



***SUPPORT VECTOR MACHINE DAN SHAPLEY ADDITIVE  
EXPLANATIONS UNTUK ANALISIS KEJADIAN  
RAWAN PANGAN DI JAWA TIMUR***

**TESIS**

Oleh

**M. Dika Saputra**

**NIM 201820101010**

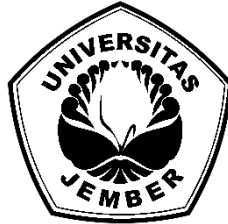
**PROGRAM STUDI MAGISTER MATEMATIKA**

**JURUSAN MATEMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS JEMBER**

**2023**



***SUPPORT VECTOR MACHINE DAN SHAPLEY ADDITIVE  
EXPLANATIONS UNTUK ANALISIS KEJADIAN  
RAWAN PANGAN DI JAWA TIMUR***

**TESIS**

Diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat untuk menyelesaikan studi pada Program Studi Magister Matematika (S-2) dan mencapai gelar Magister Sains

Oleh

**M. Dika Saputra  
NIM 201820101010**

**PROGRAM STUDI MAGISTER MATEMATIKA  
JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS JEMBER**

**2023**

## PERSEMBAHAN

Dengan menyebut nama Allah Yang Maha Pengasih dan Maha Penyayang serta sholawat dan salam selalu tercurahkan kepada Nabi Muhammad SAW, penulis persembahkan tesis ini sebagai ungkapan kebahagiaan dan rasa terima kasih kepada:

1. Ibunda Istadil Khoiriyah dan Ayahanda Sungkono serta seluruh keluarga yang selalu mendukung, memberikan doa, kasih sayang dan motivasi yang tiada henti.
2. Ibunda kedua penulis Nyai Hj. Lilik Istiqomah, S.H., M.H yang selalu mendoakan penulis.
3. Dr. Alfian Futuhul Hadi, S.Si., M.Si. selaku Dosen Pembimbing Utama dan Dr. Bagus Sartono, S.Si., M.Si. selaku Dosen Pembimbing Anggota yang telah membimbing dengan sabar dan telaten dalam penyempurnaan tugas akhir ini.
4. Prof. Drs. I Made Tirta, M.Sc., Ph.D. dan Dr. Firdaus Ubaidillah, S.Si., M.Si. selaku Dosen Penguji yang senantiasa memberikan kritik dan saran demi perbaikan tugas akhir ini.
5. Seluruh jajaran guru RA Perwanida II Tapanrejo, MI Darussholah Tapanrejo, SMPN 2 Muncar, dan SMKN Darul Ulum Muncar yang telah memberikan banyak ilmu kepada penulis.
6. Sahabat dan teman-teman seperjuangan terkhusus E.A.R yang selama ini mendukung penulis sehingga tugas akhir ini bisa terselesaikan.
7. Semua pihak yang membantu penulis dalam menyelesaikan tugas akhir.

**MOTTO**

Ruang Bahagia : ciptakanlah kebahagiaan-kebahagian kecil dalam dirimu tanpa melanggar aturan dari Tuhan.



**PERNYATAAN**

Saya yang bertanda tangan di bawah ini:

Nama : M. Dika Saputra

NIM : 201820101010

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “*Support Vector Machine dan Shapley Additive Explanations* untuk analisis kejadian rawan pangan di Jawa Timur” adalah benar-benar hasil karya ilmiah sendiri, kecuali jika dalam pengutipan substansi disebutkan sumbernya dan belum pernah diajukan pada institusi manapun serta bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, 25 Januari 2023

Yang menyatakan,

M. Dika Saputra

NIM. 201820101010

**TESIS**

***SUPPORT VECTOR MACHINE DAN SHAPLEY ADDITIVE  
EXPLANATIONS UNTUK ANALISIS KEJADIAN  
RAWAN PANGAN DI JAWA TIMUR***

Oleh  
**M. Dika Saputra**  
**NIM 201820101010**

Pembimbing

Dosen Pembimbing Utama : Dr. Alfian Futuhul Hadi, S.Si., M.Si.

Dosen Pembimbing Anggota : Dr. Bagus Sartono, S.Si., M.Si.

**PENGESAHAN**

Tesis berjudul “*Support Vector Machine dan Shapley Additive Explanations* untuk Analisis Kejadian Rawan Pangan di Jawa Timur” telah diuji dan disahkan pada:

Hari, tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Jember

Tim Penguji :

Ketua,

Anggota I,

Dr. Alfian Futuhul Hadi, S.Si., M.Si.  
NIP. 19740719 200012 1 001

Dr. Bagus Sartono, S.Si., M.Si.  
NIP. 19780411 200501 1 002

Anggota II,

Anggota III,

Prof. Drs. I Made Tirta, M.Sc., Ph.D.  
NIP. 19591220 198503 1 002

Dr. Firdaus Ubaidillah, S.Si., M.Si.  
NIP. 19700606 199803 1 003

Mengesahkan  
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Jember

Drs. Achmad Sjaifullah, M.Sc., Ph.D.  
NIP. 19591009 198602 1 001

## RINGKASAN

**SUPPORT VECTOR MACHINE DAN SHAPLEY ADDITIVE EXPLANATIONS UNTUK ANALISIS KEJADIAN RAWAN PANGAN DI JAWA TIMUR**; M. Dika Saputra, 201820101010; 2023; 63 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Pembelajaran mesin adalah metode yang memungkinkan mesin mendapatkan pengetahuan dengan memberikan *input* berupa data sebagai bahan belajar (*training*) sebelum diujikan (*testing*) untuk mendapatkan *output* yang optimal. Salah satu jenisnya adalah teknik pembelajaran mesin yang terawasi, yang mampu memetakan *input* ke *output* yang diinginkan dan mampu menghasilkan model prediktif dengan akurasi model yang sangat baik. Kemampuan pembelajaran mesin dalam model prediktif lebih unggul dari model sederhana tradisional. Terdapat berbagai macam algoritma *machine learning*, salah satunya adalah *Support Vector Machine* (SVM). Berdasarkan prinsip *Structural Risk Minimization* SVM adalah salah satu algoritma pembelajaran mesin yang paling banyak digunakan dan mencapai kinerja generalisasi yang unggul untuk klasifikasi dan regresi. SVM memiliki kelebihan dalam akurasi dan kekurangan dalam interpretasi.

Sebagian besar model yang dikembangkan oleh *data scientist* hanya fokus pada akurasi prediksi sebagai matriks kinerja tetapi jarang menjelaskan prediksi secara bermakna. Pembelajaran mesin terarah secara umum menghasilkan model yang sulit untuk dijelaskan dalam bentuk sederhana. Kesulitan model ini sering disebut dengan *black box* yaitu model dapat memberikan akurasi yang baik namun tidak mudah dalam melakukan interpretasi. Pendekatan *SHapley Additive exPlanations* (SHAP) dalam memperoleh tingkat kepentingan variabel merupakan upaya saat ini untuk menginterpretasikan model *black box*. SHAP terbukti efektif dalam menjelaskan berbagai model pembelajaran terawasi. Penjelasan SHAP yang canggih secara lokal menjelaskan dampak variabel pada prediksi individu dan baru-baru ini telah diperluas untuk memberikan penilaian global pada seluruh dataset. Interpretasi dalam model pembelajaran mesin yang terawasi dapat memberikan manfaat tambahan dalam penentuan kebijakan oleh para pengambil kebijakan.

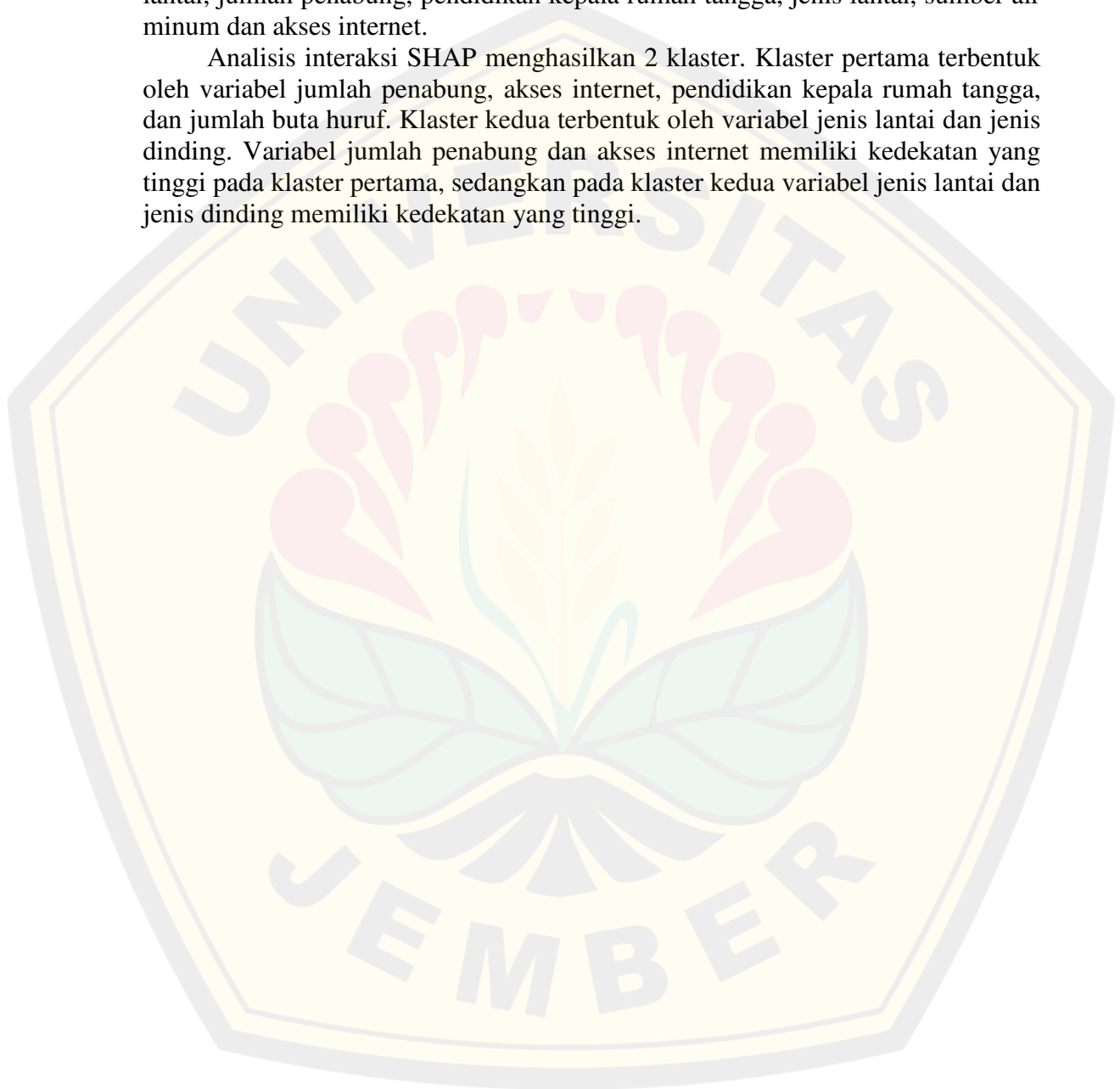
Data yang digunakan merupakan data kejadian rawan pangan berdasarkan Susenas KOR (pengumpulan data yang dilaksanakan tiap tahun) Maret 2021. Terdapat tiga tahapan analisis data, yaitu penyiapan data, pembentukan model klasifikasi, dan interpretasi model klasifikasi. Data pada penelitian ini dibagi menjadi data latih dan data uji dengan perbandingan 70:30 pada dataset rumah tangga rawan pangan, selanjutnya melakukan penyeimbangan data menggunakan teknik SMOTE pada data latih tersebut. Pembentukan model klasifikasi dilakukan pada data yang telah seimbang dengan menggunakan hiperparameter optimum. Pada metode SHAP interpretasi ditampilkan melalui SHAP *summary plot* dan *Shap Hierarchical Clustering*.

Tujuan dari penelitian adalah menjelaskan model SVM Nonlinier menggunakan *Shapley Additive Explanations*, mengetahui peran setiap variabel prediktor dari Susenas terhadap kejadian kerawanan pangan, mengkaji pengaruh dari variabel prediktor pada kejadian kerawanan pangan menggunakan *Shapley Additive Explanations*.



Berdasarkan nilai evaluasi model pada data kejadian rawan pangan Provinsi Jawa Timur memperlihatkan bahwa model SVM mampu bekerja cukup baik dalam mengklasifikasikan rumah tangga rawan pangan. Parameter optimum diperoleh pada kernel *radial basis function (rbf)*. SHAP dengan fungsi interpretasi *Permutation Explainer* mampu menjelaskan model SVM Nonlinier dengan baik. Interpretasi model SVM Nonlinier menggunakan SHAP menghasilkan urutan variabel penting penciri kejadian rawan pangan Provinsi Jawa Timur yaitu luas lantai, jumlah penabung, pendidikan kepala rumah tangga, jenis lantai, sumber air minum dan akses internet.

Analisis interaksi SHAP menghasilkan 2 kluster. Kluster pertama terbentuk oleh variabel jumlah penabung, akses internet, pendidikan kepala rumah tangga, dan jumlah buta huruf. Kluster kedua terbentuk oleh variabel jenis lantai dan jenis dinding. Variabel jumlah penabung dan akses internet memiliki kedekatan yang tinggi pada kluster pertama, sedangkan pada kluster kedua variabel jenis lantai dan jenis dinding memiliki kedekatan yang tinggi.



## PRAKATA

Puji syukur kehadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir dengan judul “*Support Vector Machine dan Shapley Additive Explanations* untuk analisis kejadian rawan pangan di Jawa Timur”. Tugas akhir ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan Magister (S-2) pada Program Studi Magister Matematika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan tugas akhir ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terimakasih kepada:

1. Dr. Alfian Futuhul Hadi, S.Si., M.Si. selaku Dosen Pembimbing Utama dan Dr. Bagus Sartono, S.Si., M.Si selaku Dosen Pembimbing Anggota yang telah meluangkan waktu, pikiran, dan perhatian dalam penulisan tugas akhir ini;
2. Prof. Drs. I Made Tirta, M.Sc., Ph.D., selaku Dosen Penguji I dan Dr. Firdaus Ubaidillah, S.Si., M.Si selaku Dosen Penguji II yang telah memberikan kritik dan saran yang membangun dalam penyempurnaan tugas akhir ini;
3. Dosen dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember;
4. Keluarga yang telah memberikan semangat dan doa tulus ikhlas penuh kasih sayangnya;
5. Teman-teman dan semua pihak yang telah membantu dan memberi semangat.

Semoga bantuan, bimbingan dan dorongan yang telah diberikan dicatat sebagai amal baik oleh Allah SWT dan mendapat balasan yang sesuai dari-Nya. Penulis juga menerima kritik dan saran dari semua pihak demi kesempurnaan tugas akhir ini. Akhirnya penulis berharap semoga tugas akhir ini dapat bermanfaat.

