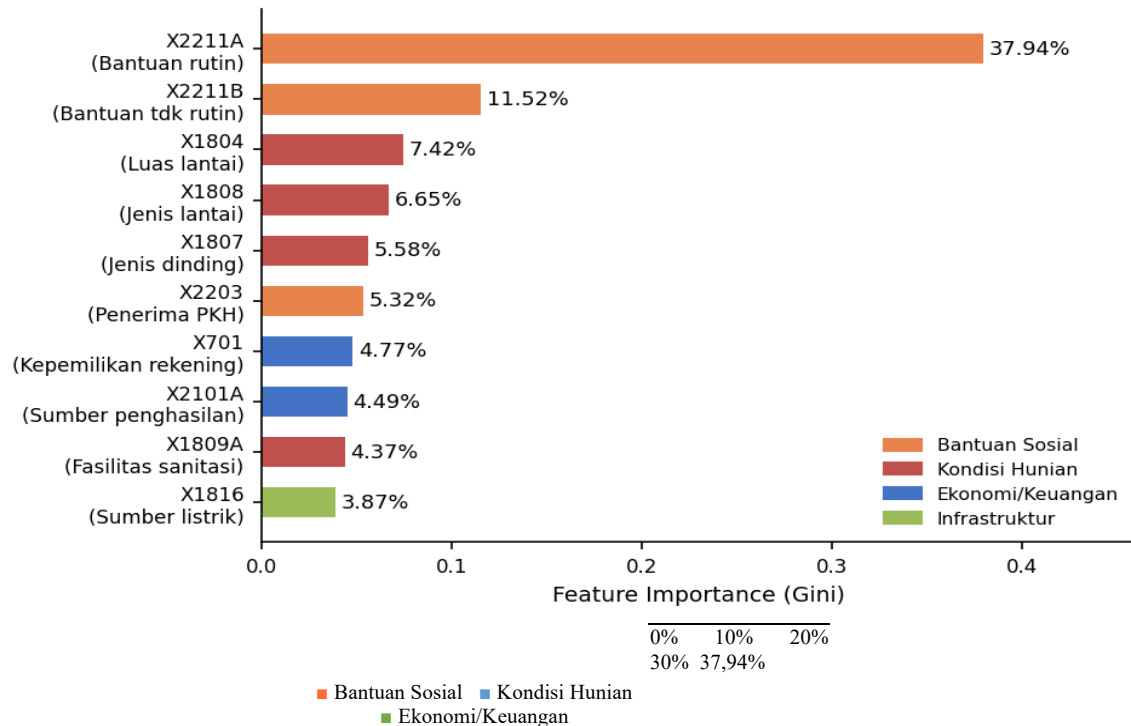
### 3.10 Analisis Feature Importance

Tabel 6 menampilkan peringkat 13 variabel berdasarkan nilai feature importance model RF + EasyEnsemble mencerminkan rata-rata penurunan impuritas Gini yang dikontribusikan tiap variabel di seluruh pohon. Gambar 6 memvisualisasikan 10 variabel teratas dalam bentuk horizontal bar chart berwarna berdasarkan kategori faktor.

Tabel 6. Peringkat Feature Importance (RF + EasyEnsemble)

No	Kode	Variabel	Imp.	%
1	X2211A	Bantuan sosial rutin	0,3794	37,94%
2	X2211B	Bantuan sosial tidak rutin	0,1152	11,52%
3	X1804	Luas lantai rumah	0,0742	7,42%
4	X1808	Jenis lantai	0,0665	6,65%
5	X1807	Jenis dinding	0,0558	5,58%
6	X2203	Penerima PKH	0,0532	5,32%
7	X701	Kepemilikan rekening	0,0477	4,77%
8	X2101A	Sumber penghasilan	0,0449	4,49%
9	X1809A	Fasilitas sanitasi	0,0437	4,37%
10	X1816	Sumber listrik	0,0387	3,87%
11–13	(3 var)	Sumber air, BB, lahan	—	8,07%





Gambar 5. Horizontal Bar Chart Feature Importance 10 Variabel Teratas

Dua kelompok faktor dominan teridentifikasi. Kelompok pertama: akses terhadap program perlindungan sosial, yang secara kolektif menyumbang 54,78% (X2211A: 37,94% + X2211B: 11,52% + X2203: 5,32%). Dominasi X2211A mencerminkan bahwa status sebagai penerima bantuan sosial rutin merupakan penanda terkuat kerentanan ekonomi memang dapat dipahami, karena program ini dirancang menyoal rumah tangga dengan kondisi paling lemah.

Kelompok kedua: kondisi fisik dan infrastruktur hunian, dengan kontribusi gabungan 28,89% (X1804 + X1808 + X1807 + X1809A + X1816). Temuan ini sejalan dengan Ramadhani et al. (2022) dan Dharmawan et al. (2022) yang menegaskan kualitas tempat tinggal sebagai cerminan langsung kemiskinan struktural [3][4]. Kepemilikan rekening bank (X701, 4,77%) menunjukkan peran inklusi keuangan, sementara sumber penghasilan utama (X2101A, 4,49%) mengindikasikan bahwa ketergantungan pada sektor informal meningkatkan ketidakpastian pemenuhan kebutuhan pangan.

