

## ALGORITMA SHAPLEY ADDITIVE EXPLANATIONS UNTUK ANALISIS KEJADIAN RAWAN PANGAN RUMAH TANGGA DI KALIMANTAN BARAT

<sup>1\*</sup>Muhammad Usman, <sup>2</sup>Andri Hidayat

<sup>1</sup>Politeknik Negeri Sambas, Jl. Raya Sejangkung Sambas

<sup>1</sup>[usman.mtaib@gmail.com](mailto:usman.mtaib@gmail.com),

<sup>2</sup>Politeknik Negeri Sambas, Jl. Raya Sejangkung Sambas

[andribise@gmail.com](mailto:andribise@gmail.com)

Email Koresponding: [usman.mtaib@gmail.com](mailto:usman.mtaib@gmail.com)

### ABSTRAK

Kerawanan pangan merujuk pada situasi dimana individu atau kelompok tidak memiliki akses yang cukup terhadap makanan bergizi dalam periode tertentu di suatu wilayah. Masalah ini merupakan tantangan signifikan baik di skala global maupun nasional, termasuk di Indonesia. Penanganan kerawanan pangan memerlukan penerapan berbagai strategi, termasuk pengembangan model klasifikasi untuk insiden kerawanan pangan di tingkat rumah tangga atau individu. Penelitian ini bertujuan untuk menjelaskan model *Support Vector Machine (SVM) Nonlinier* dengan pendekatan *Shapley Additive Explanations* dalam memodelkan kerawanan pangan di tingkat rumah tangga, serta untuk mengidentifikasi kontribusi masing-masing variabel prediktor dari data Riset Sosial Ekonomi Nasional (SUSENAS) tahun 2023, yang diperoleh dari Badan Pusat Statistik, terhadap kejadian kerawanan pangan di Provinsi Kalimantan Barat. Temuan penelitian menunjukkan bahwa variabel-variabel signifikan yang memengaruhi kerawanan pangan mencakup luas lantai rumah, tingkat pendidikan kepala keluarga, jumlah tabungan, ketersediaan air bersih, akses internet, sumber air minum, dan jenis lantai. Hasil evaluasi terhadap model yang diterapkan pada data kejadian rawan pangan di Provinsi Kalimantan Barat menunjukkan bahwa model *Support Vector Machine (SVM)* memiliki kinerja yang cukup baik dalam klasifikasi rumah tangga yang mengalami kerawanan pangan. Parameter optimal dicapai melalui penggunaan kernel *Radially Symmetric Basis Function (RBF)*, dan algoritma SHAP dengan metode interpretasi *Permutation Explainer* dapat secara efektif menjelaskan model SVM *nonlinier* yang digunakan.

**Kata Kunci:** Kerawanan Pangan Rumah Tangga, *Shapley Additive Explanations*, *Support Vector Machine*, Susenas

### 1. PENDAHULUAN

*Machine learning* adalah subbidang dalam kecerdasan buatan yang mencakup berbagai metode yang memungkinkan komputer untuk belajar menjalankan tugas secara otomatis tanpa memerlukan pemrograman yang eksplisit (Sarker, 2021). Algoritma *machine learning* memiliki efektivitas dalam mengidentifikasi pola-pola tersembunyi dalam data. Meskipun mayoritas model yang dikembangkan oleh ilmuwan data menekankan pada tingkat akurasi prediksi sebagai indikator keberhasilan, seringkali mereka kesulitan memberikan penjelasan yang memadai mengenai proses yang mendasari pembuatan prediksi tersebut (Futoma et al., 2015; Basu Roy et al., 2015), terutama pada model *machine learning* yang lebih kompleks. Banyak algoritma *Machine learning* yang cenderung menjadi *black box* dengan komponen yang kompleks dan sulit untuk dijelaskan dan dipahami. Meski pentingnya fitur dapat dievaluasi langsung dari algoritma yang digunakan, skor yang diperoleh seringkali hanya berlaku secara lokal dan tidak konsisten antar pengamatan yang berbeda (Lundberg dan Lee 2017). Lundberg menyarankan penggunaan metode *Shapley Additive Explanations (SHAP)* untuk menginterpretasikan *black box* menjadi lebih mudah dipahami dan mudah dijelaskan.