Jember, 25 Januari 2023

Penulis

DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>PERSEMBAHAN .....</b>	<b>ii</b>
<b>MOTTO .....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN .....</b>	<b>iv</b>
<b>HALAMAN PEMBIMBING .....</b>	<b>v</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>vi</b>
<b>RINGKASAN .....</b>	<b>vii</b>
<b>PRAKATA.....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xiv</b>
<b>BAB 1. PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan .....	3
1.4 Manfaat .....	3
<b>BAB 2. TINJAUAN PUSTAKA.....</b>	<b>4</b>
2.1 Kerawanan Pangan.....	4
2.2 <i>Preprocessing Data</i> .....	4
2.3 Balancing Data .....	5
2.4 <i>Machine Learning (ML)</i> .....	6
2.5 <i>Support Vector Machine (SVM)</i> .....	6
2.6 <i>Cross-Validation (CV)</i> .....	16
2.7 Evaluasi Model.....	17
2.8 <i>Shapley Value dan Shapley Additive Explanations (SHAP)</i> .....	18

<b>BAB 3. METODOLOGI PENELITIAN</b> .....	21
3.1. Sumber Data.....	21
3.2. Data Penelitian .....	21
3.3. Tahapan Penelitian .....	23
<b>BAB 4. HASIL DAN PEMBAHASAN</b> .....	25
4.1. Deskripsi Data.....	25
4.2. <i>Preprocessing</i> Data Sebelum Pemodelan .....	26
4.2.1 Persiapan Data.....	26
4.2.2 Weight of Evidence and Information Value .....	27
4.3. Pemodelan Klasifikasi Rawan Pangan Menggunakan SVM .....	28
4.4. Identifikasi variabel penting menggunakan SHAP pada Provinsi Jawa Timur.....	30
4.5. Identifikasi interaksi antar variabel menggunakan SHAP berdasarkan hierarchical clustering .....	34
<b>BAB 5. PENUTUP</b> .....	<b>36</b>
5.1. Kesimpulan .....	36
5.2. Saran.....	36
<b>DAFTAR PUSTAKA</b> .....	<b>37</b>
<b>LAMPIRAN</b> .....	<b>41</b>

**DAFTAR TABEL**

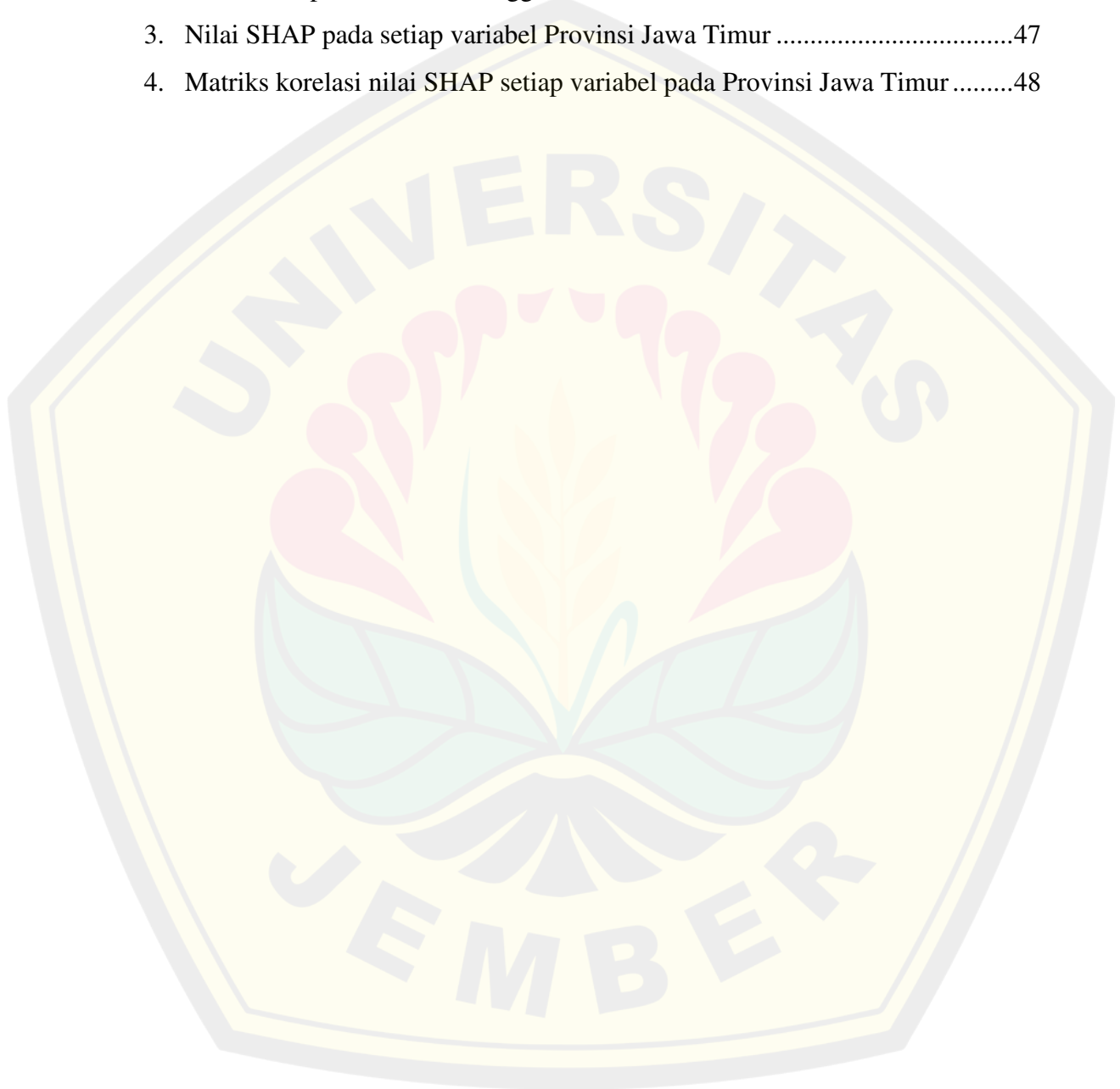
	Halaman
2.1 Fungsi kernel pada SVM .....	16
2.2 <i>Confusion Matrix</i> .....	17
3.1 Penjelasan Variabel Penelitian.....	22
4.1 Proporsi awal data latih dan data uji .....	26
4.2 Jumlah data latih awal dan data latih yang sudah diseimbangkan menggunakan teknik SMOTE .....	27
4.3 Nilai informasi variabel prediktor.....	27
4.4 Nilai informasi variabel prediktor pada data kerawanan pangan.....	28
4.5 Hiperparameter pada model SVM .....	28
4.6 Hiperparameter terbaik yang digunakan dalam pembentukan model SVM....	29
4.7 Nilai evaluasi model SVM pada data tanpa SMOTE dan data SMOTE .....	29

**DAFTAR GAMBAR**

	Halaman
2.1 Ilustrasi Hyperplane .....	7
2.2 Ilustrasi Hyperplane .....	8
2.3 Ilustrasi transformasi data dari <i>input space</i> ke <i>feature space</i> .....	15
2.4 Proses <i>splitting</i> pada 10-Cross Validation .....	16
3.1 Kerangka kerja penyiapan data penelitian .....	21
3.2 Skema tahapan proses penelitian .....	24
4.1 Perbandingan proporsi status rawan pada seluruh rumah tangga di Provinsi Jawa Timur.....	25
4.2 Urutan variabel penting Provinsi Jawa Timur .....	30
4.3 (a) <i>Scatter plot</i> nilai SHAP luas lantai Provinsi Jawa Timur.....	31
(b) <i>Scatter plot</i> nilai SHAP jumlah penabung Provinsi Jawa Timur .....	31
4.4 (a) <i>Scatter plot</i> nilai SHAP pendidikan kepala rumah tangga Provinsi Jawa Timur.....	32
(b) <i>Scatter plot</i> nilai SHAP jumlah penabung Provinsi Jawa Timur .....	32
4.5 (a) <i>Scatter plot</i> nilai SHAP sumber air minum Provinsi Jawa Timur.....	33
(b) <i>Scatter plot</i> nilai SHAP akses internet Provinsi Jawa Timur .....	33
4.6 <i>Hierarchical Clustering</i> pada Provinsi Jawa Timur .....	35

**DAFTAR LAMPIRAN**

	Halaman
1. Kuisisioner Susenas Blok XVII akses terhadap makanan pada Survei Sosial Ekonomi Nasional (SUSENAS) 2021 .....	41
2. Sebaran responden rumah tangga .....	42
3. Nilai SHAP pada setiap variabel Provinsi Jawa Timur .....	47
4. Matriks korelasi nilai SHAP setiap variabel pada Provinsi Jawa Timur .....	48



## BAB 1. PENDAHULUAN

### 1.1 Latar Belakang

Pembelajaran mesin (ML) adalah subbidang kecerdasan buatan yang mencakup berbagai pendekatan dengan memberikan komputer kemampuan untuk mempelajari tugas-tugas tanpa diprogram secara eksplisit (Sarker, 2021). Algoritma pembelajaran mesin bekerja dengan mempelajari atau mengungkap pola mendasar yang tertanam dalam data. Sebagian besar model yang dikembangkan oleh para ilmuwan data berfokus pada akurasi prediksi sebagai metrik kinerja tetapi jarang menjelaskan prediksi dengan cara yang berarti (Futoma, dkk. 2015; Basu Roy, dkk. 2015). Hal ini berlaku dengan pembelajaran mesin yang kompleks. *Support vector machine* atau SVM (Cortes dan Vapnik 1995), berdasarkan prinsip *Structural Risk Minimization* (SRM) dari teori pembelajaran statistik, adalah salah satu algoritma pembelajaran mesin yang paling banyak digunakan dan mencapai kinerja generalisasi yang unggul untuk klasifikasi dan regresi (Kang dan Cho 2014).

Pada beberapa penelitian SVM secara umum memiliki akurasi dan stabilitas prediksi yang tinggi, akan tetapi SVM merupakan model klasifikasi atau prediksi dengan kompleksitas yang tinggi sehingga sulit untuk dijelaskan (Molnar, 2019). Penentuan nilai kepentingan variabel sering dilihat dari koefisien regresi (dalam model regresi). Variabel yang memiliki koefisien paling besar dianggap sebagai variabel yang memiliki kontribusi terbesar. Pada kasus SVM linier kontribusi variabel dapat dijelaskan dengan menggunakan koefisien variabel akan tetapi, dalam SVM nonlinier kontribusi variabel menjadi lebih sulit dijelaskan karena data ditransformasi ke dalam ruang dimensi yang lebih tinggi dengan memanfaatkan kernel *trick*. Interpretasi prediksi ML dapat dicoba secara global atau pada tingkat prediksi individu. Pendekatan pembobotan *feature* biasanya bergantung pada penilaian bobot atau nilai kepentingan global untuk model dan kumpulan data pelatihan tertentu. Di sisi lain, keputusan model juga dapat dijelaskan dengan fokus pada prediksi individu dan kontribusi *feature* pada data.

*Shapley Additive Explanations (SHAP)* adalah metode interpretasi prediksi model machine learning yang berbasis teori permainan optimal melalui nilai



Shapley dan dapat dirasionalisasikan sebagai perpanjangan dari pendekatan *Local Interpretable Model-agnostic Explanations* (LIME) (Lundberg dan Lee, 2017; Rodríguez-Pérez dan Bajorath, 2020). SHAP dapat menunjukkan kontribusi setiap variabel penjelas terhadap nilai dugaan pada suatu model (Štrumbelj dan Kononenko 2014; Lundberg dan Lee 2017). Nilai Shapley adalah satu-satunya solusi yang memenuhi sifat *Efficiency*, *Symmetry*, *Dummy* dan *Additivity*. SHAP juga memenuhi karena menghitung nilai Shapley. Selain itu, dalam interpretasi SHAP menjelaskan tiga sifat yakni *Local accuracy*, *Missingness*, Konsistensi. Metode SHAP memiliki komputasi yang cepat sehingga memungkinkan untuk menghitung banyak nilai Shapley yang diperlukan untuk interpretasi model global. Metode interpretasi global mencakup *feature importance*, *feature dependence*, interaksi, pengelompokan dan *summary plot*. Metode SHAP menghasilkan tingkat kepentingan peubah yang dapat menjadi solusi dalam melakukan interpretasi model *black-box* yang berguna bagi pemerintah dalam mengambil kebijakan dan menentukan arah pembangunan. Salah satu fokus pembangunan dunia termasuk Indonesia yaitu kerawanan pangan.

Kerawanan pangan merupakan suatu kondisi ketidakmampuan individu atau sekelompok individu pada waktu tertentu dalam suatu wilayah untuk memperoleh pangan yang cukup dan bergizi untuk hidup sehat dan beraktivitas dengan baik (BKP, 2013). Kerawanan pangan menjadi permasalahan yang sangat penting karena merupakan salah satu tujuan Pembangunan Berkelanjutan atau *Sustainable Development Goals* (SDGs) yaitu *Zero Hunger*. Upaya mencegah terjadinya kerawanan pangan diperlukan sejumlah usaha untuk mengatasi kejadian rawan pangan. Usaha terkini yang dapat dilakukan adalah membuat model klasifikasi pada kejadian rawan pangan rumah tangga atau individu. Pada beberapa bidang, misalnya pada pemerintahan, tidak hanya dibutuhkan nilai kebaikan model klasifikasi tetapi juga dibutuhkan penjelasan dari model klasifikasi yang diperoleh. Oleh karena itu, penulis tertarik untuk melakukan implementasi metode SVM-SHAP terhadap klasifikasi kejadian rawan pangan level rumah tangga. Adapun data pendukung yang digunakan adalah data Survei Sosial Ekonomi Nasional (SUSENAS) Provinsi Jawa Timur tahun 2021

## 1.2 Rumusan Masalah

Berdasarkan permasalahan di atas, rumusan masalah yang akan dibahas pada penelitian ini adalah sebagai berikut.

1. Bagaimana menjelaskan model SVM Nonlinier menggunakan *Shapley Additive Explanations*.
2. Bagaimana peran variabel prediktor dari SUSENAS terhadap kejadian kerawanan pangan di Jawa Timur.
3. Bagaimana pengaruh (interpretasi) dari setiap prediktor pada kejadian kerawanan pangan di Jawa Timur.

## 1.3 Tujuan

Berdasarkan rumusan masalah di atas, tujuan penelitian ini adalah.

1. Memperoleh penjelasan model SVM Nonlinier menggunakan *Shapley Additive Explanations*.
2. Mengetahui peran setiap variabel prediktor dari SUSENAS terhadap kejadian kerawanan pangan.
3. Mengkaji pengaruh dari variabel prediktor pada kejadian kerawanan pangan menggunakan *Shapley Additive Explanations*.

## 1.4 Manfaat

Secara umum manfaat yang diharapkan dari penelitian ini adalah dapat menambah wawasan keilmuan tentang metode klasifikasi SVM-SHAP. Secara khusus manfaat penelitian ini diharapkan dapat berperan dalam keberhasilan *Sustainable Development Goals* di Indonesia. Sebagai perencanaan respon krisis pangan dalam upaya pencegahan kerawanan pangan. Selain itu, penelitian ini diharapkan mampu memberikan pertimbangan atau rekomendasi bagi pembuat kebijakan dalam mengatasi kejadian rawan pangan di Jawa Timur berdasarkan level individu maupun wilayah.

## BAB 2. TINJAUAN PUSTAKA

### 2.1 Kerawanan Pangan

Menurut BKP (2019), kerawanan pangan dapat didefinisikan sebagai suatu kondisi ketidakmampuan individu atau sekumpulan individu di suatu wilayah untuk memperoleh pangan yang cukup dan sesuai untuk hidup sehat dan aktif. Pada Permentan No 43 Tahun 2010 kerawanan pangan dapat diartikan sebagai kondisi suatu daerah, masyarakat atau rumah tangga yang tingkat ketersediaan dan keamanan pangannya tidak cukup untuk memenuhi standar kebutuhan fisiologis bagi pertumbuhan dan kesehatan sebagian masyarakat. Secara umum kerawanan pangan dipengaruhi oleh ketersediaan (*food availability*), keterjangkauan (*access to supplies*) dan konsumsi pangan (*food utilization*)

Ketersediaan pangan adalah tersedianya pangan secara fisik di daerah tertentu, yang diperoleh baik dari hasil produksi sendiri, impor atau perdagangan maupun bantuan pangan. Ketersediaan pangan dapat dihitung pada level masyarakat, kabupaten, provinsi, dan nasional. Ketersediaan pangan di suatu daerah mungkin mencukupi, tetapi tidak semua rumah tangga memiliki akses yang memadai baik secara kuantitas maupun keragaman pangan melalui mekanisme tersebut. Kemampuan rumah tangga memperoleh cukup pangan merupakan aspek yang perlu untuk diperhatikan. Pemanfaatan pangan merupakan penggunaan pangan oleh rumah tangga meliputi cara penyimpanan, pengolahan, dan penyiapan makanan dalam rumah tangga sesuai kebutuhan masing-masing individu dan status kesehatan masing-masing anggota rumah tangga.