### 3.11 Implikasi Kebijakan

Setidaknya ada tiga implikasi praktis yang dapat ditarik. Pertama, data kepesertaan program bantuan sosial yang sudah tersedia di tingkat pemerintah daerah dapat langsung dijadikan lapisan awal pemetaan risiko tanpa survei tambahan yang perlu dievaluasi adalah cakupannya, agar rumah tangga eligible yang belum terdaftar dapat segera diidentifikasi. Kedua, program peningkatan kualitas hunian seperti bedah rumah dan penyediaan sanitasi layak memiliki manfaat ganda: meningkatkan kenyamanan sekaligus secara tidak langsung menekan risiko kerawanan pangan. Ketiga, fasilitasi akses layanan keuangan formal misalnya rekening tabungan berbiaya nol merupakan strategi jangka menengah yang relevan untuk memperkuat ketahanan ekonomi rumah tangga rentan.

## 4. KESIMPULAN

Penelitian ini menghasilkan empat temuan utama. Pertama, penetapan variabel target berdasarkan ambang berat  $\text{count} \geq 3$  menghasilkan distribusi kelas yang jauh lebih proporsional (36,7% rawan vs 63,3% tidak rawan) dibandingkan definisi yang lebih longgar, sehingga lebih valid untuk dimodelkan secara statistik. Kedua, pada dataset SUSENAS Kota Langsa yang hampir seluruhnya terdiri dari variabel kategorik, DRF justru menghasilkan performa terendah: AUC 0,7904 dan sensitivitas hanya 2,62%. Keunggulan teoritis bootstrap ganda DRF tidak terwujud ketika fitur-fitur dominan bersifat biner atau nominal. Ketiga, penanganan ketidakseimbangan kelas terbukti menjadi faktor penentu. RF + EasyEnsemble mencapai performa terbaik (AUC 0,8398; sensitivitas 79,84%) dan keunggulannya dikonfirmasi secara statistik melalui ANOVA ( $F = 191,899$ ;  $p < 0,001$ ) serta uji Tukey HSD. Model ini mampu mengidentifikasi hampir delapan dari sepuluh rumah tangga yang benar-benar rawan pangan. Keempat, feature importance mengungkap bahwa kerawanan

pangan di Kota Langsa mencerminkan kemiskinan yang lebih mendasar: ketergantungan pada bantuan sosial (54,78%) dan kondisi hunian yang buruk (28,89%) adalah prediktor utama memberikan landasan empiris yang kuat untuk merancang intervensi yang lebih terarah dan efisien.

Penelitian lanjutan disarankan menguji DRF pada dataset dengan lebih banyak variabel kontinu, membandingkan dengan algoritma gradient boosting modern (XGBoost, LightGBM), serta memperluas cakupan ke seluruh kabupaten/kota di Provinsi Aceh.

## REFERENSI

- [1] Badan Pusat Statistika, "Prevalensi Penduduk Dengan Kerawanan Pangan Sedang Atau Berat," BPS, 2023.
- [2] E. Ramadhani, B. Sartono, A. F. Hadi, S. 'Ufa, dan T. Akhdansyah, "Comparison of Main Characteristics of Food Insecurity Using Classification Tree and Random Forest," *Sinkron*, vol. 7, no. 4, pp. 2486–2497, 2022, doi: 10.33395/sinkron.v7i4.11852.
- [3] H. Dharmawan, B. Sartono, A. Kurnia, A. F. Hadi, dan E. Ramadhani, "A Study of Machine Learning Algorithms To Measure The Feature Importance In Class-Imbalance Data of Food Insecurity Cases Indonesia," *Commun. Math. Biol. Neurosci.*, 2022, doi: 10.28919/cmbn/7636.
- [4] H. Du, Y. Zhang, L. Zhang, dan Y. C. Chen, "Selective Ensemble Learning Algorithm for Imbalanced Dataset," *Comput. Sci. Inf. Syst.*, vol. 20, no. 2, pp. 831–856, 2023, doi: 10.2298/CSIS220817023D.
- [5] S. Han, H. Kim, dan Y. S. Lee, "Double random forest," *Mach. Learn.*, vol. 109, no. 8, pp. 1569–1586, 2020, doi: 10.1007/s10994-020-05889-1.
- [6] X. Y. Liu dan Z. H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 39, no. 2, pp. 539–550, 2009.
- [7] A. M. S. Rais, B. Susetyo, dan B. Sartono, "Rotation Double Random Forest Algorithm to Predict the Food Insecurity Status of Households," *J. RESTI*, vol. 8, no. 1, pp. 33–41, 2024, doi: 10.29207/resti.v8i1.5540.
- [8] A. Fhonna, A. Mukhlis, dan Fadlisyah, "Prediksi Kerawanan Pangan Menggunakan Algoritma Machine Learning Berbasis Data SUSENAS," *J. Inform. Univ. Pamulang*, vol. 8, no. 2, 2023.
- [9] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [10] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001. (Feature importance via mean decrease impurity is described in the original RF paper.)
- [11] J. C. Obi, "A comparative study of several classification metrics," *World J. Adv. Eng. Technol. Sci.*, vol. 8, no. 1, pp. 308–314, 2023, doi: 10.30574/wjaets.2023.8.1.0054.
- [12] Z. Du, H. Fang, dan W. Yang, "Predicting Food Insecurity Using Machine Learning: Evidence from Household Survey Data," *Appl. Artif. Intell.*, vol. 37, no. 2, 2023, doi: 10.1080/08839514.2023.2171568.
- [13] Food and Agriculture Organization, "The Food Insecurity Experience Scale," FAO, 2022.
- [14] Food and Agriculture Organization, "The State of Food Security and Nutrition in the World 2023," FAO, 2023.
- [15] R. M. Mohana, C. K. K. Reddy, P. R. Anisha, dan B. V. R. Murthy, "Random Forest algorithms for the classification of tree-based ensemble," *Mater. Today Proc.*, 2021.