

- 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.

KAJIAN TINGKAT KEPENTINGAN PEUBAH DENGAN TREESHAP DAN ER-SHAP PADA KLASIFIKASI RESILIENSI PANGAN SELAMA PANDEMI COVID-19

NABILA SYUKRI



**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR**

**BOGOR
2026**





PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Kajian Tingkat Kepentingan Peubah dengan TreeSHAP dan ER-SHAP pada Klasifikasi Resiliensi Pangan selama Pandemi Covid-19” 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 tesis ini.

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

Bogor, Juli 2026

Nabila Syukri
M0501241086

- 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 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 2026
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.

RINGKASAN

NABILA SYUKRI. Kajian Tingkat Kepentingan Peubah dengan TreeSHAP dan ER-SHAP pada Klasifikasi Resiliensi Pangan selama Pandemi Covid-19. Dibimbing oleh SACHNAZ DESTA OKTARINA dan SEPTIAN RAHARDIANTORO

Pandemi COVID-19 menyebabkan gangguan terhadap sistem pangan dan kondisi sosial ekonomi rumah tangga di Indonesia. Penurunan pendapatan, perubahan aktivitas kerja, keterbatasan akses layanan kesehatan, serta berbagai strategi adaptasi rumah tangga berpotensi memengaruhi kemampuan rumah tangga dalam mempertahankan maupun memulihkan kondisi ketahanan pangannya (resiliensi pangan). Kompleksitas hubungan antar faktor tersebut menyebabkan pendekatan statistik konvensional kurang mampu menangkap hubungan yang bersifat nonlinier, sedangkan model *machine learning* (ML) yang memiliki kemampuan prediksi tinggi umumnya bersifat *black-box*. Oleh karena itu, penelitian ini mengintegrasikan model klasifikasi berbasis ML dengan pendekatan *Explainable Artificial Intelligence* (XAI) untuk memperoleh model yang akurat sekaligus mudah diinterpretasikan.

Penelitian ini bertujuan mengevaluasi kinerja algoritma XGBoost, LightGBM, dan CatBoost dalam mengklasifikasikan resiliensi pangan rumah tangga, menganalisis hubungan peringkat peubah penting antaralgoritma dan antarmetode XAI, serta mengidentifikasi peubah-peubah yang berperan penting dalam menentukan status resiliensi pangan rumah tangga. Data yang digunakan berasal dari High Frequency Phone Survey (HFPS) Indonesia yang dilaksanakan oleh World Bank pada delapan periode survei tahun 2020–2023. Analisis dilakukan terhadap 2.868 rumah tangga yang memiliki data lengkap pada seluruh periode pengamatan. Variabel respons dibentuk menggunakan dua definisi resiliensi pangan, yaitu *Strict* dan *Broad*, sedangkan 96 peubah hasil survei diringkas menjadi 42 peubah prediktor melalui *feature engineering*, *summary features*, dan ekstraksi fitur berbasis teks.

Pemodelan dilakukan menggunakan algoritma XGBoost, LightGBM, dan CatBoost. Ketidakseimbangan kelas pada definisi *Strict* ditangani menggunakan oversampling using propensity score (OUPS), sedangkan optimasi hyperparameter dilakukan menggunakan Optuna dengan *stratified 5-fold cross validation*. Evaluasi model dilakukan melalui sepuluh kali pengulangan pembagian data menggunakan metrik *Balanced Accuracy*, *F1-Score*, sensitivitas, dan *Area Under the Curve* (AUC). Interpretasi model dilakukan menggunakan TreeSHAP dan Ensemble Random-SHAP (ER-SHAP). Hubungan peringkat peubah dianalisis menggunakan koefisien korelasi Spearman, kemudian diagregasi menggunakan Borda Count untuk memperoleh peringkat konsensus.

Hasil penelitian menunjukkan bahwa seluruh algoritma menghasilkan performa klasifikasi yang baik dengan nilai AUC di atas 0,80 pada kedua definisi resiliensi pangan. Penerapan OUPS meningkatkan kemampuan model dalam mengenali kelas minoritas pada definisi *Strict*, yang ditunjukkan oleh peningkatan sensitivitas dan *Balanced Accuracy*. Analisis hubungan peringkat peubah menunjukkan korelasi Spearman yang tinggi, baik antaralgoritma maupun antarmetode XAI, dengan korelasi TreeSHAP dan ER-SHAP berada di atas 0,93.