### 2.2 *Preprocessing Data*

Sebuah data dikatakan kurang baik jika terdapat permasalahan berupa *noisy data* (pencilan), *missing value* (data tidak lengkap) dan *incosistent* (format atau skala data yang berbeda) karena memiliki ukuran yang besar serta berasal dari berbagai sumber yang membuat kualitas data kurang baik. Kualitas data akan berdampak pada kualitas pemodelan dan hasil analisisnya. *Preprocessing data* merupakan tahap awal pengolahan data yang dilakukan untuk meningkatkan

kualitas data sebelum melakukan analisis data lebih lanjut sehingga menghasilkan model yang baik dan meningkatkan efisiensi pemodelan. Salah satu metode dalam *preprocessing data* yaitu normalisasi data.

### 2.3 *Balancing Data*

Data tidak seimbang merupakan keadaan dimana distribusi kelas tidak seimbang, yaitu jumlah kelas data (*instance*) yang satu lebih sedikit atau lebih banyak dibanding dengan jumlah kelas lainnya. Klasifikasi kelompok kelas data yang memiliki jumlah lebih sedikit yang dikenal dengan kelompok minoritas (*minority class*), sedangkan kelompok kelas data yang memiliki lebih banyak dikenal dengan kelompok mayoritas (*majority class*). Permasalahan ketidakseimbangan kelas (*class imbalance*) merupakan permasalahan penting yang harus diatasi. Kondisi tersebut dapat menyulitkan metode klasifikasi dalam melakukan fungsi generalisasi pada algoritma *machine learning*. Hampir semua algoritma klasifikasi menunjukkan performa yang sangat buruk ketika bekerja pada data dengan kelas yang sangat tidak seimbang (Qiong, dkk., 2016; Gu, dkk., 2016). Metode klasifikasi lebih mudah mengklasifikasikan kelas mayoritas dibandingkan kelas minoritas.

Permasalahan ketidakseimbangan kelas dapat ditangani dengan dua pendekatan yaitu pendekatan level data dan pendekatan level algoritma (Peng dan Yao, 2010). Salah satu pendekatan level data yang baik digunakan adalah teknik sampling yaitu *oversampling*, *undersampling*, atau gabungan dari keduanya (Yap, dkk., 2014). *Synthetic Minority Over-sampling Technique (SMOTE)* merupakan metode yang dapat mengatasi permasalahan ketidakseimbangan data minoritas dengan membuat data sintesis dari data tersebut. Metode *SMOTE* bekerja dengan mencari *k-nearest neighbors* (ketetanggaan terdekat data sebanyak *k*) pada setiap data kelas minoritas, kemudian dibuat data sintesis sebanyak prosentase duplikasi yang diinginkan antara data minor dan *k-nearest neighbors* yang dipilih secara acak. Selain itu, metode *random undersampling* merupakan metode yang digunakan untuk mengatasi permasalahan ketidakseimbangan data mayoritas.

Proses *undersampling* bekerja dengan mengurangi data mayoritas hingga jumlah tertentu secara acak.

#### 2.4 *Machine Learning* (ML)

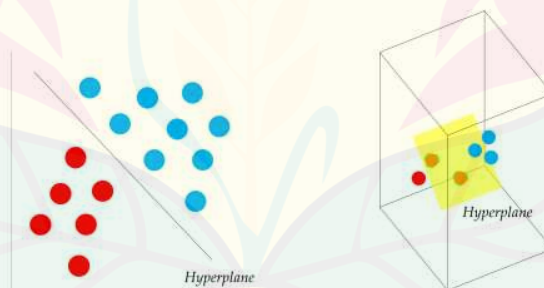
*Machine Learning* (ML) atau Mesin Pembelajaran adalah cabang dari *AI* yang fokus belajar dari data (*learn from data*), yaitu fokus pada pengembangan sistem yang mampu belajar secara “mandiri” tanpa harus berulang kali diprogram manusia (Cholissodin, dkk., 2020). *Machine Learning* membutuhkan data yang valid sebagai bahan belajar (*training*) sebelum diujikan (*testing*) untuk mendapatkan *output* yang optimal. Metode dalam ML terdiri dari dua pendekatan yaitu *Supervised Learning* dan *Unsupervised Learning*. *Supervised Learning* adalah pembelajaran terarah/terawasi, artinya pada pembelajaran ini serangkaian pengamatan dimana hasil yang diharapkan telah diketahui sebelumnya. Pada algoritma *supervised learning*, terdapat dua variabel, yaitu variabel *input* yang biasa disebut variabel  $X$  dan variabel *output* yang biasa disebut variabel  $Y$ . Tujuan algoritma *supervised learning* adalah untuk mempelajari fungsi pemetaan dari variabel  $X$  ke variabel  $Y$ . Rumus umum pemetaan variabel  $X$  dan  $Y$  adalah  $Y = f(x)$ . Tujuan akhir dari algoritma *supervised learning* adalah untuk memperkirakan fungsi pemetaan ( $f$ ) agar kita dapat memprediksi variabel  $Y$  ketika kita memiliki data *input* (variabel  $X$ ) yang baru. Algoritma *Unsupervised Learning* lebih bebas dalam proses eksplorasi data karena tidak memiliki data label dan dapat mencari karakteristik data yang tersembunyi. Algoritma *Unsupervised Learning* menggunakan titik data sebagai referensi untuk menemukan struktur dan pola yang ada di dalam data set. Salah satu bentuk dari algoritma *Unsupervised Learning* adalah *Clustering*.

#### 2.5 *Support Vector Machine* (SVM)

*Support Vector Machine* (SVM) merupakan salah satu metode pembelajaran terawasi yang biasa digunakan untuk klasifikasi (*Support Vector Classification*) dan regresi (*Support Vector Regression*) (Samsudiney, 2019). Tujuan dari SVM adalah menemukan *hyperplane* terbaik dengan memaksimalkan jarak antar kelas. SVM

memiliki prinsip dasar *linear classifier* yaitu kasus klasifikasi yang secara linier kedua kelas dapat dipisahkan menggunakan *hyperplane* dengan *soft margin*, namun SVM telah dikembangkan agar dapat bekerja pada masalah nonlinier dengan memasukkan konsep kernel *trick* pada ruang kerja berdimensi tinggi (Cholissodin, dkk., 2020 : Hamel, 2009).

Pada tahun 1992 *Support Vector Machine* (SVM) diperkenalkan oleh Vladimir Vapnik, Bernhard Boser, dan Isabelle Guyon. Sebelum tahun 1992 tepatnya tahun 1950-an dasar untuk SVM sudah ada seperti, margin *hyperplane kernel* yang diperkenalkan oleh Aronszajn pada tahun 1950 dan karya awal oleh Vapnik dan Alexei Chervonenkis pada teori belajar statistik. SVM membutuhkan *training set* positif dan negatif. *Training set* ini dibutuhkan untuk membuat keputusan terbaik dalam memisahkan data positif dengan data negatif di ruang  $n$ -dimensi, yang disebut dengan *hyperplane*. Secara sederhana konsep SVM dapat dinyatakan sebagai usaha menemukan fungsi pemisah (*hyperplane*) yang optimal yang dapat memisahkan dua set data dari dua kelas yang berbeda seperti pada Gambar 2.1.



Gambar 2.1 Ilustrasi *Hyperplane*  
(Sumber: Putra, 2020)

Misalkan diberikan dua atribut input  $A_1$  dan  $A_2$ , himpunan  $W = \{w_1, w_2, \dots, w_d\}$ ;  $W$  merupakan bobot/*weight*,  $d$  adalah jumlah atribut, tupel training  $X = (x_1, x_2)$  dimana  $x_1$  dan  $x_2$  adalah nilai-nilai atribut  $A_1$  dan  $A_2$ , dan  $y \in \{-1, +1\}$  merupakan kelas data dari atribut  $A_1$  dan  $A_2$ . Secara matematis fungsi *hyperplane* didefinisikan

$$f(x) = w_1x_1 + w_2x_2 + b$$

$$f(x) = W \cdot X + b \quad (2.1)$$

dimana  $W, X \in \mathbb{R}^d$  dan  $b$  (*threshold*) adalah bias yang berupa skalar (Hamel, 2009).

*Hyperplane* yang terletak diantara dua set objek dari kelas positif ( $y = +1$ ) dan kelas negatif ( $y = -1$ ) dapat dituliskan sebagai berikut:

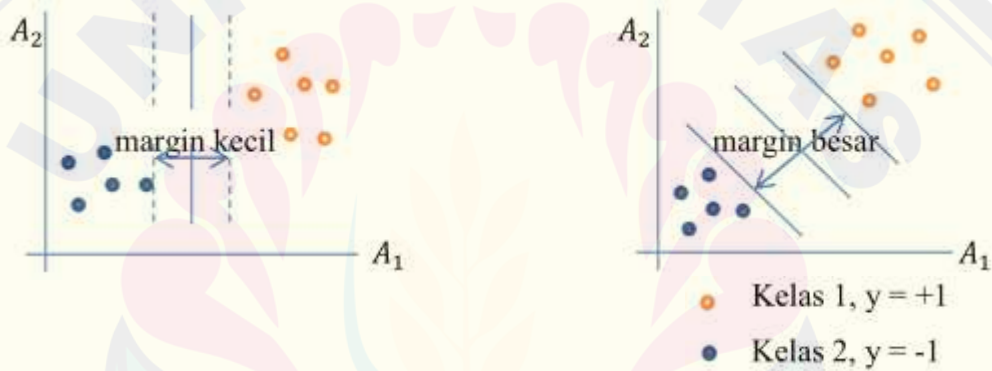
$$H_1: x_i w + b \geq 1 \text{ untuk } y_1 = +1 \quad (2.2)$$

$$H_2: x_i w + b \leq -1 \text{ untuk } y_2 = -1 \quad (2.3)$$

Penggabungan dari persamaan (2.2) dan persamaan (2.3) menghasilkan pertidaksamaan

$$y_i(x_i w + b) \geq 1 \text{ untuk } \forall i = 1, 2, \dots, n \quad (2.4)$$

dengan  $n$  adalah jumlah data *training*. Ilustrasi *hyperplane* yang memisahkan dua set objek dengan margin maksimal dapat dilihat pada Gambar 2.2.



Gambar 2.2 ilustrasi *hyperplane*  
(Sumber: Meirista, 2015)

Margin antara dua kelas dapat dihitung dengan mencari jarak antara kedua *hyperplane*  $H_1$  atau  $H_2$ . Setiap data pelatihan yang jatuh pada  $H_1$  atau  $H_2$  yang memenuhi persamaan (2.2) disebut *support vector*. Jarak terdekat suatu titik di bidang  $H_1$  terhadap pusat (*hyperplane*) dapat dihitung dengan meminimalkan  $x^T x$  dengan memperhatikan kendala  $x_i w + b \geq 1$ . Dengan menggunakan bentuk fungsi *Lagrange* dan turunan pertamanya diperoleh

$$\min x^T x - \lambda(w^T x + b - 1)$$

$$\frac{d}{dx} (x^T x - \lambda(w^T x + b - 1)) = 0$$

$$2x - \lambda w^T = 0$$

$$x = \frac{\lambda}{2} w^T$$

dengan mensubstitusikan  $x$  ke bidang  $H_1: x_i w + b = 1$  diperoleh :

$$\frac{\lambda}{2} w^T w + b = 1$$

$$\lambda = \frac{2(1-b)}{w^T w}$$

sehingga dengan mensubstitusikan kembali  $\lambda$  pada  $x$  diperoleh

$$x = \frac{(1-b)}{w^T w} w$$

$$x^T x = \frac{(1-b)^2}{(w^T w)^2} w^T w$$

$$= \frac{(1-b)^2}{w^T w}$$

maka jarak  $H_1$  terhadap pusat adalah :

$$\|x\| = \sqrt{x^T x} = \sqrt{\frac{(1-b)^2}{w^T w}} = \frac{(1-b)}{\|x\|}$$

untuk mencari jarak terdekat suatu titik pada bidang  $H_2$  terhadap titik pusat dapat dihitung dengan meminimalkan  $x^T x$  dengan memperhatikan kendala  $x_i w + b \leq -1$ . Dengan menggunakan bentuk fungsi Lagrange dan turunan pertamanya diperoleh:

$$\min x^T x + \lambda(w^T x + b + 1)$$

$$\frac{d}{dx} (x^T x + \lambda(w^T x + b + 1)) = 0$$

$$2x + \lambda w^T = 0$$

$$x = -\frac{\lambda}{2} w^T$$

dengan mensubstitusikan  $x$  ke bidang  $H_2: x_i w + b = -1$  diperoleh :

$$-\frac{\lambda}{2} w^T w + b = -1$$

$$\lambda = \frac{2(b+1)}{w^T w}$$

sehingga dengan mensubstitusikan kembali  $\lambda$  pada  $x$  diperoleh :



$$x = \frac{(-b-1)}{w^T w} w$$

$$x^T x = \frac{(-b-1)^2}{(w^T w)^2} w^T w$$

$$= \frac{(-b-1)^2}{w^T w}$$

maka jarak  $H_2$  ke pusat adalah :

$$\|x\| = \sqrt{x^T x} = \sqrt{\frac{(-b-1)^2}{w^T w}} = \frac{(-b-1)}{\|w\|}$$

margin maksimum didapatkan dengan memaksimalkan jarak antara  $H_1$  dan  $H_2$  :

$$\left| \frac{(1-b)}{\|w\|} - \frac{(-b-1)}{\|w\|} \right| = \frac{2}{\|w\|}$$

oleh karena memaksimalkan  $\frac{1}{\|w\|}$  sama dengan meminimalkan  $\|w\|^2$  dan untuk menyederhanakan penyelesaian ditambahkan faktor  $\frac{1}{2}$ . Dengan demikian, model persamaannya menjadi :

$$\text{Min } \frac{1}{2} \|w\|^2$$

$y_i(x_i w + b) \geq 1$ , untuk  $\forall i = 1, 2, \dots, n$  ( $n$  merupakan jumlah data *training*).

Untuk menyelesaikan persamaan tersebut, secara komputasi lebih sulit dan perlu waktu yang cukup panjang sehingga digunakan pengali Langrangian  $\alpha_i$ , dengan  $i = 1, 2, \dots, n$ . Model masalah sebelumnya dapat ditulis :

$$\begin{aligned} \text{Min } L_p &= \frac{1}{2} \|w\|^2 + \alpha_i (0 - \sum_{i=1}^n [y_i(x_i \cdot w + b) - 1]) \\ L_p &= \frac{1}{2} \|w\|^2 + 0 - \alpha_i (\sum_{i=1}^n [y_i(x_i \cdot w + b) - 1]) \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i(x_i \cdot w + b) - 1] \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i(x_i \cdot w + b) + \sum_{i=1}^n \alpha_i \end{aligned} \quad (2.5)$$

Solusi dapat diperoleh dengan meminimalkan  $L_p$  terhadap variabel  $w, b$  (*primal variable*):

$$\frac{\partial L_p}{\partial b} = 0$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (2.6)$$

dan

$$\begin{aligned}\frac{\partial L_p}{\partial w} &= 0 \\ w - \sum_{i=1}^n \alpha_i y_i x_i &= 0 \\ w &= \sum_{i=1}^n \alpha_i y_i x_i\end{aligned}\quad (2.7)$$

Dan memaksimalkan  $L_D$  terhadap variabel  $\alpha_i$  (*dual variables*). untuk mendapatkan problem dual ( $L_D$ ), substitusikan persamaan (2.6) dan (2.7) ke (2.4)

$$\begin{aligned}\text{Maks } L_D &= L_p = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i(x_i \cdot w + b) - 1] \\ \|w\|^2 &= (w \cdot w) \\ &= \left( \sum_{i=1}^n \alpha_i y_i x_i \cdot \sum_{j=1}^n \alpha_j y_j x_j \right) \\ &= \underbrace{\sum_{i=1, j=1}^n \alpha_i y_i x_i \alpha_j y_j x_j}_A\end{aligned}\quad (2.8)$$

dan

$$\begin{aligned}\sum_{i=1}^n \alpha_i [y_i(x_i \cdot w + b) - 1] &= \sum_{i=1}^n \alpha_i y_i x_i w + \sum_{i=1}^n \alpha_i y_i b - \sum_{i=1}^n \alpha_i \\ &= \sum_{i=1}^n \alpha_i y_i x_i + \sum_{j=1}^n \alpha_j y_j x_j + 0 - \sum_{i=1}^n \alpha_i \\ &= \underbrace{\sum_{i=1, j=1}^n \alpha_i y_i x_i \alpha_j y_j x_j}_A - \underbrace{\sum_{i=1}^n \alpha_i}_B\end{aligned}\quad (2.9)$$

sehingga diperoleh

$$\begin{aligned}L_D &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i(x_i \cdot w + b) - 1] \\ &= \frac{1}{2} A - (A - B) \\ &= \frac{1}{2} A - A + B \\ &= -\frac{1}{2} A + B \\ &= B - \frac{1}{2} A \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i y_i x_i \alpha_j y_j x_j\end{aligned}\quad (2.10)$$

$\sum_{i=1}^n \alpha_i y_i = 0$ , dan  $0 \leq \alpha_i, i = 1, 2, \dots, n$  dengan demikian, dapat diperoleh nilai  $\alpha_i$  yang akan digunakan untuk menentukan  $w$  dan  $b$  dengan rumus :

$$\begin{aligned}w &= \sum_{i=1}^n \alpha_i y_i x_i \\ b &= y_i - w x_i\end{aligned}\quad (2.11)$$

terdapat nilai  $\alpha_i$  untuk setiap data pelatihan. Data pelatihan yang memiliki nilai  $\alpha_i \geq 0$  adalah *support vector*. Dengan demikian fungsi keputusan yang dihasilkan hanya dipengaruhi oleh *support vector*.

