

Analisis Faktor Dominan Terhadap PHBS Rumah Tangga Menggunakan SVM dan SHAP Wilayah Puskesmas Tepusen, Temanggung

Drajad Bima Ajipangestu¹, Novia Dewi Putri Ayuningtyas², Maesaroh³, Arum Helmi Manggala Putri⁴, Subinarto⁵, Ayu Fadhila⁶

^{1,3}Bisnis Digital, Institut Teknologi dan Bisnis Muhammadiyah Grobogan, Grobogan, Indonesia

²Promosi Kesehatan, Puskesmas Tepusen, Kaloran, Temanggung, Indonesia

⁴Sains Data, Institut Teknologi dan Bisnis Muhammadiyah Grobogan, Grobogan, Indonesia

⁵Rekam Medis dan Informasi Kesehatan, Akademi Kesehatan Muhammadiyah Temanggung, Temanggung, Indonesia

⁶Magister Kesehatan Masyarakat, Universitas Negeri Semarang, Semarang, Indonesia

Email: ¹bimadrajad@gmail.com, ²noviadewipa@gmail.com, ³maesaroh@itbmg.ac.id, ⁴arum.manggala@gmail.com,

⁵subinarto@akmt.ac.id, ⁶ayufadhilaaa@students.unnes.ac.id

Email Penulis Korespondensi: ¹bimadrajad@gmail.com

Abstrak- Perilaku Hidup Bersih dan Sehat (PHBS) merupakan indikator penting dalam penilaian status kesehatan masyarakat. Penelitian ini bertujuan untuk mengklasifikasikan tingkat PHBS rumah tangga serta mengidentifikasi faktor dominan yang memengaruhinya di wilayah kerja Puskesmas Tepusen, Kecamatan Kaloran, Kabupaten Temanggung. Data diperoleh dari survei rumah tangga tahun 2024 sebanyak 4.425 responden, dengan pengisian kuesioner yang dilakukan oleh kader dan tenaga *surveilans* terlatih. Pengukuran mengacu pada 16 indikator PHBS terstandarisasi, yang telah divalidasi oleh Dinas Kesehatan, mencakup: persalinan oleh tenaga kesehatan, pemeriksaan kehamilan, pemberian ASI eksklusif, penimbangan balita, konsumsi gizi seimbang, penggunaan air bersih, kepemilikan jamban sehat, pengelolaan sampah, tidak merokok di dalam rumah, aktivitas fisik, mencuci tangan pakai sabun, menyikat gigi secara teratur, tidak mengonsumsi alkohol/narkoba, kepesertaan jaminan kesehatan, pemberantasan sarang nyamuk, serta konsumsi tablet tambah darah (TTD). Metode *Support Vector Machine* (SVM) digunakan untuk klasifikasi multikelas (Sehat Pratama, Madya, Utama, dan Paripurna), dengan skema pembagian data 80:20 dan validasi silang *5-fold* untuk menghindari *overfitting*. Kernel *linear* menunjukkan akurasi tertinggi, sebesar 97,74%. Untuk mengevaluasi kontribusi setiap fitur terhadap hasil prediksi, digunakan metode interpretabilitas *Shapley Additive Explanations* (SHAP). Indikator “TIDAK_MEROKOK” memiliki nilai SHAP tertinggi (0.118313), diikuti “CEK_KESEHATAN” (0.037) dan “MINUM_TTD” (0.030). Hasil ini menunjukkan pentingnya intervensi berbasis data untuk pengendalian rokok, peningkatan layanan kesehatan preventif, dan pemenuhan gizi perempuan dalam meningkatkan status PHBS secara operasional.

Kata Kunci : PHBS, SVM, SHAP, *Machine Learning*, Kesehatan Masyarakat

Abstract- Clean and Healthy Living Behavior (PHBS) is a crucial indicator for assessing public health status. This study aims to classify household PHBS levels and identify the dominant factors influencing them within the Tepusen Public Health Center (*Puskesmas*) area, Kaloran District, Temanggung Regency. Data were collected from a 2024 household survey involving 4,425 respondents, with questionnaire administration conducted by trained community health workers and surveillance officers. Measurement was based on 16 standardized PHBS indicators validated by the Health Office, including: delivery by health personnel, antenatal care, exclusive breastfeeding, routine child weighing, balanced nutrition, clean water usage, healthy latrine ownership, waste management, no smoking inside the house, regular physical activity, handwashing with soap, regular tooth brushing, no alcohol or drug use, participation in health insurance, mosquito nest eradication, and iron tablet consumption. The Support Vector Machine (SVM) method was employed for multiclass classification (*Sehat Pratama*, *Sehat Madya*, *Sehat Utama*, and *Sehat Paripurna*), using an 80:20 training-test data split and 5-fold cross-validation to prevent overfitting. The linear kernel achieved the highest accuracy of 97.74%. To assess the contribution of each feature to prediction outcomes, the Shapley Additive Explanations (SHAP) model interpretation technique was applied. “TIDAK_MEROKOK” indicator yielded the highest SHAP value (0.118313), followed by “CEK_KESEHATAN” (0.037) and “MINUM_TTD” (0.030). These findings suggest that smoking behavior significantly impacts PHBS classification. The results support data-driven policy recommendations, such as implementing smoking cessation programs, improving preventive health service access, and ensuring targeted distribution of iron supplements to improve PHBS levels effectively at the community level.