Metode *Shapley Additive Explanations (SHAP)* memungkinkan pemanfaatan nilai *Shapley* secara luas untuk tujuan interpretasi model secara keseluruhan. SHAP memberikan informasi tentang tingkat kepentingan variabel yang mendukung pemahaman model *black box*, sehingga sangat bermanfaat dalam menangani isu kerawanan pangan. Pemilihan fitur adalah

elemen krusial dalam *machine learning* yang memastikan hanya variabel yang relevan dan terkait dengan variabel target yang dimasukkan ke dalam model, untuk menjaga tingkat akurasi (Sivasankar & Vijaya, 2019). Penggunaan variabel yang tidak relevan dapat menyebabkan penurunan akurasi model. Sebaliknya, pemilihan variabel yang tepat dapat meningkatkan kualitas prediksi (Heinze et al., 2018). Dengan mengurangi jumlah variabel, diharapkan model dapat mencapai akurasi yang lebih tinggi dan mengidentifikasi variabel yang paling signifikan, sehingga menghasilkan prediksi yang lebih akurat untuk masalah yang dihadapi (Nilsson et al., 2007).

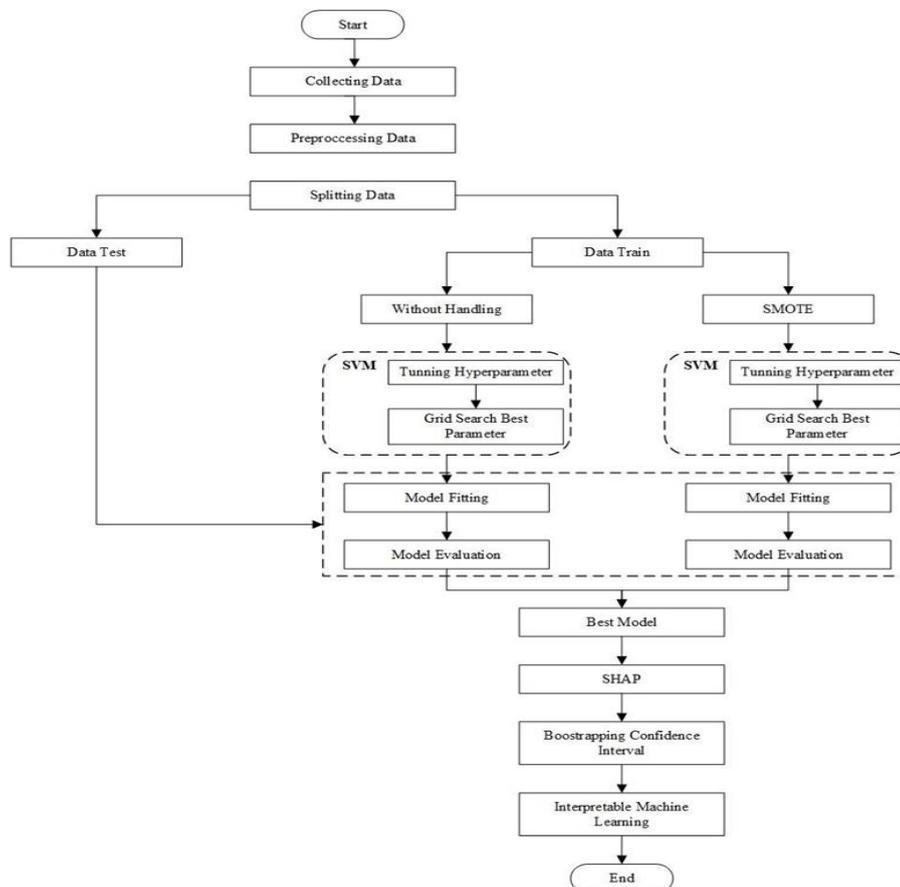
Kerawanan pangan adalah situasi di mana individu atau kelompok tidak dapat memperoleh makanan yang memadai dan sehat pada suatu waktu di wilayah tertentu (BKP, 2013). Kerawanan pangan merupakan isu kritis global dan nasional, termasuk di Indonesia. Ini bertentangan dengan tujuan *Sustainable Development Goals* (SDGs) yang berupaya mengeliminasi kelaparan, meningkatkan keamanan pangan dan nutrisi, serta memajukan pertanian yang berkelanjutan, dengan target pencapaian pada tahun 2030. Indonesia menempati posisi ke-65 dari 113 negara secara global, ke-12 di kawasan Asia Pasifik, dan ke-4 di Asia Tenggara dalam hal skor ketahanan pangan (*Police Brief FAO*, 2006).