Berdasarkan formula *Lagrangian, maximum marginal hyperplane* (MMH) dapat didefinisikan sebagai *boundary decision*:

$$f(x_d) = \sum_{i=1}^n \alpha_i y_i x_i + b \quad (2.12)$$

dimana  $y_i$  adalah label kelas *support vector*  $X_i$ , untuk  $X_i$  adalah data uji,  $\alpha_i$  dan  $b$  adalah parameter numerik yang ditentukan secara otomatis oleh optimasi algoritma SVM dan  $l$  adalah jumlah *support vector*.

Pada pengklasifikasian data tidak semua dataset dapat dipisahkan secara linier, sehingga diperlukan penyelesaian khusus untuk permasalahan ini. Pada data-data yang tidak dapat dipisahkan secara linier ditambahkan variabel *slack*  $\xi_i \geq 0$  ke pertidaksamaan (2.3) sehingga kendala dan fungsi tujuan menjadi :

$$y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0, \text{ untuk } \forall_i = 1, 2, \dots, n \quad (2.13)$$

dengan  $n$  merupakan jumlah data *training* dan

$$\text{Min } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (2.14)$$

dengan  $y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0, \xi_i \geq 0$  untuk  $\forall_i$

Dengan meminimalkan  $\|w\|^2$  (Persamaan 2.14), berarti memaksimalkan margin antara dua kelas. Hal ini dilakukan untuk mengurangi jumlah kesalahan klasifikasi (*missclassification error*) dengan adanya variabel *slack*  $\xi_i$ . Variabel *slack* digunakan untuk mengatasi kasus ketidaklayakan (*infeasibility*) dan pembatas (*constraints*) dengan memberikan pinalti untuk data yang tidak memenuhi pembatas tersebut. Pinalti ini dinotasikan dengan C (Vijayakumar dan Wu, 1999)

Selanjutnya, bentuk primal problem pada persamaan (2.5) berubah menjadi

$$\begin{aligned} \text{Min } L_p = & \text{Min } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n a_i [y_i(x_i \cdot w + b) - 1 + \xi_i] \\ & - \sum_{i=1}^n q_i \xi_i \end{aligned} \quad (2.15)$$

dengan cara penurunan  $L_p$  (persamaan 2.6 dan 2.7) pada data linier, maka persamaan untuk data nonlinier adalah sebagai berikut

$$\begin{aligned} \text{Min } L_p &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(x_i \cdot w + b) - 1 + \xi_i] - \sum_{i=1}^n q_i \xi_i \\ \|w\|^2 &= (w \cdot w) \\ &= \sum_{i=1}^n \alpha_i y_i x_i \cdot \sum_{j=1}^n \alpha_j y_j x_j \\ &= \sum_{i=1, j=1}^n \alpha_i y_i x_i \alpha_j y_j x_j \end{aligned}$$

dan

$$\begin{aligned} C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(x_i \cdot w + b) - 1 + \xi_i] - \sum_{i=1}^n q_i \xi_i \\ &= C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i y_i x_i w + \sum_{i=1}^n \alpha_i y_i b - \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \alpha_i \xi_i - \\ &\quad \sum_{i=1}^n q_i \xi_i \\ &= C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i y_i x_i \sum_{i=1}^n \alpha_i y_i x_i + 0 - \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n q_i \xi_i \\ &= C \sum_{i=1}^n \xi_i - \sum_{i=1, j=1}^n \alpha_i y_i x_i \alpha_j y_j x_j - \sum_{i=1}^n \alpha_i \end{aligned}$$

sehingga diperoleh

$$\begin{aligned} L_p(w, x, \xi_i) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(x_i \cdot w + b) - 1 + \xi_i] \\ &\quad - \sum_{i=1}^n q_i \xi_i \end{aligned} \tag{2.16}$$

untuk meminimumkan  $L_p(w, x, \xi_i)$  digunakan  $\frac{\partial L_p}{\partial b} = 0$ ,  $\frac{\partial L_p}{\partial w} = 0$ , dan  $\frac{\partial L_p}{\partial \xi_i} = 0$

dimana  $i = 1, 2, \dots, n$  ( $n$  merupakan jumlah data training). maka

$$\frac{\partial L_p}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0$$

$$\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0$$

$$\text{dengan } w = \sum_{i=1}^n \alpha_i y_i x_i$$

$$\frac{\partial L_p}{\partial \xi_i} = C - \alpha_i - q_i = 0$$

$$\text{akibatnya } C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n q_i \xi_i = 0$$

dengan memanfaatkan turunan parsial  $L_p$  terhadap  $w, x$ , dan  $\xi_i$  masalah dual menjadi

$$\begin{aligned}
\text{Maks } L_D &= \frac{1}{2} (\sum_{i=1}^n \alpha_i y_i x_i) (\sum_{j=1}^n \alpha_j y_j x_j) + C \sum_{i=1}^n \xi_i \\
&\quad - \sum_{i=1}^n \alpha_i y_i ((\sum_{i=1}^n \alpha_i y_i x_i) \cdot x_i + b) + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i - \\
&\quad \sum_{i=1}^n q_i \xi_i \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j (x_i \cdot x_j) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j (x_i \cdot x_j) \\
&\quad - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n \alpha_i \xi_i \\
&= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j (x_i \cdot x_j) \tag{2.17}
\end{aligned}$$

dengan kendala

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Karena

$$\frac{\partial L_p}{\partial \xi_i} = C - \alpha_i - q_i = 0$$

dan semua nilai  $q_i$  tidak diperhitungkan pada fungsi tujuan, maka kendalanya menjadi

$$0 \leq \alpha_i \leq C, \text{ untuk } i = 1, 2, \dots, n$$

Bias yang sebelumnya telah didefinisikan sebagai  $b$ , dapat didefinisikan secara implisit sebagai atribut tambahan sehingga data input  $X_i = [x_{1i}, x_{2i}, \dots, x_{mi}]$  menjadi  $X_j^T = [x_{i1}, x_{i2}, \dots, x_{im}]$ , vektor bobot  $w$  menjadi  $w^T = [w_1, w_2, \dots, w_m]$ , dan  $b/\lambda$  dimana  $\lambda$  merupakan konstanta.

Bentuk permasalahan menjadi meminimumkan :

$$L_p = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i (1 - y_i) (x_i^T \cdot w^T) (-\xi_i) + \sum_{i=1}^n q_i (-\xi_i)$$

untuk meminimumkan  $L_p$  digunakan  $\frac{\partial L_p}{\partial w^T} = 0$ , dan  $\frac{\partial L_p}{\partial \xi_i} = 0$  dengan  $i = 1, 2, \dots, n$

dengan  $n$  merupakan jumlah data *training* maka

$$\frac{\partial L_p}{\partial w^T} = w^T - \sum_{i=1}^n \alpha_i y_i X_i = 0$$

$$w = \sum_{i=1}^n \alpha_i y_i X_i$$

$$\frac{\partial L_p}{\partial \xi_i} = C - \alpha_i - q_i = 0$$

Sehingga diperoleh

$$\begin{aligned}
L_D &= \frac{1}{2} (\sum_{i=1}^n \alpha_i y_i X_i) (\sum_{j=1}^n \alpha_j y_j X_j^T) + C \sum_{i=1}^n \xi_i \\
&\quad + \sum_{i=1}^n \alpha_i (1 - y_i (\sum_{j=1}^n \alpha_j y_j X_j \cdot X_i^T) - \xi_i) + \sum_{i=1}^n q_i (-\xi_i)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} \sum_{i=1}^n \alpha_i y_i X_i \sum_{j=1}^n \alpha_j y_j X_j^T + \sum_{i=1}^n \alpha_i \left( y_i \left( \sum_{j=1}^n \alpha_j y_j X_j \cdot X_i^T \right) \right) + \\
&\quad + \sum_{i=1}^n \alpha_i + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n \alpha_i \xi_i \\
&= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i y_i X_i \sum_{j=1}^n \alpha_j y_j X_j^T \\
&= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j (X_i \cdot X_j^T) \\
&= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i \alpha_j y_j K(X_i \cdot X_j^T) \tag{2.18}
\end{aligned}$$

dengan kendala

$$0 \leq \alpha_i \leq C, \text{ untuk } i = 1, 2, \dots, n$$

#### 4.2.1. Fungsi Kernel

Pada umumnya, kasus klasifikasi di dunia nyata kebanyakan bersifat non linier sehingga digunakan fungsi kernel untuk mengatasi masalah tersebut. Dalam penggunaan fungsi kernel, suatu data  $x$  di *input space* dipetakan ke *feature space*  $F$  dengan dimensi yang lebih tinggi melalui *map*  $\varphi(\varphi: x \rightarrow \varphi(x))$ . Fungsi  $\varphi(x)$  tidak dapat dihitung tetapi, *dot product* dari dua vektor dapat dihitung baik dalam *input space* maupun di *feature space*

$$K(X_i, X_j^T) = \varphi(X_i) \cdot \varphi(X_j) \tag{2.19}$$

pada *feature space*, *hyperplane* yang memisahkan kedua kelas tersebut dapat dikonstruksikan ilustrasi dari konsep tersebut dapat dilihat pada Gambar 2.3



Gambar 2.3 Ilustrasi transformasi data dari *input space* ke *feature space* (sumber: Putra, 2020)

Beberapa fungsi kernel lain yang umum dipakai dalam literasi SVM dapat dilihat pada Tabel 2.1

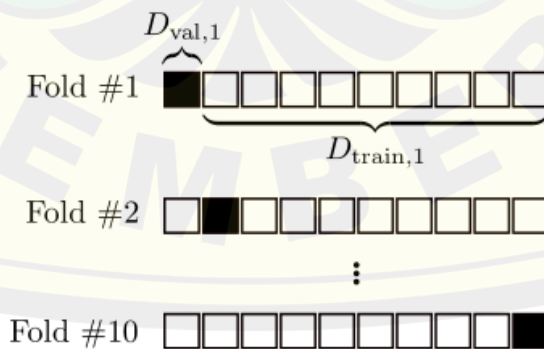
Tabel 2.1 Fungsi kernel pada SVM

Jenis Kernel	Definisi
Derajat Polinomial $h$	$K(X_i, X_j^T) = (X_i \cdot X_j^T + 1)^h$ (2.20)
Gaussian Radial Basis	$K(X_i, X_j^T) = \frac{e^{-\ x_i - x_j\ ^2}}{2\sigma^2}$ (2.21)
Sigmoid	$K(X_i, X_j^T) = \tanh(\alpha X_i \cdot X_j^T - \delta)$ (2.22)

(Nugroho, dkk., 2003)

### 2.6 Cross-Validation (CV)

*Cross Validation* (CV) merupakan metode statistik yang dapat digunakan untuk mengevaluasi dan membandingkan kinerja model dalam algoritma pembelajaran (Refaeilzadeh, dkk., 2016). Secara umum, metode ini digunakan dalam memprediksi model dan memperkirakan seberapa presisi sebuah model prediktif ketika dijalankan. Konsep kerja dari metode ini adalah dengan membagi data menjadi dua segmen. Segmen satu digunakan untuk mempelajari atau melatih model dan segmen 2 digunakan untuk memvalidasi model. Pada *cross validation* dikenal sebagai estimasi rotasi, dengan membagi data menjadi himpunan bagian k dengan ukuran yang hampir sama, data dilatih dan diuji sebanyak k. Setiap pengulangan, salah satu himpunan bagian akan digunakan sebagai data pengujian dan himpunan data k lainnya sebagai data pelatihan. Ilustrasi penggunaan *cross validation* dapat dilihat pada Gambar 2.4. Penggunaan jumlah fold terbaik untuk uji validitas, dianjurkan menggunakan *10-fold cross validation*.



Gambar 2.4 Proses *splitting* pada 10-Cross Validation (sumber : Berrar, 2018)

## 2.7 Evaluasi Model

Model klasifikasi yang dihasilkan dari berbagai metode diharapkan mampu mengklasifikasikan seluruh data dengan benar, namun hampir tidak mungkin suatu sistem klasifikasi memiliki kinerja akurat dan presisi. Pengujian performansi dari model klasifikasi dapat diukur melalui tabel matriks konfusi (*confusion matrix*) yaitu tabulasi silang antara data variabel respon yang masuk ke dalam kelas prediksi dan observasi (Kuhn dan Johnson, 2013). Pada kasus dengan 2 kelas, matriks konfusi yang terbentuk dapat dilihat pada Tabel 2.2.

Tabel 2.2 Matriks konfusi pada klasifikasi 2 kelas

		Prediksi	
		<i>Positive</i>	<i>Negative</i>
Aktual	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	<i>Negative</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Kinerja suatu model klasifikasi dapat diukur melalui tiga nilai yaitu akurasi, F1-Score, dan *Area Under Curve* (AUC). Akurasi dapat didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual (Kristina, 2020). F1-Score merupakan perbandingan rata-rata presisi dan sensitivitas (*recall*) yang dibobotkan. Akurasi, presisi, sensitivitas dan F1-Score dapat dihitung dengan persamaan :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.23)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2.24)$$

$$\text{Sensitivitas / Recall} = \frac{TP}{TP+FN}$$

$$\text{F1-Score} = 2 \frac{\text{Sensitivitas} \times \text{Presisi}}{\text{Sensitivitas} + \text{Presisi}} \times 100\% \quad (2.25)$$

Nilai *Area Under Curve* (AUC) merupakan luas area yang berada di bawah kurva *Receiver Operating Characteristics* (ROC) yang menunjukkan tingkat keakuratan diskriminatif dengan memperkirakan probabilitas output dari sampel acak (Qadrini, dkk., 2021). ROC digunakan untuk mengukur keakuratan model klasifikasi dalam menentukan *threshold*. Nilai AUC selalu berada diantara 0 dan 1



dimana semakin besar nilai AUC maka semakin akurat model klasifikasi yang digunakan.