Keywords: PHBS, SVM, SHAP, Machine Learning, Public Health

1. PENDAHULUAN

Perilaku Hidup Bersih dan Sehat (PHBS) merupakan indikator penting dalam upaya peningkatan kualitas kesehatan masyarakat. Kementerian Kesehatan Republik Indonesia telah menetapkan sejumlah indikator PHBS yang perlu diterapkan pada level rumah tangga, sebagai bagian dari strategi promotif dan preventif dalam sistem pelayanan kesehatan nasional. Indikator-indikator tersebut mencakup aspek-aspek dasar dalam kehidupan sehari-hari, seperti tidak merokok di dalam rumah, mencuci tangan dengan sabun, menggunakan air bersih, memiliki jamban sehat, mengikuti program imunisasi, serta melakukan pemeriksaan kesehatan secara rutin. Penerapan PHBS yang konsisten di tingkat rumah tangga berperan besar dalam mencegah penyakit menular maupun tidak menular, serta meningkatkan produktivitas dan kesejahteraan masyarakat secara umum[1]. Namun, implementasi PHBS di lapangan sering kali tidak merata, terutama di wilayah pedesaan seperti yang terdapat di wilayah kerja Puskesmas Tepusen, Kecamatan Kaloran, Kabupaten Temanggung. Tantangan-tantangan seperti keterbatasan akses informasi kesehatan, rendahnya tingkat pendidikan, budaya lokal, serta keterbatasan infrastruktur sanitasi menjadi faktor penghambat utama. Meskipun terdapat program penyuluhan dan intervensi dari puskesmas serta Dinas Kesehatan, masih ditemukan rumah tangga yang belum memenuhi sebagian besar indikator PHBS. Hal ini mencerminkan adanya kebutuhan untuk pendekatan yang lebih kontekstual dan berbasis data dalam merancang strategi intervensi kesehatan masyarakat.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi tingkat Perilaku Hidup Bersih dan Sehat (PHBS) rumah tangga berdasarkan data survei di wilayah kerja Puskesmas Tepusen dengan menggunakan algoritma *Support Vector Machine* (SVM), serta mengidentifikasi indikator PHBS yang memiliki kontribusi paling signifikan terhadap hasil klasifikasi berdasarkan metode interpretabilitas model *Shapley Additive Explanations* (SHAP). Secara khusus, penelitian ini ingin menjawab dua pertanyaan utama: (1) sejauh mana model SVM dapat mengklasifikasikan tingkat PHBS rumah tangga secara akurat dibandingkan pendekatan klasifikasi statis berbasis jumlah indikator yang terpenuhi; dan (2) indikator mana saja yang paling dominan dalam memengaruhi klasifikasi tingkat PHBS menurut hasil pembobotan SHAP. Hipotesis awal yang diangkat adalah bahwa model SVM dengan kernel yang sesuai mampu menghasilkan akurasi klasifikasi di atas 90%, serta terdapat sejumlah kecil indikator kunci yang memiliki kontribusi lebih besar dibandingkan indikator lainnya dalam memengaruhi status PHBS rumah tangga. *Support Vector Machine* (SVM) merupakan salah satu algoritma klasifikasi yang terbukti efektif dalam menangani masalah klasifikasi dengan dimensi tinggi dan data yang kompleks[2]. Dalam penelitian ini, digunakan pendekatan klasifikasi multikelas dengan tiga jenis kernel: *linear*, *radial basis function* (RBF), dan *polynomial*[3].

Pendekatan ini memungkinkan pemodelan hubungan non-linier antara indikator PHBS dengan kategori tingkat PHBS rumah tangga, yang mungkin tidak terdeteksi oleh metode statistik konvensional. Beberapa studi sebelumnya telah memanfaatkan *machine learning* dalam konteks prediksi status kesehatan masyarakat. Lebih spesifik ke perilaku hidup sehat, *Data & Knowledge Engineering* (juli 2023) memaparkan penggunaan kombinasi SVM, ANN, *Random Forest*, dan SHAP untuk mengevaluasi faktor-faktor yang memengaruhi angka kematian ibu di India secara eksplanatif dan berbasis kebijakan kesehatan presisi[4]. Pendekatan ini relevan dengan konteks studi PHBS rumah tangga, karena melibatkan penggunaan model klasifikasi non-linier dan interpretasi fitur yang transparan untuk mendukung intervensi kesehatan berbasis data.

Untuk meningkatkan interpretabilitas model, metode *Shapley Additive Explanations* (SHAP) digunakan sebagai alat bantu analisis. SHAP merupakan pendekatan berbasis teori permainan yang memberikan penjelasan kontribusi setiap fitur terhadap hasil prediksi model[5]. Melalui analisis SHAP, peneliti dapat mengetahui indikator mana yang paling signifikan dalam menentukan klasifikasi PHBS, baik secara agregat (*global feature importance*) maupun pada level individu (*local explanations*). Dalam konteks kesehatan masyarakat, interpretabilitas ini penting untuk mendukung transparansi dan akuntabilitas dalam pengambilan keputusan berbasis data.

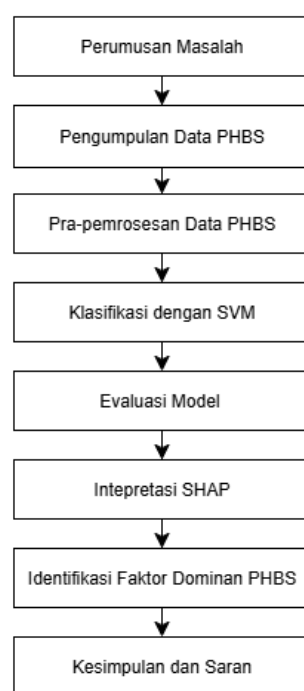
Meskipun sistem klasifikasi PHBS rumah tangga secara konvensional menggunakan pendekatan rule-based dengan menghitung jumlah indikator yang terpenuhi, penerapan metode SVM memiliki signifikansi tersendiri[6]. Pertama, pendekatan *machine learning* mampu mengenali pola-pola kompleks yang mungkin tidak tertangkap oleh sistem klasifikasi berbasis jumlah semata. SVM dapat mempelajari kombinasi fitur-fitur tertentu yang secara statistik lebih berpengaruh dalam menentukan kelas PHBS, bahkan ketika jumlah indikator yang terpenuhi sama. Hal ini membuka kemungkinan untuk menemukan pola baru yang sebelumnya tidak diperhitungkan dalam kerangka kebijakan konvensional[7].