Untuk mengatasi isu kerawanan pangan, diperlukan berbagai strategi, salah satunya adalah pengembangan model klasifikasi untuk insiden kerawanan pangan di tingkat rumah tangga atau individu. Dalam hal ini, memiliki model klasifikasi yang tidak hanya akurat tetapi juga disertai penjelasan yang komprehensif mengenai cara kerjanya sangatlah krusial. Penelitian ini menerapkan metode *Shapley Additive Explanations* (SHAP) dalam klasifikasi insiden kerawanan pangan di tingkat rumah tangga, dengan memanfaatkan data dari Survei Sosial Ekonomi Nasional (KOR SUSENAS) Provinsi Kalimantan Barat tahun 2023.

## 2. METODE

### 2.1 Tahapan Penelitian

Adapun tahapan penelitian dapat di lihat pada gambar 1.



Gambar 1. Tahapan Penelitian

### 2.2 Pengumpulan data

Penelitian ini menggunakan data sekunder dari Survei Sosial Ekonomi Nasional (KOR SUSENAS) Provinsi Kalimantan Barat pada tahun 2023. SUSENAS adalah survei yang dilakukan oleh Badan Pusat Statistik (BPS) Indonesia untuk mengumpulkan data sosial dan ekonomi dari rumah tangga di seluruh Indonesia. Data sekunder dari SUSENAS biasanya mencakup informasi tentang Demografi, Pendidikan, Kesehatan, Ketenagakerjaan, Konsumsi dan Pengeluaran, Perumahan, dan Akses terhadap Layanan Publik. Variabel prediktor yang digunakan dalam studi ini adalah indikator kerawanan pangan, dengan variabel respons yang berkaitan dengan status kejadian rawan pangan. Indikator tersebut mencakup peristiwa yang disebabkan oleh kerawanan pangan, serta kejadian yang berfungsi sebagai faktor penyebab atau dampak dari kerawanan pangan itu sendiri.

### 2.3 Pemrosesan data

Data dianggap tidak optimal jika menghadapi isu-isu seperti keberadaan data *outlier*, kehilangan nilai, serta ketidakselarasan, terutama dalam konteks data berukuran besar yang berasal dari sumber yang beraneka ragam. Kualitas data berperan krusial dalam menentukan mutu model dan analisis yang dihasilkan. Memastikan kualitas data yang tinggi adalah langkah awal yang penting dalam setiap proyek analisis data atau pengembangan model, karena data yang berkualitas rendah dapat mengarah pada hasil yang tidak akurat dan keputusan yang salah. Dengan demikian, tahap pra-pemrosesan data menjadi sangat penting untuk meningkatkan integritas data sebelum dilakukan analisis lebih lanjut, yang pada gilirannya akan menguatkan akurasi dan efisiensi model yang dihasilkan. Normalisasi merupakan salah satu teknik pra-pemrosesan yang lazim diterapkan. Secara keseluruhan, dataset penelitian terdiri dari 24 variabel dan 1 variabel terkait kerawanan pangan.

### 2.4 Deskripsi data

Dalam proses ini, langkah yang diambil meliputi penyusunan deskripsi untuk data numerik dan kategorik. Dengan mendeskripsikan data numerik dan kategorik secara tepat, dapat diperoleh wawasan awal yang penting tentang karakteristik dan distribusi data, yang akan membantu dalam analisis lebih lanjut dan pengambilan keputusan. Selanjutnya, dilakukan eksplorasi data guna memperoleh pemahaman yang lebih mendalam tentang variabel-variabel yang akan dianalisis.

### 2.5 Imbalanced Data dan Splitting Data

Tahap pengolahan data yang tidak seimbang bertujuan untuk menyeimbangkan jumlah instansi dalam berbagai kelas data melalui teknik *oversampling* dan *undersampling*. *Oversampling* dan *undersampling* merupakan teknik yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset, terutama dalam konteks pembelajaran mesin. Ketidakseimbangan kelas terjadi ketika satu kelas memiliki jauh lebih banyak contoh daripada kelas lainnya, yang dapat menyebabkan model bias terhadap kelas mayoritas. Selanjutnya, pada tahap pemisahan data, variabel penjelas akan dibagi menjadi dua set, yaitu data pelatihan dan data pengujian. Data pelatihan itu sendiri kemudian akan dipisahkan lebih lanjut menjadi dua subset, yakni *subtraining* dan validasi.