## 2.8 *Shapley Value dan Shapley Additive Explanations (SHAP)*

Pada tahun 1953, Lyold Shapley memperkenalkan *Shapley Value* yang mendeskripsikan hadiah (*payoff*) dari setiap pemain pada *coalitional games* secara proposional dengan mempertimbangkan kontribusi pemain terhadap nilai yang dikumpulkan. Pemain bekerjasama dalam sebuah koalisi dan menerima keuntungan tertentu dari kerjasama tersebut. Pemain diibaratkan sebagai variabel penjelas, sedangkan hadiah adalah nilai dugaan. Nilai Shapley diperoleh dari konsep permainan tersebut. Semua kemungkinan kombinasi dari pemain atau variabel penjelas harus dievaluasi dengan dan tanpa variabel penjelas ke-*i* untuk menghitung nilai Shapley. Menurut Shapley (1953), sebuah permainan koalisi adalah  $\{N, v\}$  dengan *payoff* yang dapat ditransfer dan terdiri dari :

1. Sebuah set pemain  $N$  yang bersifat terbatas (*finite*)
2. Sebuah fungsi  $v$  yang merupakan *payoff* bilangan riil dari koalisi  $S$  dan merupakan semua kemungkinan koalisi dari  $N$ .

Terdapat tiga aksioma yang menyusun *Shapley Value* yang berlandaskan pada definisi keadilan. Aksioma *Shapley Value* diformulasikan secara matematis sebagai berikut.

1. Efisiensi

Total hadiah terdistribusikan ke semua pemain :

$$\sum_{i=0}^N \varphi_i = v(N) \quad (2.26)$$

2. *Symmetry*

Menyatakan bahwa untuk setiap  $v$ , jika  $i$  dan  $j$  bersifat *interchangeable* maka  $\varphi_i(N, v) = \varphi_j(N, v)$ . Artinya jika pemain  $i$  dan  $j$  dapat diganti dan menghasilkan nilai kontribusi yang sama, maka nilai akhir yang diperoleh oleh  $i$  harus sama dengan  $j$

3. *Dummy Player*

Menyatakan bahwa untuk setiap  $v$ , jika  $i$  adalah *dummy player* maka  $\varphi_i(N, v) = 0$ . *Dummy player* seharusnya tidak mendapat bagian dari hasil koalisi.

Ketika pemain  $i$  tidak memberi nilai kontribusi, pemain tersebut tidak mendapat bagian hasil koalisi.

#### 4. Additivity

Menyatakan bahwa untuk setiap nilai permainan yang dapat dipecah menjadi dua pemain independen  $v_1$  dan  $v_2$ ,  $\varphi_i(N, v_1 + v_2) = \varphi_i(N, v_1) + \varphi_i(N, v_2)$ . Hal ini berarti jika sebuah permainan dapat dipisah menjadi dua bagian maka kita dapat memecah bagian hasil koalisi tersebut pada setiap permainan.

Dari ketiga aksioma tersebut nilai yang memenuhi kriteria ketiga aksioma di atas disebut shapley value. Nilai ini didefinisikan dengan persamaan (2.7)

$$\varphi_i(N, v) = \varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(N-|S|-1)!}{N!} (v(S \cup \{i\}) - v(S)), i = 1, \dots, N \quad (2.27)$$

dimana  $\varphi_i(N, v)$  adalah nilai shapley pemain ke- $i$  dan  $S$  adalah subset dari jumlah pemain  $v$  adalah nilai dari fungsi karakteristik setiap pemain dan  $N$  adalah jumlah pemain.

*Shapley Additive Explanations* (SHAP) adalah metode interpretasi prediksi model machine learning yang berbasis teori permainan optimal melalui nilai Shapley (Lundberg dan Lee, 2017). Ide utama SHAP adalah menghitung nilai Shapley untuk setiap variabel sampel yang akan diinterpretasikan, dimana setiap nilai shapley mewakili dampak yang dihasilkan variabel terkait dalam prediksi. Berdasarkan beberapa persamaan untuk menentukan kontribusi masing-masing variabel, nilai shapley pada SHAP diperoleh sebagai berikut :

$$\varphi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(N-|S|-1)!}{N!} (v(S \cup \{j\}) - v(S)) \quad (2.28)$$

dimana  $\varphi_i \in \mathbb{R}$  adalah nilai shapley variabel ke- $j$  dan  $S$  adalah subset dari jumlah prediktor,  $v(S)$  adalah nilai dari fungsi karakteristik setiap prediktor dan  $N$  adalah jumlah prediktor. Fungsi linier pada fungsi  $g(z')$  dapat diperoleh berdasarkan metode penambahan sebagai berikut :

$$g(z') = \varphi_0 + \sum_{i=1}^N \varphi_i z'_i \quad (2.29)$$

dimana  $g(z')$  adalah model interpretasi,  $z' \in \{0,1\}^N$  nilai 1 jika variabel ikut diamati, dan nilai 0 jika variabel tidak ikut diamati.  $N$  adalah jumlah variabel yang digunakan dalam pembuatan model,  $\varphi_0$  adalah nilai basis atau basis value dari suatu

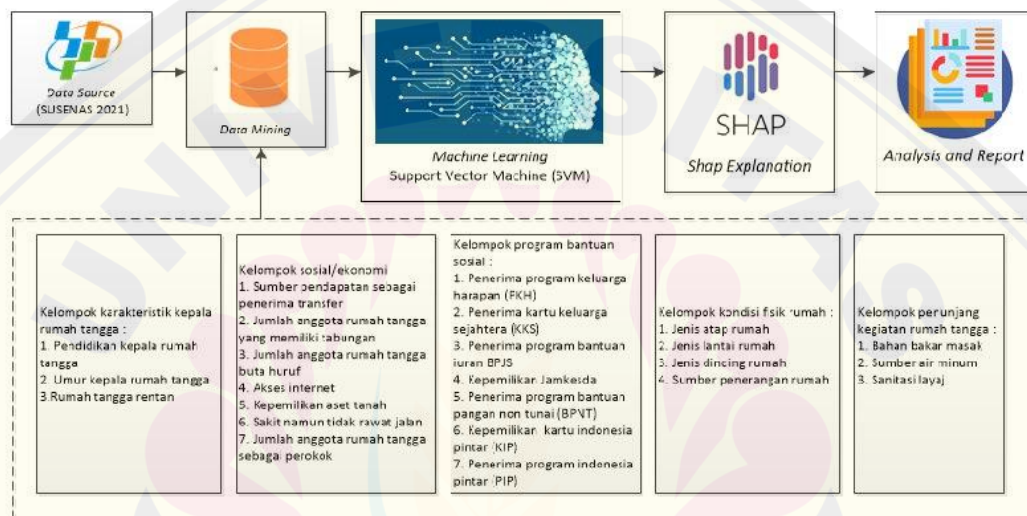
model klasifikasi, dan  $\varphi_j \in \mathbb{R}$  adalah nilai Shapley. Salah satu jenis SHAP yaitu *Permutation Explainer* (Dataman, 2019). *Permutation Explainer* adalah model-agnostik, sehingga dapat menghitung nilai Shapley untuk model apa pun. Model ini bekerja dengan mengulangi permutasi lengkap dari fitur maju dan mundur.



**BAB 3. METODOLOGI PENELITIAN**

**3.1. Sumber Data**

Data yang digunakan dalam penelitian ini adalah data Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2021 Provinsi Jawa Timur. Data tersebut merupakan data sekunder yang diperoleh dari Sistem Informasi Layanan Statistik (SILaStik). Pengumpulan data dilakukan sesuai dengan *framework* data penelitian yang terangkum pada Gambar 3.1



Gambar 3.1 Kerangka kerja persiapan data penelitian

**3.2. Data Penelitian**

Data yang digunakan pada penelitian ini mencakup 24.792 rumah tangga sampel. Variabel respon yang digunakan adalah status kerawanan pangan yang diukur melalui *Food Insecurity Experience Scale* (FIES) pada SUSENAS. Variabel penjelas pada penelitian ini terdiri atas kelompok karakteristik kepala rumah tangga, kelompok sosial ekonomi, kelompok program perlindungan sosial, kelompok kondisi fisik rumah, dan kelompok penunjang kegiatan rumah tangga. Beberapa variabel yang digunakan dapat dilihat pada Tabel 3.1

Tabel 3.1 Penjelasan Variabel Penelitian

Kelompok	Variabel	Skala	Kategori
Karakteristik Kepala Rumah Tangga (KRT)	Pendidikan Kepala Rumah Tangga	Ordinal	Tidak Tamat SD SD SMP SMA Perguruan Tinggi
	Rumah Tangga Rentan	Nominal	Tidak Ya
Sosial/ Ekonomi	Sumber pendapatan sebagai penerima transfer	Nominal	Tidak Ya
	Kepemilikan aset tanah	Nominal	Tidak Ya
	Akses internet	Nominal	Tidak Ya
	Sakit namun tidak rawat jalan	Nominal	Tidak Ya
	Jumlah anggota rumah tangga yang memiliki tabungan	Nominal	Tidak ada 1 2 atau lebih
	Jumlah anggota rumah tangga buta huruf	Nominal	Tidak ada Ada
Program Bantuan Sosial	Penerima Program Keluarga Harapan (PKH)	Nominal	Tidak Ya
	Penerima Kartu Keluarga Sejahtera (KKS)	Nominal	Tidak Ya
	Penerima Bantuan Pemda	Nominal	Tidak Ya
	Penerima Bantuan Pangan Non Tunai (BPNT)	Nominal	Tidak Ya
	Penerima Program Bantuan Iuran (PBI) BPJS	Nominal	Tidak Ya
	Kepemilikan Jamkesda	Nominal	Tidak Ya
	Kepemilikan KIP	Nominal	Tidak Ya
Kondisi Fisik Rumah	Jenis atap rumah	Ordinal	Asbes Seng Beton Genteng Lainnya
	Jenis lantai rumah	Ordinal	Ubin/tegel/teraso Semen/bata merah Parquet/vinil/karpet Marmer/keramik Lainnya
	Jenis dinding rumah	Ordinal	Kayu/papan Tembok/plesteran Lainnya
	Sumber penerangan rumah	Ordinal	Non listrik Listrik
	Luas Lantai	Rasio	< 50 50 – 100 101 – 200 > 200
Penunjang Kegiatan Rumah Tangga	Bahan bakar masak	Ordinal	Tidak memasak Kayu bakar Minyak tanah LPG 3Kg LPG 12Kg/5.5Kg/Bluegaz Lainnya
	Sumber air minum	Ordinal	Sumur bor/pompa Leding Air kemasan/isi ulang Lainnya
	Air minum layak	Nominal	Tidak Ya
	Sanitasi layak	Nominal	Tidak Ya

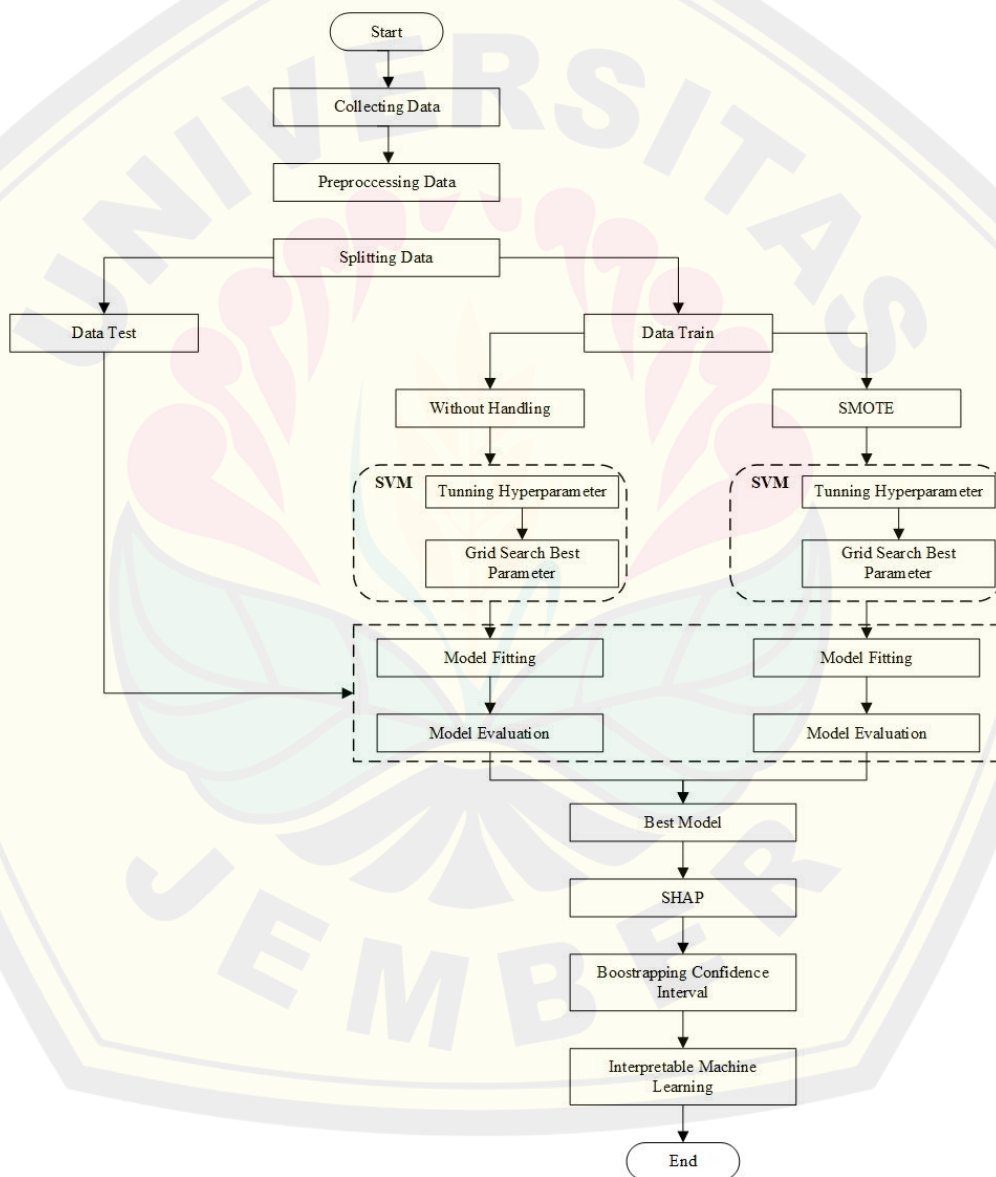
### 3.3. Tahapan Penelitian

Tahapan penelitian dapat dijelaskan melalui langkah-langkah sebagai berikut:

1. Pengumpulan data : secara keseluruhan data penelitian yang digunakan terdiri dari 24 variabel dan 1 variabel kelas kerawanan pangan. Data yang diperoleh dari SILaStik berbentuk database (.dbf). Selanjutnya data difilter berdasarkan nomor urut rumah tangga se-Indonesia.
2. Kontruksi data: membuat struktur data dari Silastik ke dalam bentuk format excel atau csv. Proses ini dilakukan pada tiap komponen variabel sehingga akan diperoleh 24 variabel dan 1 variabel kelas kerawanan pangan.
3. Penentuan variabel : mendefinisikan 24 variabel sebagai variabel penjelas (X) dan 1 variabel kelas kerawanan pangan sebagai variabel respon (y)
4. Deskripsi data : membuat deskripsi data numerik maupun kategorik. melakukan eksplorasi data untuk mendapatkan gambaran umum mengenai variabel-variabel yang akan dianalisis seperti adanya *missing value*, pencilan dan data anomali.
5. *Imbalanced* data : proses menyeimbangkan data pada jumlah kelas data (*instance*) baik secara *oversampling* maupun *undersampling*.
6. *Splitting* data : membagi data variabel penjelas ke dalam data *training* dan *testing*. Data *training* sendiri akan dibagi lagi menjadi data *subtraining* dan *validation*.
7. Membangun model SVM menggunakan data *subtraining* dan *validation* dengan menemukan nilai terbaik dari parameter C dan  $\gamma$ . *Parameter tuning* dilakukan dengan menggunakan GSCV.
8. Melatih model dengan parameter terbaik dengan menggunakan CV. Data yang digunakan adalah data *subtraining* dan *validation* pada poin 7 dengan n split sebanyak 10 kali
9. Menguji model terbaik dengan melakukan prediksi menggunakan data *testing* pada poin 7.
10. Melakukan analisa dan *interpretable* model menggunakan SHAP dengan tahapan sebagai berikut :

- Menghitung nilai Shapley
- Melihat variabel penjelas yang paling tinggi berdasarkan nilai SHAP menggunakan *plot shap summary*
- Melihat hubungan variabel penjelas dengan pengaruhnya terhadap hasil prediksi nilai Shapley menggunakan SHAP *hierarchical clustering*.