Integrasi metode SHAP dalam penelitian ini memberikan nilai tambah dalam hal interpretasi hasil model. SHAP memungkinkan identifikasi fitur-fitur mana yang paling berkontribusi terhadap klasifikasi setiap rumah tangga[8]. Hal ini dapat digunakan oleh pemangku kebijakan untuk memprioritaskan indikator-indikator tertentu dalam intervensi program kesehatan, berdasarkan pengaruh nyatanya terhadap status PHBS di lapangan. Misalnya, jika indikator "tidak merokok di dalam rumah" menunjukkan kontribusi paling tinggi terhadap peningkatan status PHBS, maka edukasi dan intervensi yang difokuskan pada perilaku merokok dapat diutamakan[9].

Hasil dari model klasifikasi ini berpotensi untuk digunakan dalam otomatisasi proses penilaian PHBS di masa mendatang, dan bisa ditampilkan dalam bentuk media informasi online [10]. Jika data dapat dikumpulkan secara digital melalui sistem informasi kesehatan masyarakat atau survei berbasis aplikasi, maka klasifikasi PHBS dapat dilakukan secara *real-time* dan dengan akurasi yang lebih tinggi. Selain itu, pendekatan ini juga dapat digunakan untuk validasi terhadap sistem klasifikasi konvensional yang bersifat manual dan rawan bias, serta syarat sentimen[11]. Dengan demikian, penelitian ini tidak hanya memiliki kontribusi akademik dalam hal penerapan metode *machine learning* dan interpretasi model, tetapi juga memberikan nilai praktis bagi penguatan sistem informasi dan kebijakan promosi kesehatan. Dengan pendekatan berbasis data dan teknologi, hasil penelitian ini diharapkan mampu memberikan rekomendasi yang lebih tepat sasaran bagi program-program promosi kesehatan, khususnya di wilayah-wilayah dengan keterbatasan sumber daya. Implementasi hasil ini dapat diarahkan pada upaya peningkatan PHBS secara berkelanjutan, dengan fokus pada indikator-indikator yang memiliki dampak paling besar terhadap peningkatan kualitas kesehatan masyarakat[12]. Penelitian ini menjadi langkah awal dalam menjembatani antara teknologi *data science* dan praktik kesehatan masyarakat yang berbasis komunitas.

2. METODOLOGI PENELITIAN

Desain penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data berbasis *machine learning*. Tujuan utama adalah mengklasifikasikan tingkat Perilaku Hidup Bersih dan Sehat (PHBS) rumah tangga serta mengidentifikasi faktor-faktor dominan yang memengaruhinya. Proses penelitian terdiri dari beberapa tahapan yang digambarkan pada Gambar 1 berikut:



Gambar 1. Alur Tahapan Penelitian

2.1 Perumusan Masalah

Tahap awal berupa identifikasi kesehatan masyarakat terkait PHBS di wilayah Puskesmas Tepusen, Kaloran, Temanggung, kebutuhan analisis data terkait hasil data PHBS pada tahun 2024, serta keinginan untuk meningkatkan persentase hasil dari PHBS Utama ke PHBS Paripurna di tahun selanjutnya. Permasalahan diformulasikan untuk analisis berbasis komputasi data dan *machine learning* menggunakan metode klasifikasi dan interpretasi data.

2.2 Pengumpulan Data PHBS

Data primer dikumpulkan pada tahun 2024 melalui survei rumah tangga menggunakan kuesioner terstruktur yang mencakup 16 indikator PHBS sesuai standar dari Kementerian Kesehatan. Survei dilaksanakan di wilayah

kerja Puskesmas Tepusen, yang terdiri atas enam desa: Keblukan, Kemiri, Tepusen, Tegowanuh, Tempuran, dan Gandulan. Sebanyak 4.425 rumah tangga berhasil didata secara menyeluruh menggunakan metode *total sampling*, di mana seluruh rumah tangga yang menjadi sasaran pendataan diwawancarai langsung oleh kader kesehatan dan tim surveilans Puskesmas. Pendekatan ini dipilih karena cakupan wilayah yang terjangkau serta untuk memastikan representasi penuh populasi target.

Setiap rumah tangga diklasifikasikan ke dalam empat kategori tingkat PHBS berdasarkan jumlah indikator yang dipenuhi, yakni Sehat Pratama (1–5 indikator), Sehat Madya (6–10), Sehat Utama (11–15), dan Sehat Paripurna (16). Data rekapitulasi awal dan distribusi kategori dapat dilihat pada Gambar 2.