### 2.6 Membangun Model

Model *Shapley Additive Explanations* (SHAP) dibangun dengan memanfaatkan data *subtraining* dan validasi, melalui proses pencarian nilai optimal untuk parameter  $C$  dan  $\gamma$ . Prosedur penyesuaian parameter dilakukan menggunakan metode *Grid Search Cross Validation*. *Grid Search Cross Validation* merupakan teknik yang digunakan dalam pemodelan *machine learning* untuk menemukan kombinasi terbaik dari hyperparameter model. Metode ini membantu dalam meningkatkan kinerja model dengan melakukan pencarian sistematis di seluruh ruang *hyperparameter*. Setelah menentukan parameter yang paling optimal, model dilatih dengan menerapkan teknik *Cross Validation*. Data yang digunakan untuk pembangunan model berasal dari *subtraining* dan validasi, dengan pembagian data dilakukan sebanyak 10 kali.

### 2.7 Pengujian Model

*Confusion matrix* merupakan alat analisis yang digunakan dalam proses evaluasi untuk menentukan akurasi prediksi objek. *confusion matrix* sangat berguna dalam evaluasi model

klasifikasi. Ini memberikan gambaran yang jelas tentang kinerja model dengan menunjukkan bagaimana prediksi model dibandingkan dengan label sebenarnya. Dalam matriks ini, kategori yang diprediksi diletakkan di bagian atas, sedangkan kategori yang telah diamati ditempatkan di sisi kiri.

**Tabel 1. Confusion Matrix**

Nilai	Nilai Aktual	
Prediksi	TP	TN
	FP	FN

TP merujuk pada true positif yang secara akurat diklasifikasikan sebagai positif. TN mengacu pada true negatif yang dengan tepat dikategorikan sebagai negatif. FP didefinisikan sebagai false positif yang keliru diklasifikasikan sebagai positif. FN adalah false negatif yang salah diidentifikasi sebagai negatif.

## 2.8 Melakukan Analisa dan Interpretasi Hasil SHAP

Analisis dan interpretasi hasil SHAP memberikan wawasan yang mendalam tentang bagaimana fitur mempengaruhi prediksi model. Dengan menggunakan visualisasi dan analisis yang tepat, dapat meningkatkan pemahaman tentang model, menjelaskan prediksi kepada pemangku kepentingan, dan melakukan perbaikan yang diperlukan untuk meningkatkan kinerja model. Tahapan ini dilakukan untuk menganalisis dan menginterpretasikan model yang paling optimal berdasarkan berbagai tahap proses sebelumnya, sehingga dapat diidentifikasi variabel-variabel yang berkontribusi terhadap kerawanan pangan.

## 3. HASIL DAN PEMBAHASAN

Sebelum melakukan pemodelan klasifikasi dengan menggunakan *Support Vector Machine* (SVM), langkah awal yang perlu diambil adalah membagi data menjadi dua kelompok, yaitu data latih dan data uji. Data latih berfungsi untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi kinerja model yang dihasilkan. Proses pembagian data dilakukan dengan proporsi 70% dialokasikan untuk data latih dan 30% untuk data uji.

**Tabel 2. Proporsi awal data latih dan data uji**

Kategori	Provinsi Kalimantan Barat	Data Uji
	Data latih	
Proporsi	70%	30%
Jumlah	17.300	7.412
Rawan	2.800	1.258
Tidak rawan	14.450	6.116

Data pelatihan yang diperoleh kemudian dimanfaatkan untuk membangun model pembelajaran mesin *Support Vector Machine* (SVM). Penyeimbangan data dilakukan dengan menerapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) khusus untuk kelas minoritas, yaitu kelompok rumah tangga yang mengalami kerawanan pangan. Modifikasi dilakukan dengan memanfaatkan informasi dari lima tetangga terdekat ( $k=5$ ), dan pengambilan sampel dilaksanakan dengan strategi pengambilan sampel 1 (satu) agar jumlah kelas minoritas setara dengan kelas mayoritas pada data pelatihan yang baru. Jumlah data latih yang sudah diseimbangkan dengan teknik SMOTE dapat dilihat pada Tabel 3.