Agregasi menggunakan Borda Count mengidentifikasi tujuh peubah yang secara konsisten memiliki tingkat kepentingan tinggi pada kedua definisi resiliensi pangan, yaitu kondisi kerawanan pangan sebelum pandemi, tingkat kekhawatiran terhadap kondisi keuangan, rata-rata pendapatan rumah tangga, guncangan harga beras, frekuensi hambatan memperoleh layanan kesehatan, variasi jam kerja, dan frekuensi perubahan pekerjaan.

Penelitian ini menunjukkan bahwa pendekatan *Explainable Machine Learning* mampu menghasilkan model klasifikasi dengan performa yang baik sekaligus memberikan interpretasi yang kuat terhadap faktor-faktor yang memengaruhi resiliensi pangan rumah tangga. Hasil penelitian ini diharapkan dapat menjadi dasar dalam penyusunan kebijakan perlindungan sosial dan ketahanan pangan yang lebih tepat sasaran untuk meningkatkan ketahanan rumah tangga dalam menghadapi krisis di masa mendatang.

Kata kunci: COVID-19, explainable machine learning, gradient boosting, resiliensi pangan, XAI

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.

SUMMARY

NABILA SYUKRI. Study of Variable Importance Using TreeSHAP and ER-SHAP in Food Resilience Classification during the Covid-19 Pandemic. Supervised by SACHNAZ DESTA OKTARINA and SEPTIAN RAHARDIANTORO

The COVID-19 pandemic has disrupted the food system and socioeconomic conditions of households in Indonesia. Declining income, changes in work activities, limited access to healthcare, and various household adaptation strategies have the potential to impact households' ability to maintain and restore food security (food resilience). The complexity of the relationships between these factors makes conventional statistical approaches incapable of capturing nonlinear relationships, while machine learning models with high predictive capabilities are generally black-box. Therefore, this study integrates a machine learning-based classification model with an Explainable Artificial Intelligence (XAI) approach to obtain a model that is both accurate and easy to interpret.

This study aims to evaluate the performance of the XGBoost, LightGBM, and CatBoost algorithms in classifying household food resilience, analyze the relationship between the rankings of important variables between the algorithms and the XAI methods, and identify variables that play a significant role in determining household food resilience status. The data used comes from the High Frequency Phone Survey (HFPS) Indonesia, conducted by the World Bank over eight survey periods from 2020 to 2023. The analysis was conducted on 2,868 households with complete data for the entire observation period. The response variable was formed using two definitions of food resilience: Strict and Broad. The 96 survey variables were summarized into 42 predictor variables through feature engineering, summary features, and text-based feature extraction.

Modeling was performed using the XGBoost, LightGBM, and CatBoost algorithms. Class imbalance in the Strict definition was addressed using oversampling using propensity score (OUPS), while hyperparameter optimization was performed using Optuna with stratified 5-fold cross-validation. Model evaluation was performed through ten iterations of data splitting using Balanced Accuracy, F1-Score, sensitivity, and Area Under the Curve (AUC) metrics. Model interpretation was performed using TreeSHAP and Ensemble Random-SHAP (ER-SHAP). The relationship between variable ranks was analyzed using Spearman's correlation coefficient, then aggregated using the Borda Count to obtain a consensus ranking.

The results showed that all algorithms produced good classification performance, with AUC values above 0.80 for both definitions of food resilience. The application of OUPS improved the model's ability to recognize minority classes in the Strict definition, as demonstrated by increased sensitivity and Balanced Accuracy. Variable rank correlation analysis demonstrated high Spearman correlations, both between algorithms and between XAI methods, with TreeSHAP and ER-SHAP correlations above 0.93. Aggregation using the Borda Count identified seven variables with consistently high importance for both definitions of food resilience: pre-pandemic food insecurity, level of financial concern, average household income, rice price shocks, frequency of barriers to accessing healthcare, variation in working hours, and frequency of job changes.



This study demonstrates that the Explainable Machine Learning approach is capable of producing classification models with good performance while providing robust interpretations of the factors influencing household food resilience. These results are expected to inform the development of more targeted social protection and food security policies to improve household resilience in the face of future crises.

Keywords: COVID-19, explainable machine learning, gradient boosting, food resilience, XAI

1. Hak Cipta Dilindungi Undang-undang
 - a. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - b. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - c. 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.