NO	Desa	DESA DI DATA	RUMAH TANGGA YG ADA (RK)	RUMAH TANGGA DI DATA	CAKUPAN KUMULATIF PER INDIKATOR																CAPAIAN STRATA			
					PERSALINAN NAKES	PERUSAHAAN KEMAMLAHAN (JG)	ASI EKSKLUSIF	IM BANG BALITA	GIZI SEM BANG	AIR BERSIH	JAM BAK SEHAT	SAM PAH	MINUM TTD	AKTIFITAS FISIK	TIDAK MEROKOK	CUCI TANGAN	GOSOK GIGI	GARAM YODIUM	CBK KESEHATAN	PSN	SEHAT PRATAMA	SEHAT MADYA	SEHAT UTAMA	SEHAT PARIPURNA
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16				
1	keblukan	1	412	428	428	428	425	428	428	427	427	427	299	420	90	428	427	428	359	427	-	-	383	45
2	kemiri	1	802	752	725	724	698	718	750	751	740	724	480	721	108	752	752	752	617	724	1	18	684	49
3	tepusen	1	769	762	746	749	724	747	759	761	761	762	445	580	136	699	755	732	487	759	-	4	742	16
4	tegowanuh	1	750	699	662	684	646	678	699	697	693	693	553	690	97	698	699	697	431	695	-	4	657	37
5	tempuran	1	1026	1.019	1.018	1.018	1.014	1.018	1.016	998	975	90	830	1.011	184	1.009	1.019	1.018	734	1.016	-	1	1.016	2
6	gandulan	1	760	766	761	755	710	754	766	766	763	765	656	739	146	766	766	766	513	757	-	3	704	59

Gambar 2. Data Rekap Capaian Indikator Wilayah Kerja Puskesmas Tepusen, Kaloran, Temanggung

2.3 Pra-pemrosesan Data PHBS

Pada Tahapan ini, peneliti melakukan tahapan *preprocessing* data dengan mempersiapkan data agar bisa diolah dengan maksimal; 1. Memastikan tidak ada data kosong, 2. Pelabelan dengan *indicator* sehat pratama (1–5) dengan Kode 1, madya (6–10) dengan Kode 2, utama (11–15) dengan Kode 3, dan paripurna (16) dengan Kode 4. Total 4425 Data siap untuk diuji dan diklasifikasikan ke tahap berikutnya. Data contoh hasil tahap Praproses nosa dilihat pada gambar 3.

RT	KK_ID	ALUNAN	NSA	KEHAJI	EKSKLUBANG	BAJI	SEIMBAUR	BERSIMBAN	SEHSAMPALHINUM	TTIFITAS	FAK	MEROICI	TANGAOSOK	GIKAM	YODI	KESEHAJ	PSN	JAH	INDIKABEL	PHBS
1 01-001		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	Sehat Paripurna
1 01-002		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	Sehat Paripurna
1 01-003		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	Sehat Paripurna
1 01-004		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	Sehat Paripurna
1 01-005		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	Sehat Paripurna
1 01-006		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	Sehat Paripurna
1 01-007		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	Sehat Paripurna
1 01-008		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	15	Sehat Utama
1 01-009		1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	15	Sehat Utama
1 01-010		1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	15	Sehat Utama
1 01-011		1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	15	Sehat Utama

Gambar 3. Potongan Screenshot Data Hasil Pra-proses untuk Masuk Kedalam Machine Learning

2.4. Klasifikasi dengan SVM

Penelitian ini menggunakan metode klasifikasi berbasis *machine learning* dengan algoritma *Support Vector Machine* (SVM) untuk mengelompokkan perilaku rumah tangga berdasarkan tingkat pencapaian indikator Perilaku Hidup Bersih dan Sehat (PHBS). Data dikumpulkan dari wilayah kerja Puskesmas Tepusen, Kecamatan Kaloran, Kabupaten Temanggung.

Tahapan awal meliputi pra-pemrosesan data, yaitu normalisasi fitur menggunakan teknik *StandardScaler*, serta pengkodean label target menggunakan *LabelEncoder* untuk memastikan kompatibilitas dengan algoritma klasifikasi. Model SVM dilatih dengan menggunakan *kernel linear*, *polynomial*, dan *radial basis function (RBF)* guna membandingkan performa masing-masing pendekatan terhadap data klasifikasi multikelas (Sehat Pratama, Madya, Utama, dan Paripurna).

2.5. Evaluasi Model

Selanjutnya, evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*, baik pada pengujian langsung maupun dengan validasi silang (*5-fold cross validation*) untuk memastikan generalisasi model terhadap data baru. Untuk mengetahui kontribusi masing-masing indikator dalam memengaruhi hasil klasifikasi, digunakan metode *Shapley Additive Explanations* (SHAP). Visualisasi SHAP *summary plot* digunakan untuk mengidentifikasi indikator paling dominan yang berkontribusi terhadap prediksi tingkat PHBS rumah tangga.

2.6. Interpretasi SHAP

Untuk memberikan interpretasi terhadap hasil klasifikasi, digunakan metode *Shapley Additive Explanations* (SHAP). SHAP menjelaskan kontribusi masing-masing fitur (indikator PHBS) terhadap output klasifikasi. SHAP

menghitung nilai kontribusi setiap fitur (indikator PHBS) terhadap output model untuk masing-masing data uji. Nilai positif menunjukkan fitur tersebut mendorong ke arah kelas tertentu, sedangkan nilai negatif menunjukkan fitur tersebut mengurangi kemungkinan prediksi terhadap kelas tersebut. Hubungan relasi kedua metode bisa dilihat pada tabel 1.

Tabel 1. Peran Metode

Komponen	Peran
SVM (<i>Support Vector Machine</i>)	Model klasifikasi yang digunakan untuk menentukan kategori PHBS rumah tangga (misalnya: Sehat Paripurna, Sehat Utama, dst.) berdasarkan 16 indikator PHBS (fitur biner 1/0).
SHAP (<i>Shapley Additive Explanations</i>)	Alat interpretasi yang digunakan untuk menjelaskan kontribusi masing-masing indikator PHBS terhadap hasil klasifikasi yang dihasilkan oleh SVM.