Secara umum langkah yang digunakan pada penelitian ini terangkum pada Gambar 3.2

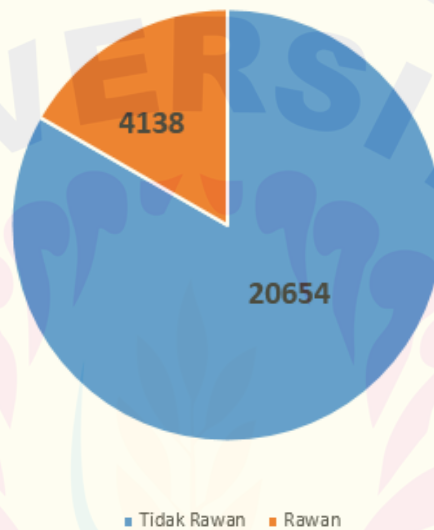


Gambar 3.2 Skema tahapan proses penelitian

## BAB 4. HASIL DAN PEMBAHASAN

### 4.1. Deskripsi Data

Data penelitian diperoleh dari SUSENAS tahun 2021 oleh BPS. Berdasarkan data 32.462 rumah tangga sampel dalam SUSENAS Maret 2021 Provinsi Jawa Timur, sebanyak 16 rumah tangga menyatakan tidak tahu (kode 8) atau menolak menjawab (kode 9) setidaknya satu dari 8 pertanyaan FIES. Selain itu terdapat 7.654 nilai data yang hilang, sehingga sejumlah 24.792 rumah tangga yang akan dianalisis dalam penelitian ini.



Gambar 4.1 Perbandingan proporsi status rawan pada seluruh rumah tangga di Provinsi Jawa Timur

Proporsi rumah tangga rawan pangan dan tidak rawan pangan ditunjukkan oleh gambar 4.1 Jumlah rumah tangga tidak rawan pangan pada seluruh rumah tangga di Jawa Timur sebanyak 83,31% rumah tangga, sedangkan rumah tangga rawan pangan sebanyak 16,69%. Data penelitian ini dibagi menjadi 5 kelompok dengan unit dasar rumah tangga dan unit dasar individu. Jumlah variabel prediktor yang berasal dari kelompok karakteristik kepala rumah tangga terdiri dari 2 variabel diantaranya pendidikan kepala rumah tangga dan rumah tangga rawan/rentan. Jumlah variabel prediktor yang berasal dari kelompok sosial/ekonomi terdiri dari 6 variabel diantaranya sumber pendapatan sebagai penerima transfer, kepemilikan



aset tanah, akses internet, sakit namun tidak rawat jalan, jumlah anggota rumah tangga yang memiliki tabungan, jumlah anggota rumah tangga buta huruf. Jumlah variabel prediktor yang berasal dari kelompok program bantuan sosial terdiri dari 7 variabel diantaranya PKH, KKS, Penerima Bantuan Pemda, BPNT, Bantuan Iuran BPJS, Kepemilikan Jamkesda, Kepemilikan KIP. Jumlah variabel prediktor yang berasal dari kelompok kondisi fisik rumah terdiri dari 5 variabel diantaranya jenis atap rumah, jenis lantai rumah, jenis dinding rumah, sumber penerangan rumah dan luas lantai. Empat variabel prediktor lainnya adalah kelompok penunjang kegiatan rumah tangga yang meliputi bahan bakar masak, sumber air minum, air minum layak dan sanitasi layak.

## 4.2. *Preprocessing* Data Sebelum Pemodelan

### 4.2.1. Persiapan Data

Sebelum melakukan pemodelan klasifikasi menggunakan SVM, dilakukan pembagian data menjadi data latih dan data uji. Data latih digunakan untuk membentuk model klasifikasi sedangkan data uji digunakan untuk mengukur kinerja dari model yang telah didapatkan. Data dibagi dengan proporsi 70% untuk data latih dan 30% untuk data uji. Berdasarkan Tabel 4.1 dapat diketahui dari total data set yang ada, pembagian data sebanyak 17.354 data latih dan 7.438 untuk data uji. Pembagian data latih dan data uji pada kumpulan data penelitian dilakukan secara acak.

Tabel 4.1 Proporsi awal data latih dan data uji

Kategori	Provinsi Jawa Timur	
	Data latih	Data uji
Proporsi	70%	30%
Jumlah	17.354	7.438
Rawan	2.845	1.293
Tidak rawan	14.509	6.147

Data latih yang telah didapatkan selanjutnya digunakan untuk membentuk model pembelajaran mesin SVM. Jumlah persentase data rawan pangan pada data latih tidak seimbang, diketahui kelas terbanyak terdapat pada kelompok rumah tangga tidak rawan pangan (lihat pada Tabel 4.1). Ketidakseimbangan distribusi kelas data pada variabel respon ini yang disebut dengan data tidak seimbang.

Permasalahan data tidak seimbang perlu dilakukan penanganan lebih lanjut karena dapat mengakibatkan kecilnya akurasi pada kelas minoritas, padahal kelas data minoritas merupakan kelas yang akan diteliti lebih lanjut.

Tabel 4.2 Jumlah data latih awal dan data latih yang sudah diseimbangkan menggunakan teknik SMOTE

Kategori	Data latih		Data SMOTE	
	Observasi	%	Observasi	%
Rawan	2.845	16,40	14.509	50,00
Tidak rawan	14.509	83,60	14.509	50,00

Penyeimbangan data menggunakan metode *Syntetic minority over-sampling technique* (SMOTE) yang dilakukan pada kelas minoritas yaitu kelompok rumah tangga rawan pangan. Modifikasi dilakukan dengan menggunakan informasi lima tetangga terdekat ( $k=5$ ) dan pengambilan sampel dilakukan menggunakan sampling strategi sebesar 1 (satu) atau kelas minoritas akan menjadi sama banyak dengan kelas mayoritas pada data latih baru. Jumlah data latih yang sudah diseimbangkan dengan teknik SMOTE dapat dilihat pada Tabel 4.2 data latih yang sudah diseimbangkan selanjutnya digunakan untuk membangun model pembelajaran mesin *Support Vector Machine* (SVM).

#### 4.2.2. *Weight of Evidence and Information Value*

*Weight of Evidenve* (WoE) *and Information Value* (IV) memberikan kerangka kerja yang bagus untuk eksplorasi data, penyaringan variabel, dan transformasi variabel dalam klasifikasi biner. Nilai informasi dapat membantu mengukur kekuatan prediksi dalam memisahkan kelas biner. Ukuran kekuatan variabel prediktor dapat dilihat pada Tabel 4.3

Tabel 4.3 Nilai informasi variabel prediktor

Nilai informasi	Kekuatan prediksi
$X_i \leq 0,02$	Tidak berguna untuk prediksi
$0,02 < X_i \leq 0,1$	Daya prediksi lemah
$0,1 < X_i \leq 0,3$	Daya prediksi sedang
$0,3 < X_i \leq 0,5$	Daya prediksi kuat
$0,5 < X_i$	Daya prediksi mencurigakan

Hasil penelitian yang telah dilakukan oleh Fitria, dkk. (2023) variabel prediktor pada data kerawanan pangan Provinsi Jawa Timur tahun 2021 memiliki daya prediksi sedang dalam pengklasifikasian. Hal ini diperkuat dengan nilai informasi pada Tabel 4.4

Tabel 4.4 Nilai informasi variabel prediktor pada data kerawanan pangan

Provinsi Jawa Timur		Provinsi Jawa Barat	
Nilai informasi	Variabel	Nilai informasi	Variabel
0,21	Pendidikan KRT	0,38	Pendidikan KRT
0,20	Bahan bakar masak	0,26	Bahan bakar masak
0,20	Jumlah penabung	0,17	Jumlah penabung
0,20	Jenis lantai	0,13	Luas lantai

#### 4.3. Pemodelan Klasifikasi Rawan Pangan Menggunakan SVM

Pemodelan klasifikasi kejadian rawan pangan menggunakan model SVM diterapkan pada data latih yang telah diseimbangkan. Saat melakukan pemodelan, optimasi hiperparameter dilakukan menggunakan proses *10 fold cross validation* untuk memperoleh kestabilan kinerja model dengan melihat nilai parameter optimum. Pada proses *10 fold cross validation*, data latih yang teracak dibagi menjadi 10 bagian. Setiap satu bagian dijadikan data uji dan sisanya dijadikan sebagai data latih. Pencarian hiperparameter menggunakan *grid search optimization*. Hiperparameter yang digunakan dapat dilihat pada Tabel 4.5

Tabel 4.5 Hiperparameter pada model SVM

Model	Kernel	Hiperparameter	Nilai
SVM	Linier	C	0,01 0,1 1 10 20 50 100 500 1000 1500
	polinomial	Degree	3 4
		C	0,01 0,1 1 10 20 50 100 500 1000 1500
	Rbf	Gamma	auto, scale
		C	0,01 0,1 1 10 20 50 100 500 1000 1500
Sigmoid	C	0,01 0,1 1 10 20 50 100 500 1000 1500	

Pencarian hiperparameter didapatkan melalui proses pengujian menyeluruh terhadap semua nilai yang diujikan. Model pembelajaran mesin SVM digunakan terhadap data tanpa SMOTE dan data yang sudah diseimbangkan menggunakan

teknik SMOTE. Hiperparameter terbaik pada masing-masing model dapat dilihat pada Tabel 4.6

Tabel 4.6 Hiperparameter terbaik yang digunakan dalam pembentukan model SVM

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

Hiperparameter optimum yang dihasilkan metode SVM pada data tanpa SMOTE yaitu gamma sebesar 0,04 dan C sebesar 200, sedangkan pada data SMOTE yaitu gamma sebesar 0,04 dan C sebesar 500. Model dengan hiperparameter optimal yang sudah didapatkan selanjutnya dilakukan evaluasi model dilakukan evaluasi untuk memperoleh kebaikan model dengan menghitung nilai akurasi, sensitivitas dan AUC dari masing-masing data. Nilai evaluasi model dilakukan pada data latih dan data uji dari masing-masing data. Klasifikasi rumah rawan pangan pada SVM menggunakan data SMOTE lebih baik dibandingkan dengan nilai evaluasi pada data tanpa SMOTE. Hasil evaluasi model disajikan pada Tabel 4.7

Tabel 4.7 Nilai evaluasi model SVM pada data tanpa SMOTE dan data SMOTE

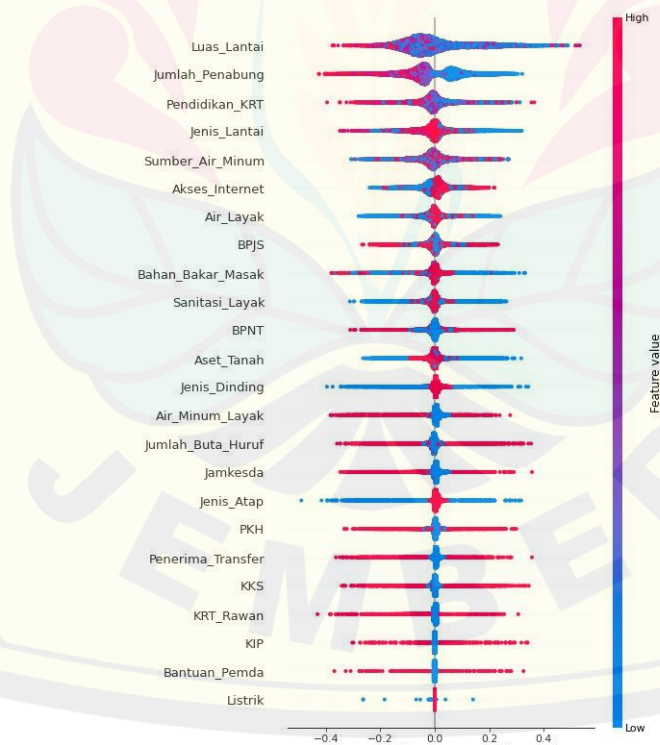
Data	Model	Akurasi	Sensitivitas	AUC
Data uji	Data tanpa SMOTE	72%	24%	0,48
	SMOTE	61%	68%	0,66

Nilai akurasi model SVM sebesar 0,72 untuk data tanpa SMOTE pada data uji dan 0,61 untuk data SMOTE pada data uji. Nilai akurasi ini membuktikan bahwa kemampuan model untuk mendeteksi secara benar seluruh data uji sebesar 72% dan 61%. Nilai sensitivitas perlu diperhitungkan, pemerintah akan terfokus untuk melihat rumah tangga rawan pangan agar dapat memberikan bantuan atau solusi dengan tepat sasaran. Pada penelitian ini, analisis menunjukkan bahwa sensitivitas sebesar 24% untuk data tanpa SMOTE pada data uji dan 68% untuk data SMOTE pada data uji, artinya kesalahan model memprediksi rumah tangga tidak rawan pangan padahal rumah tangga tersebut terkategori rawan pangan adalah sebesar 24% untuk data tanpa SMOTE dan 50% untuk data SMOTE. Nilai AUC juga

mendukung model pembelajaran mesin SVM sebesar 0,48 untuk data tanpa SMOTE pada data uji dan 0,66 untuk data SMOTE pada data uji, maka model ini cukup efektif digunakan dalam analisis identifikasi kejadian rawan pangan Provinsi Jawa Timur.

#### 4.4. Identifikasi variabel penting menggunakan SHAP pada Provinsi Jawa Timur

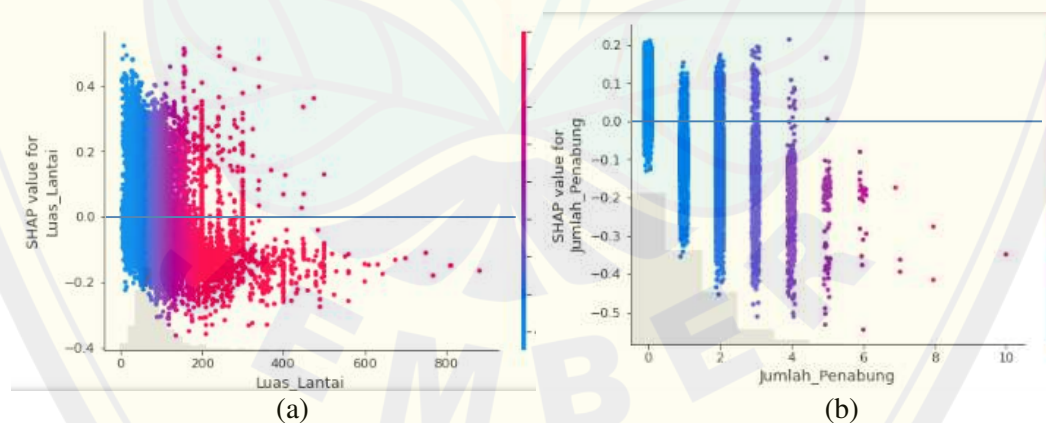
Nilai SHAP merupakan nilai kontribusi pada masing-masing variabel dalam memprediksi kejadian rawan pangan. Nilai SHAP dihitung menggunakan fungsi *Permutation Explainer* yang dapat membantu menjelaskan model SVM. Metode SHAP dapat melihat urutan variabel terpenting melalui fungsi *shap.summary\_plot()*. Urutan variabel penting SHAP berdasarkan pada seberapa mampu variabel memprediksi kejadian rawan pangan pada variabel respon. Variabel yang paling penting akan mampu membedakan prediksi dengan baik antar kategorinya dan berkontribusi terhadap prediksi.



Gambar 4.2 Urutan variabel penting Provinsi Jawa Timur

Pada *summary plot* sumbu  $x$  merupakan nilai SHAP dimana semakin besar nilainya maka semakin besar peluang rumah tangga tersebut mencirikan rawan pangan. Sumbu  $y$  merupakan urutan variabel penting, dimana urutan prediktornya berdasarkan tingkat kontribusi dalam memprediksi kejadian rawan pangan berdasarkan nilai SHAP yang dihasilkan dan setiap titik pada plot merupakan objek atau rumah tangga. Gradasi warna biru menuju merah pada nilai variabel menunjukkan nilai yang bergerak dari rendah menuju tinggi. Nilai SHAP dari setiap variabel dapat dilihat pada Lampiran 3. Urutan variabel penting yang mencirikan kejadian rawan pangan dapat dilihat pada Gambar 4.2.