KAJIAN TINGKAT KEPENTINGAN PEUBAH DENGAN TREESHAP DAN ER-SHAP PADA KLASIFIKASI RESILIENSI PANGAN SELAMA PANDEMI COVID-19

NABILA SYUKRI

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

**PROGRAM STUDI MAGISTER 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.

Penguji Luar Komisi Pembimbing pada Ujian Tesis:
Dr. Ir. Budi Susetyo, M.S.



Judul Tesis : Kajian Tingkat Kepentingan Peubah dengan TreeSHAP dan ER-SHAP pada Klasifikasi Resiliensi Pangan selama Pandemi Covid-19
Nama : Nabila Syukri
NIM : M0501241086

Disetujui oleh

Pembimbing 1:

Sachnaz Desta Oktarina, S.Stat., M.Agr.Sc, Ph.D



Pembimbing 2:

Dr. Septian Rahardiantoro, S.Stat., M.Si



Diketahui oleh

Ketua Program Studi:

Dr. Agus Mohamad Soleh, S.Si, MT

NIP 197503151999031004



Dekan Sekolah Sains Data, Matematika dan Informatika:

Prof. Dr. Ir. Agus Buono, M.Si., M.Kom

NIP 19660702 1993021001



Tanggal Ujian: 23 Juli 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga tesis ini dapat diselesaikan dengan baik. Tesis ini disusun sebagai bagian dari pemenuhan tugas akhir dengan judul “Kajian Tingkat Kepentingan Peubah dengan TreeSHAP dan ER-SHAP pada Klasifikasi Resiliensi Pangan selama Pandemi Covid-19”.

Dalam proses penyusunan tesis ini, penulis memperoleh banyak bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih dan penghargaan setinggi-tingginya kepada:

- Ibu Sachnaz Desta Oktarina S.Stat., M.Agr.Sc, Ph.D dan Bapak Dr. Septian Rahardianto, M.Si selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta saran yang sangat berharga selama proses penyusunan makalah penelitian tesis ini.
- Bapak Dr. Ir. Budi Susetyo, M.S selaku penguji yang telah memberikan masukan dan saran penyempurnaan makalah penelitian ini.
- Seluruh Tenaga Pendidik dan Kependidikan Departemen Statistika IPB University atas ilmu, dukungan, dan fasilitas yang telah diberikan kepada penulis.
- Orang tua, sahabat, kerabat dan seluruh keluarga yang memberikan dukungan, doa, dan kasih sayang yang senantiasa menyertai penulis selama proses akademik dan penyusunan penelitian ini.
- Lembaga Pengelola Dana Pendidikan (LPDP) yang telah memberikan dukungan beasiswa selama masa studi.
- Teman-teman Program Studi Statistika dan Sains Data Angkatan 2024 yang telah memberikan dukungan, semangat, serta kebersamaan selama proses pembelajaran dan penyelesaian penelitian.

Penulis menyadari bahwa penelitian ini masih memiliki keterbatasan. Oleh karena itu, penulis membuka diri terhadap saran dan kritik yang membangun demi kesempurnaan karya ini di masa mendatang.

Akhir kata, semoga penelitian ini dapat memberikan manfaat bagi semua pihak yang membutuhkan dan berkontribusi nyata bagi pengembangan ilmu pengetahuan dan perumusan kebijakan berbasis data di Indonesia.

Bogor, Juli 2026

Nabila Syukri



DAFTAR ISI

DAFTAR ISI	xii
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
TINJAUAN PUSTAKA	5
2.1 Keamanan dan Kerawanan Pangan	5
2.2 Resiliensi Pangan	5
2.3 <i>Gradient Boosting</i>	6
2.4 <i>Extreme Gradient Boosting (XGBoost)</i>	7
2.5 <i>Light Gradient Boosting Machine (LightGBM)</i>	8
2.6 <i>Categorical Boosting (CatBoost)</i>	10
2.7 <i>Hyperparameter</i>	10
2.8 Ekstraksi Peubah	11
2.9 <i>Oversampling Using Propensity Score (OUPS)</i>	12
2.10 <i>Shapley Additive Explanation (SHAP)</i>	13
2.11 TreeSHAP	14
2.12 <i>Ensemble Random – SHAP (ER-SHAP)</i>	15
2.13 Evaluasi Model Klasifikasi	16
2.14 Korelasi Spearman	17
2.15 Agregasi Peringkat Borda	18
III METODE	19
3.1 Sumber Data	19
3.2 Peubah dan Unit Pengamatan	19
3.3 Tahapan Analisis	21
IV HASIL DAN PEMBAHASAN	24
4.1 Eksplorasi Peubah Respons	24
4.2 Eksplorasi Dinamika Sosial Ekonomi Rumah Tangga	25
4.3 Proses Ekstraksi Peubah Prediktor	28
4.4 Pemodelan Klasifikasi	31
4.5 Analisis Interpretabilitas Model Klasifikasi	36
4.6 Interpretasi Faktor yang Memengaruhi Status Resiliensi Pangan	40
4.7 Implikasi Faktor yang memengaruhi Resiliensi Pangan	43
V KESIMPULAN DAN SARAN	44
5.1 Kesimpulan	44
5.2 Saran	44
DAFTAR PUSTAKA	46