2.7. Identifikasi Faktor Dominan PHBS

Berdasarkan nilai SHAP, diidentifikasi indikator-indikator mana yang paling berpengaruh dalam menentukan kategori PHBS rumah tangga. Informasi ini penting sebagai dasar pengambilan kebijakan intervensi kesehatan masyarakat.

2.8. Kesimpulan dan Saran

Tahap akhir merangkum temuan dari seluruh proses penelitian dan memberikan saran strategis untuk penguatan program PHBS di tingkat rumah tangga.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengidentifikasi faktor dominan dan mengklasifikasikan tingkat Perilaku Hidup Bersih dan Sehat (PHBS) rumah tangga berdasarkan 16 indikator yang dianalisis menggunakan algoritma *Support Vector Machine* (SVM) dan interpretasi melalui metode SHAP (*Shapley Additive Explanations*). Dataset terdiri dari 4.425 rumah tangga yang telah dikategorikan ke dalam empat tingkat PHBS, yaitu Sehat Pratama, Sehat Madya, Sehat Utama, dan Sehat Paripurna, namun hanya tiga kategori dominan yang masuk ke dalam data uji: Sehat Madya, Sehat Utama, dan Sehat Paripurna.

3.1 Hasil Klasifikasi PHBS dengan SVM

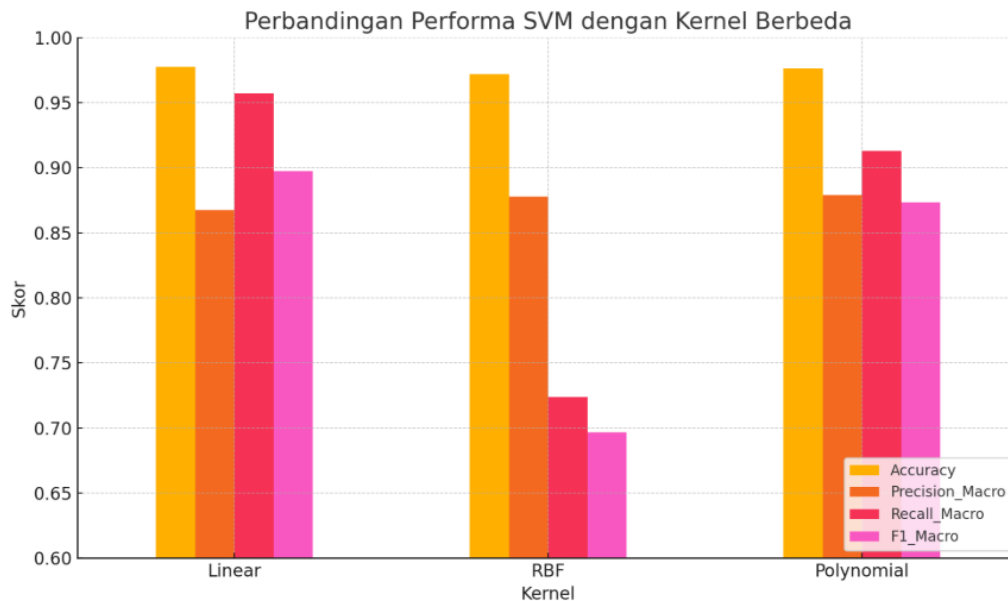
Setelah dilakukan preprocessing dan normalisasi, model *Support Vector Machine* (SVM) dilatih dan dievaluasi menggunakan dua skema validasi, yaitu split data 80:20 dan *5-fold cross-validation* untuk meningkatkan keandalan generalisasi model. Pada skema split, sebanyak 3.540 data (80%) digunakan sebagai data pelatihan dan 885 data (20%) sebagai data pengujian. Pembagian data dilakukan secara *stratified sampling*, untuk mempertahankan proporsi distribusi kelas yang tidak seimbang pada kedua subset.

Kategori Sehat Pratama tidak disertakan dalam proses pelatihan karena hanya terdapat satu data, sehingga tidak memenuhi syarat minimum untuk validasi model yang adil dan stabil. Pemodelan dilakukan dengan membandingkan tiga jenis kernel, yaitu *Linear*, *Radial Basis Function (RBF)*, dan *Polynomial*, guna menentukan performa klasifikasi yang optimal.

Hasil klasifikasi menunjukkan bahwa model SVM dengan kernel *Linear* memberikan performa terbaik dengan akurasi 97,74%, *macro precision* 0,8675, *macro recall* 0,9570, dan *macro F1-score* 0,8975. Kernel *Linear* mampu mengklasifikasikan ketiga kelas dengan baik, termasuk kelas minoritas seperti Sehat Madya. Kernel *Polynomial* menunjukkan performa tinggi (akurasi 97,63%, recall 0,9127), namun sedikit di bawah kernel *Linear* dalam hal *F1-score*. Sementara itu, kernel RBF meskipun mencetak akurasi 97,18%, memiliki *macro recall* yang lebih rendah (0,7239) dan *F1-score* 0,6967, mengindikasikan kecenderungan *overfitting* terhadap kelas mayoritas. Hasil lengkap dapat dilihat pada Tabel 2 dan Gambar 4.