Pada Gambar 4.2 Plot SHAP dari model SVM menunjukkan bahwa luas lantai, jumlah penabung, dan pendidikan kepala rumah tangga menjadi variabel penting dengan peringkat yang berada tiga besar. Hal ini menunjukkan bahwa faktor luas lantai, jumlah penabung, dan pendidikan kepala rumah tangga memiliki peluang besar mencirikan terjadinya kejadian rawan pangan di Provinsi Jawa Timur. Peubah tertinggi selanjutnya adalah jenis lantai dan sumber air minum. Untuk melihat distribusi nilai SHAP setiap kategori pada peubah penting selanjutnya akan dilihat dalam bentuk *scatter plot*. Sumbu  $x$  pada *scatter plot* merupakan nilai atau kategori dari variabel, sedangkan sumbu  $y$  merupakan nilai SHAP yang diperoleh variabel tersebut.



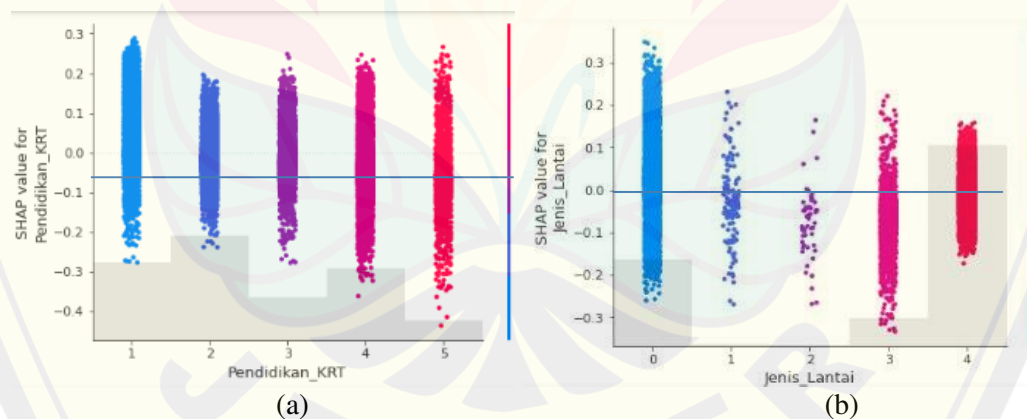
Gambar 4.3 (a) *Scatter plot* nilai SHAP luas lantai Provinsi Jawa Timur

(b) *Scatter plot* nilai SHAP jumlah penabung Provinsi Jawa Timur

Variabel penting urutan pertama yang mencirikan kejadian rawan pangan Provinsi Jawa Timur yaitu luas lantai. Sebaran luas lantai terhadap nilai SHAP pada Gambar 4.3 (a) terlihat bahwa ukuran luas lantai yang besar lebih mendominasi area

SHAP negatif. Hal ini berarti semakin besar luas lantai yang dimiliki oleh rumah tangga tersebut maka semakin kecil peluang rumah tangga terprediksi sebagai rawan pangan. Namun, luas lantai besar terlihat juga berada pada nilai SHAP positif. luas lantai besar terdapat pada nilai SHAP positif dan luas lantai kecil berada pada nilai SHAP negatif salah satunya terjadi ketika rumah tangga di pedesaan yang memiliki luas lantai besar tetapi cenderung tergolong rawan pangan, ataupun ketika rumah tangga perkotaan memiliki luas lantai kecil tetapi rumah tangga tersebut cenderung tidak tergolong rawan pangan.

Variabel penting urutan kedua yang mencirikan kejadian rawan pangan Provinsi Jawa Timur yaitu jumlah penabung. Gambar 4.3 (b) menunjukkan bahwa semakin sedikit jumlah penabung pada suatu rumah tangga maka semakin besar peluang rumah tangga tersebut terprediksi sebagai rumah tangga rawan pangan. Rumah tangga tidak memiliki penabung memiliki distribusi paling banyak dan cenderung teridentifikasi sebagai rawan pangan. Hasil ini juga sejalan dengan penelitian sebelumnya, bahwa banyaknya anggota rumah tangga yang memiliki tabungan di lembaga keuangan menjadi salah satu faktor yang mencirikan kejadian rawan pangan pada rumah tangga (Endang, dkk. 2022).

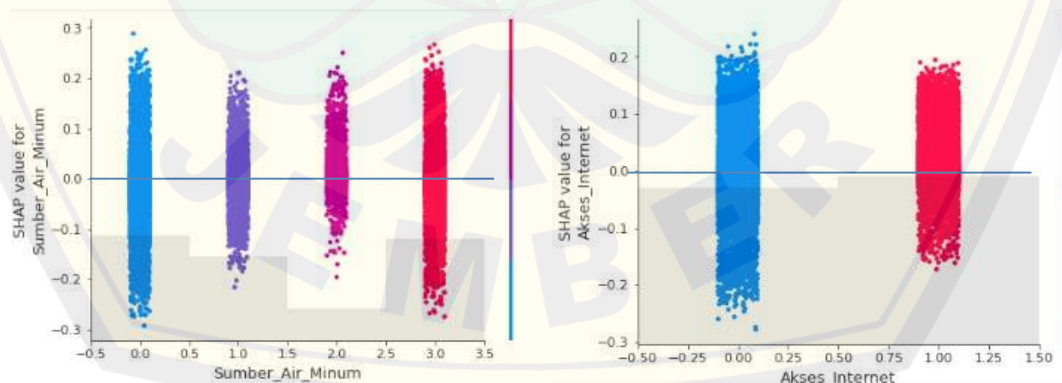


Gambar 4.4 (a) *Scatter plot* nilai SHAP pendidikan kepala rumah tangga Provinsi Jawa Timur, (b) *Scatter plot* nilai SHAP jumlah penabung Provinsi Jawa Timur

Pendidikan kepala rumah tangga menjadi variabel penting urutan ketiga yang mencirikan kejadian rawan pangan di Provinsi Jawa Timur. Terlihat baik pada Gambar 4.4 (a) semakin rendah pendidikan kepala keluarga maka semakin tinggi

peluang rumah tangga tersebut terindikasi rawan pangan. Pendidikan rumah tangga dengan tingkat pendidikan SD dan SMP memiliki nilai SHAP yang seimbang, artinya terdapat rumah tangga yang tergolong rawan pangan maupun tidak rawan pangan. Kepala rumah tangga dengan tingkat pendidikan SMA dan perguruan tinggi nilai SHAP mendominasi pada daerah negatif, hal itu berarti rumah tangga tersebut berpeluang besar untuk tidak tergolong rawan pangan. Hasil ini sejalan dengan penelitian sebelumnya yaitu pendidikan kepala rumah tangga yang memiliki pendidikan rendah merupakan faktor yang mencirikan suatu rumah tangga tergolong rawan pangan (Dharmawan, dkk. 2022)

Variabel penting urutan keempat adalah jenis lantai. Gambar 4.4 (b) menunjukkan bahwa sebagian besar titik-titik merah berada pada nilai SHAP negatif yang menunjukkan bahwa rumah tangga dengan jenis lantai yang bagus/tinggi, memiliki peluang kecil mengalami rawan pangan. Hal ini dapat menunjukkan bahwa kategori lainnya merupakan jenis lantai yang tidak bagus, karena pemberian nilai untuk kategori lainnya adalah 0. Titik-titik berwarna biru menyebar di sepanjang nilai SHAP hal ini dapat menjadi dugaan bahwa jenis lantai yang rendah yang digunakan rumah tangga belum tentu tergolong ke dalam rumah tangga rawan pangan. Namun, hasil analisis ini masih dapat disimpulkan bahwa semakin rendah kualitas jenis lantai yang digunakan rumah tangga, maka nilai SHAP semakin besar. Hal ini berarti bahwa semakin besar rumah tangga tersebut terprediksi sebagai rumah tangga rawan pangan.



Gambar 4.5 (a) *Scatter plot* nilai SHAP sumber air minum Provinsi Jawa Timur,

(b) *Scatter plot* nilai SHAP akses internet Provinsi Jawa Timur



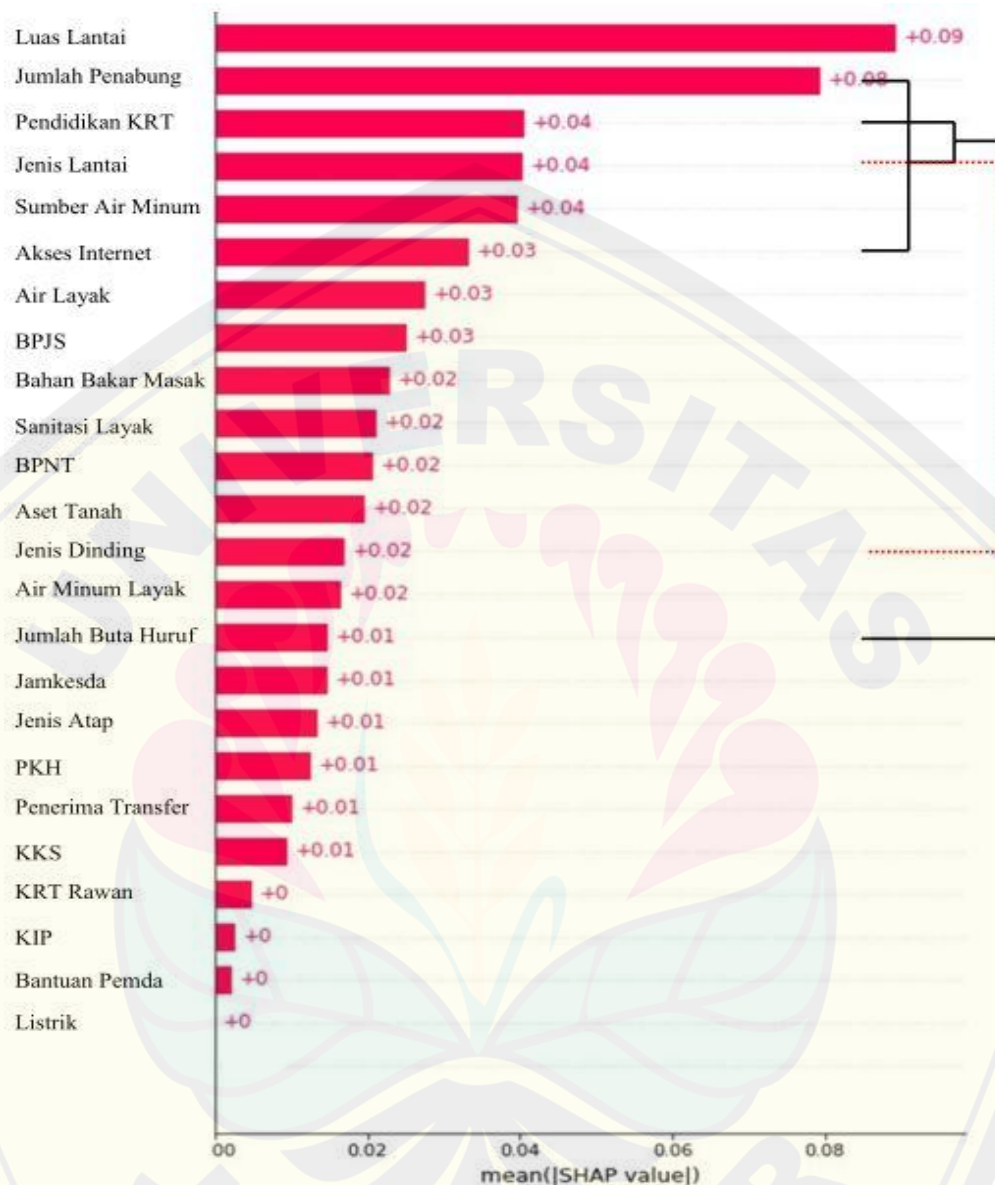
Variabel penting urutan kelima adalah sumber air minum. Pada Gambar 4.5 (a) menunjukkan bahwa sebagian besar titik-titik merah nilai SHAP negatif yang menunjukkan bahwa rumah tangga dengan sumber air minum dari air kemasan atau isi ulang, memiliki peluang kecil untuk terprediksi sebagai rawan pangan. Titik-titik berwarna ungu berada pada nilai SHAP positif yang menunjukkan bahwa rumah tangga dengan sumber air minum dari leding maupun sumur bor atau pompa semakin besar resiko mengalami rawan pangan. Titik-titik berwarna biru menyebar di sepanjang nilai SHAP dan mengumpul di sekitar nilai 0. Artinya sumber air minum dari sumur bor atau pompa hampir tidak memiliki *impact* dalam memprediksi rumah tangga rawan pangan atau tidak.

Variabel penting urutan keenam adalah akses internet. Gambar 4.5 (b) menunjukkan bahwa sebagian besar titik-titik merah berada pada nilai SHAP positif. Ini menunjukkan bahwa banyak rumah tangga yang memiliki fasilitas akses internet memiliki peluang besar terprediksi sebagai rawan pangan. Banyaknya rumah tangga yang memiliki fasilitas akses internet namun terprediksi ke dalam rawan pangan dikarenakan proporsi datanya cukup besar pada kelas rawan pangan. Pada dasarnya perkembangan teknologi saat ini mampu memaksa setiap orang untuk mengikutinya, sehingga kebutuhan fasilitas internet menjadi kebutuhan yang harus terpenuhi. Gambar 4.2 terlihat jelas bahwa titik-titik berwarna biru menyebar di sepanjang nilai SHAP hal ini dapat menjadi dugaan bahwa rumah tangga yang tidak memiliki fasilitas akses internet dapat tergolong ke dalam rumah tangga tidak rawan pangan. Pembiayaan besar rumah tangga dialihkan pada tabungan di lembaga keuangan.

#### **4.5. Identifikasi interaksi antar variabel menggunakan SHAP berdasarkan *hierarchical clustering***

Interaksi antar variabel dilakukan dengan menghitung nilai kekuatan penjelasannya (sebagaimana diukur dengan  $R^2$ ) antara satu dengan yang lainnya. Fungsi yang digunakan adalah *shap.utils.hclust()*. Pada fungsi ini *clustering cutoff* yang digunakan sebesar 0,90, artinya lebih dari 90 persen kekuatan penjelasannya

antara variabel satu dengan variabel lainnya sehingga semua bagian lain dari pohon pengelompokan akan dihapus secara *default* jika tidak memenuhi syarat.



Gambar 4.6 *Hierarchical Clustering* pada Provinsi Jawa Timur

Terlihat baik pada Gambar 4.6 bahwa pembentukan kluster berdasarkan nilai kekuatan penjelasannya adalah 2 kluster, dimana kluster pertama terbentuk oleh variabel jumlah penabung, akses internet, pendidikan kepala rumah tangga, dan jumlah buta huruf. Pada kluster kedua terbentuk oleh variabel jenis lantai dan jenis dinding. Variabel jumlah penabung dan akses internet memiliki kedekatan yang tinggi, hal ini juga menunjukkan bahwa pendugaan rawan pangan pada rumah

tangga jika dilihat dari jumlah penabung maka dapat dilihat juga dari pengeluaran anggota rumah tangga dalam pembiayaan fasilitas akses internet. Selain itu, jenis lantai dan jenis dinding memiliki kedekatan yang tinggi, artinya jika pendugaan rawan pangan pada rumah tangga dilihat dari jenis lantai maka dapat dilihat juga dari jenis dinding yang digunakan.