**Tabel 3. Jumlah data latih awal dan data latih yang sudah diseimbangkan menggunakan teknik SMOTE**

Kategori	Data latih		Data SMOTE	
	Observasi	%	Observasi	%
Rawan	2.800	16,35	14.469	50,00
Tidak rawan	14.450	83,65	14.469	50,00

Proses pencarian hiperparameter dilakukan melalui pengujian menyeluruh terhadap semua nilai yang telah diuji. Model mesin pembelajaran *Support Vector Machine* (SVM)

diterapkan pada data tanpa teknik SMOTE dan pada data yang telah diseimbangkan menggunakan metode SMOTE. Hiperparameter terbaik pada masing-masing model dapat dilihat pada Tabel 4.

**Tabel 4. Hiperparameter terbaik dalam pembentukan SVM**

Model	Hiperparameter	Data SMOTE	TanpaData SMOTE
SVM	Gamma	0,04	0,04
	C	200	500
	Kernel	rbf	rbf

Perhitungan nilai SHAP dilaksanakan dengan memanfaatkan fungsi Permutation Explainer yang bertujuan untuk memberikan penjelasan terhadap model *Support Vector Machine* (SVM). Variabel yang memiliki kontribusi tertinggi menunjukkan kemampuan yang lebih baik dalam membedakan prediksi secara akurat. Penafsiran model SVM *nonlinier* menggunakan SHAP menghasilkan urutan variabel penting yang menjadi indikator kejadian rawan pangan di Provinsi Kalimantan Barat, yaitu luas lantai yang kecil, pendidikan terakhir kepala rumah tangga yang rendah (tidak tamat SD), rumah tangga yang tidak memiliki penabung, sumber air minum yang berasal dari sumur bor dan air tadah hujan, serta tidak adanya akses internet.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan sebagai berikut:

1. Hasil evaluasi terhadap model yang diterapkan pada data kejadian rawan pangan di Provinsi Kalimantan Barat menunjukkan bahwa model *Support Vector Machine* (SVM) memiliki kinerja yang cukup baik dalam klasifikasi rumah tangga yang mengalami kerawanan pangan. Parameter optimal dicapai melalui penggunaan kernel *radially symmetric basis function* (RBF). Selain itu, algoritma SHAP dengan metode interpretasi *Permutation Explainer* dapat secara efektif menjelaskan model SVM nonlinier yang digunakan.
2. Analisis model SVM *nonlinier* yang diterapkan dengan metode SHAP menghasilkan peringkat variabel yang dianggap signifikan dalam mengidentifikasi kejadian rawan pangan di Provinsi Kalimantan Barat. Variabel-variabel penting tersebut meliputi ukuran lantai yang kecil, tingkat pendidikan terakhir kepala rumah tangga yang rendah, yakni tidak menyelesaikan pendidikan dasar, serta kondisi rumah tangga yang tidak memiliki penabung sumber air minum yang bersumber dari sumur bor dan air hujan, serta keterbatasan akses terhadap internet.

#### REFERENSI

- Badan Pusat Statistik. 2016. Potret Awal Tujuan Pembangunan Berkelanjutan (Sustainable Development Goals) di Indonesia. Badan Pusat Statistik. ISBN: 978-602-438-071-7
- Basu Roy S, Teredesai A, Zolfaghar K, Liu R, Hazel D, Newman S, et al. 2015. Dynamic hierarchical classification for patient risk-of-readmission. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining: ACM. pp. 1691–700.
- Futoma J, Morris J, dan Lucas J. 2015. A comparison of models for predicting early lundberghospital readmissions. *J Biomed Inform.* 56:229–38.
- Heinze G, Wallisch C, Dunkler D. 2018. Variable Selection - A Review and Recommendations for the Practicing Statistician. *Biometrical Journal.* 60(3):431-449. doi:10.1002/bimj.201700067.
- Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I. 2020. From Local Explanations to Global Understanding with Explainable AI For Trees. *Nature Machine Intelligence.* 2(1):56-67. doi:10.1038/s42256-019-0138-9.
- Lundberg SM, Lee S-I. 2017. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems.*

- Nilsson R, M. Peña J, Björkegren J, Tegnér J. 2007. Consistent Feature Selection for Pattern Recognition in Polynomial Time. *The Journal of Machine Learning Research*. 8:589–612.
- Sarker IH. 2021. Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Comput Sci*. Vol. 1: 1-20
- Shapley LS. 2016. A Value for n-Person Games. In: *Contributions to the Theory of Games (AM-28)*, Volume II.
- Vijayakumar, S. dan Wu, S. 1999. Sequential Support Vector Classifier and Regression. *Prociding International Conference on Soft Computing (SOCO'99)*. Hal. 610-619.