Tabel 2. Hasil Laporan Klasifikasi

<i>Kernel</i>	<i>Accuracy</i>	<i>Precision (Macro)</i>	<i>Recall (Macro)</i>	<i>F1 Score (Macro)</i>
Linear	0.9740	0.8675	0.9570	0.8975
RBF	0.9718	0.8776	0.7239	0.6967
Polynomial	0.9763	0.8790	0.9127	0.8735



Gambar 4. Hasil Laporan Grafik Perbandingan SVM dengan Tiga Kernel

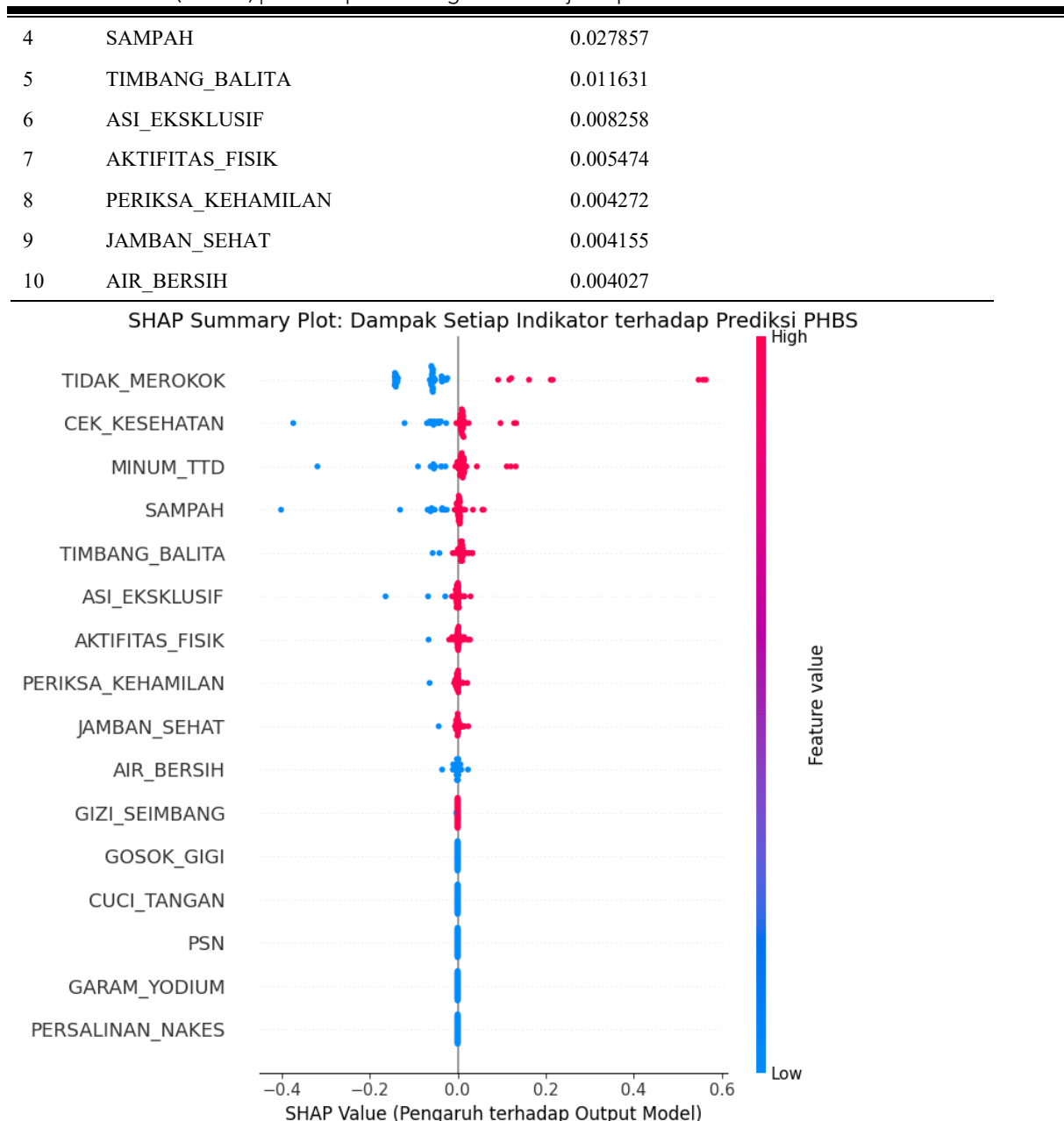
3.2 Interpretasi Model dengan SHAP

Analisis SHAP (*Shapley Additive Explanations*) dilakukan untuk menginterpretasikan model klasifikasi *Support Vector Machine* (SVM) yang telah dibangun pada data PHBS rumah tangga. SHAP memberikan nilai kontribusi setiap fitur (indikator) terhadap prediksi yang dihasilkan oleh model, dalam bentuk nilai absolut rata-rata untuk setiap fitur. Dari tabel 3, indikator "TIDAK_MEROKOK" menunjukkan nilai SHAP tertinggi (0.118313), yang berarti memiliki pengaruh paling signifikan dalam menentukan klasifikasi perilaku PHBS.

Disusul oleh indikator "CEK_KESEHATAN" (0.037) dan "MINUM_TTD" (0.030), yang juga memberikan kontribusi bermakna terhadap model. Nilai SHAP yang lebih rendah pada indikator seperti "JAMBAN_SEHAT" dan "AIR_BERSIH" dapat menunjukkan bahwa variabel tersebut sudah hampir homogen di sebagian besar data (karena cakupan tinggi), sehingga kurang berperan dalam membedakan kategori. Secara teknis, nilai-nilai SHAP ini dihitung dari hasil *KernelExplainer*, yang mendekati fungsi prediktif model dengan sampling dan memodelkan efek marginal dari masing-masing fitur. Model SVM dengan kernel *linear* digunakan dalam proses ini karena bersifat interpretable dan memungkinkan penggunaan SHAP secara langsung. Hasil ini mendukung penggunaan *explainable AI (XAI)* dalam konteks kesehatan masyarakat untuk membantu pengambil kebijakan memahami variabel-variabel yang paling berdampak dalam perilaku hidup bersih dan sehat. Hasil lengkap dapat dilihat pada tabel 3 dan gambar 5.