LAMPIRAN	50
RIWAYAT HIDUP	54

DAFTAR TABEL

1	Item Pertanyaan <i>Food Insecurity Experience Scale</i> (FIES)	5
2	<i>Hyperparameter</i> yang digunakan pada model <i>boosting</i>	11
3	<i>Confusion matrix</i>	17
4	Klasifikasi metrik AUC	17
5	Peubah mentah HFPS yang digunakan dalam penelitian	20
6	Proporsi rumah tangga berdasarkan peubah bantuan sosial, strategi <i>coping</i> , dan akses kesehatan per periode	28
7	Hasil ekstraksi peubah berdasarkan <i>feature engineering</i>	29
8	Hasil ekstraksi peubah berdasarkan <i>summary features</i>	29
9	Ringkasan jumlah peubah sebelum dan setelah ekstraksi	30
10	Inisialisasi <i>hyperparameter</i>	31
11	Distribusi data sebelum dan setelah penanganan OUPS	33
12	Evaluasi kinerja model <i>Strict</i>	34
13	Evaluasi kinerja model <i>Broad</i>	34
14	Koefisien korelasi spearman antar-algoritma	36
15	Koefisien korelasi spearman antar-metode XAI	37
16	Daftar sepuluh peubah teratas hasil agregasi Borda	38
17	Proporsi rumah tangga resilien berdasarkan peubah status rawan sebelum pandemi dan khawatir keuangan	40
18	Proporsi rumah tangga resilien berdasarkan peubah <i>shock</i> harga beras	41
19	Proporsi rumah tangga resilien berdasarkan frekuensi sakit tidak dirawat	42
20	Proporsi rumah tangga resilien berdasarkan frekuensi ganti kerja	42

DAFTAR GAMBAR

1	Kerangka konseptual resiliensi pangan	6
2	Ilustrasi <i>gradient boosting</i>	6
3	<i>Level-wise growth vs leaf-wise growth</i>	9
4	Ilustrasi GOSS dan EFB pada LightGBM	9
5	Ilustrasi <i>ensemble random-SHAP</i> (ER-SHAP)	15
6	Kriteria pembentukan peubah respons	19
7	Diagram alir analisis	21
8	Proporsi status resiliensi pangan (a) <i>Strict</i> dan (b) <i>Broad</i>	24
9	Proporsi rumah tangga berstatus bekerja pada setiap periode	25
10	Distribusi total jam kerja rumah tangga per periode	26
11	Distribusi pendapatan rumah tangga per periode	26

12	Tingkat kekhawatiran rumah tangga terhadap COVID-19 dan keuangan	27
13	Eksplorasi persepsi kenaikan harga komoditas dengan <i>word cloud</i> (a) dan frekuensi kata (b)	30
14	Distribusi skor propensitas pada data latih	32
15	Perbandingan kinerja model <i>Strict</i> dan <i>Broad</i>	35
16	Perbandingan skor borda <i>Strict</i> dan <i>Broad</i>	39
17	Distribusi rata-rata pendapatan menurut status resiliensi pangan pada definisi (a) <i>Strict</i> dan (b) <i>Broad</i>	41
18	Distribusi variasi jam kerja menurut status resiliensi pangan (a) definisi <i>strict</i> dan (b) definisi <i>broad</i>	43

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 LAMPIRAN

1	Daftar ketersediaan modul pada setiap periode	51
2	Peubah yang digunakan dalam pemodelan	51
3	<i>Hyperparameter</i> optimal	52
4	Daftar peringkat peubah <i>Strict</i> dan <i>Broad</i> hasil agregasi	53