Tabel 3. Tabel Hasil Interpretasi Model dengan SHAP

No	Indikator	Nilai SHAP
1	TIDAK_MEROKOK	0.118313
2	CEK_KESEHATAN	0.037038
3	MINUM_TTD	0.030555



Gambar 5. Hasil Laporan dengan *Summary Plot* untuk Setiap Indikator terhadap Prediksi PHBS

Gambar 5. menunjukkan dampak masing-masing indikator terhadap output model klasifikasi PHBS berdasarkan nilai SHAP. Masing-masing titik pada grafik mewakili satu sampel data, dengan: sumbu horizontal (SHAP *Value*): menunjukkan kontribusi suatu fitur terhadap prediksi model. Nilai positif berarti mendorong prediksi ke arah kelas tertentu (misalnya "Sehat Paripurna"), sedangkan nilai negatif mengurangi kemungkinan prediksi ke kelas tersebut.

Warna titik: merepresentasikan nilai asli dari fitur tersebut untuk masing-masing data. Merah: nilai fitur tinggi (indikator dipenuhi), Biru: nilai fitur rendah (indikator tidak dipenuhi atau nol).

Visualisasi ini memperkuat pemahaman bahwa indikator TIDAK_MEROKOK adalah penentu utama klasifikasi PHBS berdasarkan model SVM. Nilai SHAP ini memberikan justifikasi kuantitatif mengapa fitur tertentu memiliki bobot penting dalam pengambilan keputusan model, dan membantu pemangku kebijakan untuk menetapkan intervensi prioritas dalam program PHBS.

3.3 Identifikasi Faktor Dominan PHBS

Berdasarkan hasil perhitungan *Shapley Additive Explanations* (SHAP), diperoleh bahwa indikator TIDAK_MEROKOK merupakan fitur paling dominan dalam mempengaruhi prediksi model terhadap kategori PHBS, dengan nilai kontribusi tertinggi sebesar 0,1183. Hal ini menunjukkan bahwa kebiasaan tidak merokok di lingkungan rumah tangga memiliki peran penting dalam pembentukan perilaku hidup bersih dan sehat.

Indikator lain seperti CEK_KESEHATAN, MINUM_TTD, dan SAMPAH juga memberikan kontribusi signifikan, yang mencerminkan pentingnya akses pelayanan kesehatan dan kebersihan lingkungan. Sementara itu, indikator seperti AIR_BERSIH dan JAMBAN_SEHAT memiliki nilai kontribusi yang relatif lebih kecil, diduga karena telah dipenuhi oleh sebagian besar rumah tangga sehingga variabilitasnya rendah dalam data.

Studi sebelumnya di *Healthcare Analytics* (2024) berjudul “*A novel machine learning approach for diagnosing diabetes with a self-explainable interface*” menunjukkan bahwa SVM dan model lainnya bersama SHAP menyoroti fitur perilaku (misalnya pemeriksaan kesehatan rutin) sebagai prediktor utama penyakit metabolik [13]. Temuan ini mengindikasikan bahwa model cenderung memberikan bobot lebih tinggi pada indikator perilaku aktif dan preventif dalam kesehatan. Dengan demikian, hasil SHAP pada studi ini memperkuat bukti bahwa pendekatan intervensi berbasis edukasi dan promosi perilaku—seperti larangan merokok, pemeriksaan kesehatan preventif, serta distribusi TTD—mempunyai potensi dampak lebih signifikan ketimbang pengembangan infrastruktur yang sudah tersebar luas. Temuan ini memberikan dasar ilmiah yang kuat untuk merancang kebijakan kesehatan masyarakat yang lebih berbasis data dan tepat sasaran.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis faktor-faktor dominan yang memengaruhi Perilaku Hidup Bersih dan Sehat (PHBS) pada tingkat rumah tangga di wilayah kerja Puskesmas Tepusen, Kecamatan Kaloran, Kabupaten Temanggung. Dengan menerapkan metode klasifikasi *Support Vector Machine* (SVM) serta pendekatan interpretabilitas model melalui *Shapley Additive Explanations* (SHAP), Model dilatih menggunakan skema 80:20 dan *5-fold cross-validation* untuk memastikan hasil yang stabil dan *generalizable*. Kategori Sehat Pratama dikecualikan dari pelatihan model karena hanya terdapat satu data, sehingga tidak memungkinkan validasi yang adil. Diperoleh model yang mampu mengklasifikasikan kategori PHBS dengan tingkat akurasi tinggi, mencapai 97,74%. Berdasarkan analisis SHAP, ditemukan bahwa indikator TIDAK_MEROKOK, CEK_KESEHATAN, dan MINUM_TTD merupakan tiga faktor paling dominan dalam memengaruhi hasil klasifikasi model. Hal ini menunjukkan bahwa kebiasaan tidak merokok di dalam rumah, kepatuhan terhadap pemeriksaan kesehatan berkala, dan konsumsi tablet tambah darah merupakan indikator yang paling konsisten membedakan antara kategori PHBS rendah dan tinggi. Indikator lainnya seperti PENGELOLAAN SAMPAH, TIMBANG_BALITA, dan ASI_EKSKLUSIF juga turut berkontribusi namun dalam tingkat yang lebih moderat.

Temuan ini menggaris bawahi pentingnya pendekatan berbasis data dalam memetakan prioritas intervensi kesehatan masyarakat. Hal ini sejalan dengan penelitian Li et al. [14], yang menekankan bahwa keberhasilan metode explainable AI (XAI) tidak cukup hanya dari segi akurasi model, tetapi juga dari kemampuan interpretasi yang sebanding dengan kedalaman logika manusia, seperti diukur dengan *depth of decision tree*. Dalam penelitian ini, model menggunakan SHAP berhasil menyoroti indikator “TIDAK_MEROKOK” sebagai fitur paling dominan dalam klasifikasi tingkat PHBS rumah tangga. Ini memperlihatkan bagaimana model klasifikasi dapat dieksplorasi lebih lanjut untuk menghasilkan pengetahuan yang bisa ditindaklanjuti, bukan hanya prediksi mentah. Keterbatasan studi ini terletak pada ketidakseimbangan distribusi kelas dan keterbatasan indikator yang mungkin belum mencakup seluruh determinan sosial PHBS. Penelitian lanjutan dapat mengintegrasikan lebih banyak variabel kontekstual atau pendekatan berbasis data spasial. Hasil penelitian ini menjadi dasar yang kuat bagi penentuan prioritas edukasi dan intervensi oleh pemangku kebijakan lokal, khususnya dalam penguatan perilaku tidak merokok dan peningkatan layanan preventif.

REFERENCES

- [1] N. Novika, D. Sayati, and N. S. Murni, “Faktor-Faktor Yang Berhubungan dengan PHBS,” *Citra Delima Sci. J. Citra Int. Inst.*, vol. 7, no. 2, 2023, doi: 10.33862/citradelima.v7i2.370.
- [2] M. A. Assiva, H. A. Santoso, and C. Supriyanto, “Metode Fastica Untuk Reduksi Data Dimensi Tinggi Pada Analisis Sentimen Pariwisata Kota Semarang Menggunakan Algoritma Support Vector Machine,” *J. Teknol. Inf.*, vol. 15, no. 1, 2019.
- [3] D. B. Ajipangestu and R. Sarno, “Event classification in surabaya on twitter with support vector machine,”

- 2020, doi: 10.1109/iSemantic50169.2020.9234205.
- [4] S. S. Patel, "Explainable machine learning models to analyse maternal health," *Data Knowl. Eng.*, vol. 146, 2023, doi: 10.1016/j.datak.2023.102198.
- [5] G. Van den Broeck, A. Lykov, M. Schleich, and D. Suciu, "On the Tractability of SHAP Explanations," *J. Artif. Intell. Res.*, vol. 74, 2022, doi: 10.1613/jair.1.13283.
- [6] D. M. Abdullah and A. M. Abdulazeez, "Machine Learning Applications based on SVM Classification: A Review," *Qubahan Acad. J.*, vol. 1, no. 2, 2021, doi: 10.48161/qaj.v1n2a50.
- [7] A. Mastropietro, C. Feldmann, and J. Bajorath, "Calculation of exact Shapley values for explaining support vector machine models using the radial basis function kernel," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-46930-2.
- [8] Y. Zhou, W. Wu, H. Wang, X. Zhang, C. Yang, and H. Liu, "Identification of Soil Texture Classes Under Vegetation Cover Based on Sentinel-2 Data With SVM and SHAP Techniques," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, 2022, doi: 10.1109/JSTARS.2022.3164140.
- [9] C. Nawalia, F. Ningsih, and L. N. Tambunan, "Relationship Between Clean And Healthy Living Behavior (PHBS) With The Incidence Of Diarrhea In Toddlers Abstrak," *J. Surya Med.*, vol. 8, no. 2, 2022.
- [10] D. B. Ajipangestu, "View of PERANCANGAN WEBSITE DESA KRANGGANHARJO UNTUK MEDIA INFORMASI DAN PENGEMBANGAN POTENSI DESA," *Journal of Economics, Business, Management, Accounting And Social Sciences (JEBMASS)*, 2024. <http://putrajawa.co.id/ojs/index.php/jebmass/article/view/122/76> (accessed Jul. 21, 2025).
- [11] D. B. Ajipangestu and R. Sarno, "Sentiment Analysis Based on The Aspect of Culinary and Restaurant Review Using Latent Dirichlet Allocation and Support Vector Machine to Improve the Profitability of Culinary Business and Restaurant in Surabaya," in *Proceedings of the 3rd International Conference on Business and Management of Technology (ICONBMT 2021)*, 2022, vol. 202, doi: 10.2991/aebmr.k.211226.011.
- [12] Y. Yaslina, B. Andini, and N. Nofriadi, "HUBUNGAN SIKAP DAN MOTIVASI DENGAN PENERAPAN PERILAKU HIDUP BERSIH DAN SEHAT (PHBS) PADA KELUARGA DI WILAYAH KERJA PUSKESMAS GULAI BANCAH KOTA BUKITTINGGI TAHUN 2018," *J. Kesehat. PERINTIS (Perintis's Heal. Journal)*, vol. 5, no. 1, 2018, doi: 10.33653/jkp.v5i1.97.
- [13] G. Dharmarathne, T. N. Jayasinghe, M. Bogahawaththa, D. P. P. Meddage, and U. Rathnayake, "A novel machine learning approach for diagnosing diabetes with a self-explainable interface," *Healthc. Anal.*, vol. 5, 2024, doi: 10.1016/j.health.2024.100301.
- [14] N. Ahmed and A. Alpkoçak, "A quantitative evaluation of explainable AI methods using the depth of decision tree," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 30, no. 6, 2022, doi: 10.55730/1300-0632.3924.