## BAB 5. PENUTUP

### 5.1. Kesimpulan

Berdasarkan hasil analisis data dan pembahasan maka diperoleh kesimpulan sebagai berikut :

1. Nilai evaluasi model pada data kejadian rawan pangan Provinsi Jawa Timur memperlihatkan bahwa, model SVM mampu bekerja cukup baik dalam mengklasifikasikan rumah tangga rawan pangan. Parameter optimum diperoleh pada kernel *rbf*. SHAP dengan fungsi interpretasi *Permutation Explainer* mampu menjelaskan model SVM Nonlinier dengan baik.
2. Interpretasi model SVM Nonlinier menggunakan SHAP menghasilkan urutan variabel penting penciri kejadian rawan pangan Provinsi Jawa Timur yaitu luas lantai yang kecil, rumah tangga tidak memiliki penabung, pendidikan terakhir kepala rumah tangga yang rendah yaitu tidak tamat SD, jenis lantai yang digunakan berupa semen/bata merah, sumber air minum berasal dari sumur bor/pompa dan tidak memiliki akses internet.
3. Analisis interaksi SHAP menghasilkan 2 kluster. Kluster pertama terbentuk oleh variabel jumlah penabung, akses internet, pendidikan kepala rumah tangga, dan jumlah buta huruf. Kluster kedua terbentuk oleh variabel jenis lantai dan jenis dinding. Variabel jumlah penabung dan akses internet memiliki kedekatan yang tinggi pada kluster pertama, sedangkan pada kluster kedua variabel jenis lantai dan jenis dinding memiliki kedekatan yang tinggi.

### 5.2. Saran

Pada penelitian ini pembentukan model klasifikasi hanya menggunakan algoritma *support vector machine*, penelitian selanjutnya dapat menambahkan metode machine learning lainnya sebagai pembandingan. Selain itu pada penelitian ini tidak menggunakan seleksi fitur untuk memperkecil jumlah variabel yang digunakan, disarankan pada penelitian selanjutnya dapat menggunakan seleksi fitur seperti *Feature Ablation*, *Shrinkage*, *Principal Components Analysis*, dan metode lainnya.

## DAFTAR PUSTAKA

- Badan Ketahanan Pangan. 2013. *Petunjuk Pelaksanaan Sistem Kewaspadaan Pangan dan Gizi*. Badan Ketahanan Pangan, Kementerian Pertanian. Jakarta
- 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.
- Berrar D. 2018. Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*. Vol. 1, Elsevier, pp. 542–545.
- Cholissodin, I., Sutrisno, S., Soebroto, AA., Hasanah, U. dan Febiola, YI.. 2020. *AI : Machine Learning & Deep Learning*. Fakultas Ilmu Komputer. Malang: Universitas Brawijaya.
- Cortes, C. dan Vapnik, V. 1995. Support-Vector Networks. *Machine Learning*. 20. 273-297.
- Dataman. 2019. Explain Your Model With the SHAP Values. <https://towardsdatascience.com/explain-any-models-with-the-shap-values-use-the-kernelexplainer-79de9464897a> [diakses pada 12 Maret 2022]
- Endang Yuliani, Bagus Sartono, Hari Wijayanto, Alfian Futuhul Hadi, and Evi Ramadhani. 2022. "Study of features importance level identification of machine learning classification model in sub-populations for food insecurity", AIP Conference Proceedings 2668, 070003. <https://doi.org/10.1063/5.0111784>
- Fitria, Z., Saputra, M. D., Sartono, B., Ramadhani, E., dan Hadi, A.F. 2023. Weight of Evidence and Information Value on Support Vector Machine Classifier. *Advance In Intelligent System Research*. (in press)
- Futoma J, Morris J, dan Lucas J. 2015. A comparison of models for predicting early hospital readmissions. *J Biomed Inform*. 56:229–38.
- G. Qiong., Xian-Ming, W., Zhao, W., Bing, N., dan Chung-Sheng, X. 2016. An Improved SMOTE Algorithm Based on Genetic Algorithm for Imbalanced data classification. *Digital Inf. Management*. Vol. 142: 92 - 103

- Gu, X.-M. Wang, Z. Wu, B. Ning, dan C.-S. Xin. 2016. An improved SMOTE algorithm based on genetic algorithm for imbalanced data classification. *J Dig Inf Manag.* vol.14, no. 2, pp. 92–103.
- Dharmawan, H., Sartono, B., Kurnia, A., Hadi, A. F., Ramadhani, E. 2022. A study of machine learning algorithms to measure the feature importance in class-imbalance data of food insecurity cases in Indonesia, *Commun. Math. Biol. Neurosci.* Article ID 101
- H. Sahlaoui et al. 2021. Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanation. *IEEE Access.* Vol. 9 No. 152688-152703
- Hamel, L. 2009. *Knowledge Discovery with Support Vector Machine.* JohnWiley & Sons, Inc., Hoboken, New Jersey.
- Hammer, B. dan Gersmann, K. 2003. A note on the universal approximation capability of support vector machines. *Neural Process Lett* 17(1):43–53
- Iman, Q. dan Wijayanto, A.W. 2021. Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode Random Forest dan Support Vector Machine. *Jurnal Sistem dan Teknologi Informasi.* Vol. 09, No. 2. Pp. 178-184
- Kang, S. dan Cho, S. 2014. Approximating support vector machine with artificial neural network for fast prediction. *Expert Syst Appl* 41(10):4989–4995
- Kementerian\_PPN/Bappenas. 2017. *Metadata Indikator Pilar Pembangunan Sosial.* Jakarta: Kementerian Perencanaan Pembangunan Nasional / BAPPENAS.
- Kristina , B. C., et al. 2020. The visualization and classification method of support vector machine in lymphoma cancer. *J. Phys.: Conf. Ser.* 1613 012065 Vol. 161
- Kuhn, M., dan Johnson, K. 2013. *Applied Predictive Modeling.* New York: Springer. <http://dx.doi.org/10.1007/978-1-4614-6849-3>
- Lundberg SM, Lee SI. 2017. A unified approach to interpreting model predictions. Di dalam: *Advances in Neural Information Processing Systems.* Volume ke-2017-December: 4768–4777. arXiv:1705.07874v2
- Meirista, E. 2015. Aplikasi metode Support Vector Machine (SVM) untuk klasifikasi tanaman berdaun menjari dan gulma berdasarkan fitur bentuk dan tekstur daun. *Masters thesis.* Surabaya : Institut Teknologi Sepuluh Nopember.

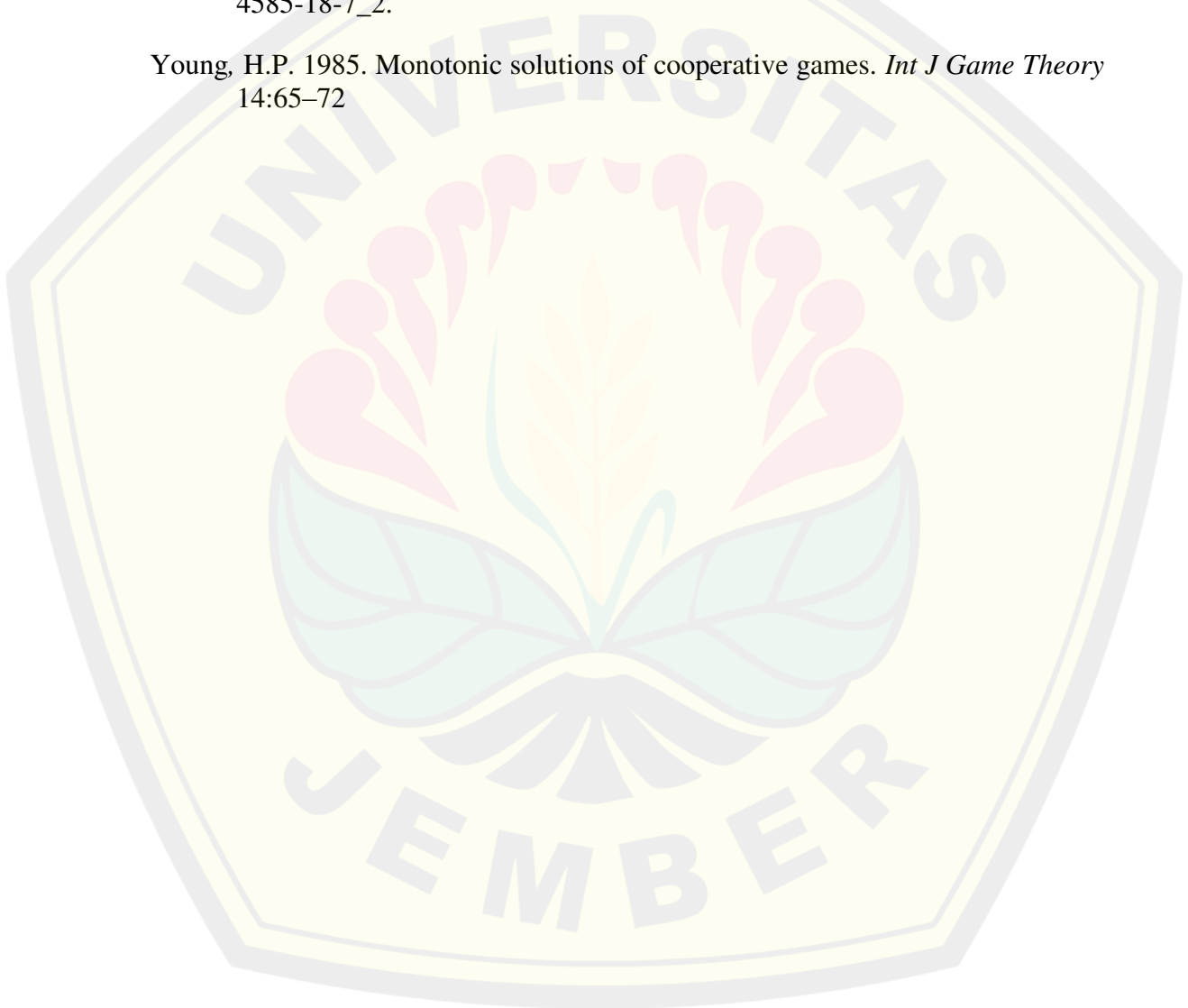
- Navia, R. L. 2015. Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *Jurnal Inovtek POLBENG – Seri Informatika*. 1(2), 126-132.
- Nugroho, A. S., A. B. Witarto, dan H. Dwi. 2003. *Support Vector Machine, Teori dan Aplikasinya dalam Bioinformatika*. <http://www.ilmukomputer.com> [diakses pada 12 Maret 2022]
- Osborne, M.J. dan Rubinstein, A. 1994. *A course in game theory*. The MIT Press, Cambridge, MA
- Peng, Y., dan Yao, J. 2010. AdaOUBoost: Adaptive Over-sampling and Under-sampling to Boost the Concept Learning in Large Scale Imbalanced Data Sets. *Proceedings of the international conference on Multimedia information retrieval* pp. 111-118. Philadelphia, Pennsylvania, USA: ACM.
- Putra, J.W.G. 2020. *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning* edisi 1.4. Tokyo, Jepang.
- Qadrini, L., Sepewali, A. dan Aina A. 2021. Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial. *Jurnal Inovasi Penelitian*. Vol. 2, No. 7. pp. 1959-1966
- Refaeilzadeh P., Tang L., dan Liu H. 2016. Cross-Validation. In: Liu L., Özsu M. (eds) *Encyclopedia of Database Systems*. New York: Springer. [https://doi.org/10.1007/978-1-4899-7993-3\\_565-2](https://doi.org/10.1007/978-1-4899-7993-3_565-2)
- Ribeiro MT, Singh S, Guestrin C. 2016. “Why should I trust you?” Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowl-edge discovery and data mining* 1:1135–1144
- Rodríguez-Pérez, R. dan Bajorath, J. 2020. Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley values. *J Med Chem*. Vol. 63 (16): 8761-8777
- Samsudiney. 2019. Penejelasan Sederhana tentang Apa itu SVM?. <https://medium.com/> [diakses pada 12 Maret 2022]
- Sarker IH. 2021. Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Comput Sci*. Vol. 1: 1-20
- Shapley LS. 1953. *A value for N-person games*. *Contributions to the theory of games*. In: Kuhn HW, Tucker AW (eds) *Annals of mathematical studies*. Princeton University Press, Princeton, pp 307–317
- Shapley LS. 2016. *A Value for n-Person Games*. In: *Contributions to the Theory of Games (AM-28)*, Volume II.

Vapnik V.N. 1999. *The Nature of Statistical Learning Theory 2nd edition*. Springer-Verlag, New York Berlin Heidelberg.

Vijayakumar, S. dan Wu, S. 1999. Sequential Support Vector Classifier and Regression. *Prociding International Conference on Soft Computing (SOCO'99)*. Hal. 610-619.

Yap, B. W., Rani, K. A., Rahman, H. A., Fong, S., Khairudin, Z. dan Abdullah, N. N. 2014. An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets. *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*. 285, pp. 13-22. Singapore: Springer. doi:10.1007/978-981-4585-18-7\_2.

Young, H.P. 1985. Monotonic solutions of cooperative games. *Int J Game Theory* 14:65–72





Lampiran

Lampiran 1. Kuisisioner Susenas Blok XVII akses terhadap makanan pada Survei Sosial Ekonomi Nasional (SUSENAS) 2021

<b>BLOK XVII. AKSES TERHADAP MAKANAN (DITANYAKAN PADA KRT/PASANGAN/ART 15 TAHUN KE ATAS)</b>	
Nama dan No. Urut pemberi informasi: ..... <input type="checkbox"/> <input type="checkbox"/>	
SEKARANG SAYA AKAN MENGAJUKAN BEBERAPA PERTANYAAN MENGENAI AKSES TERHADAP MAKANAN. <b>DALAM SETAHUN TERAKHIR, APAKAH ADA SAAT DI MANA:</b>	
1701. SELAMA SETAHUN TERAKHIR, APAKAH ANDA/ART LAINNYA <b>KHAWATIR TIDAK AKAN MEMILIKI CUKUP MAKANAN</b> UNTUK DISANTAP KARENA KURANGNYA UANG ATAU SUMBER DAYA LAINNYA?	Ya ..... 1 Tidak ..... 5 Tidak tahu ..... 8 Menolak menjawab ..... 9
1702. SELAMA SETAHUN TERAKHIR, APAKAH ADA SAAT DI MANA ANDA/ART LAINNYA <b>TIDAK DAPAT MENYANTAP MAKANAN SEHAT DAN BERGIZI</b> KARENA KURANGNYA UANG ATAU SUMBER DAYA LAINNYA?	Ya ..... 1 Tidak ..... 5 Tidak tahu ..... 8 Menolak menjawab ..... 9
1703. SELAMA SETAHUN TERAKHIR, APAKAH ANDA/ART LAINNYA HANYA <b>MENYANTAP SEDIKIT JENIS MAKANAN</b> KARENA TIDAK MEMILIKI UANG ATAU SUMBER DAYA LAINNYA?	Ya ..... 1 Tidak ..... 5 Tidak tahu ..... 8 Menolak menjawab ..... 9
1704. SELAMA SETAHUN TERAKHIR, APAKAH ANDA/ART LAINNYA PERNAH <b>MELEWATKAN SATU WAKTU MAKAN</b> PADA SUATU HARI <b>TERTEWAS</b> KARENA TIDAK MEMILIKI UANG ATAU SUMBER DAYA LAIN YANG CUKUP UNTUK MENDAPATKAN MAKANAN?	Ya ..... 1 Tidak ..... 5 Tidak tahu ..... 8 Menolak menjawab ..... 9
1705. SELAMA SETAHUN TERAKHIR, APAKAH ANDA/ART LAINNYA <b>MAKAN LEBIH SEDIKIT DARIPADA SEHARUSNYA</b> KARENA KURANGNYA UANG ATAU SUMBER DAYA LAINNYA?	Ya ..... 1 Tidak ..... 5 Tidak tahu ..... 8 Menolak menjawab ..... 9
1706. SELAMA SETAHUN TERAKHIR, APAKAH RUMAH TANGGA <b>KEHABISAN MAKANAN</b> KARENA KURANGNYA UANG ATAU SUMBER DAYA LAINNYA?	Ya ..... 1 Tidak ..... 5 Tidak tahu ..... 8 Menolak menjawab ..... 9
1707. SELAMA SETAHUN TERAKHIR, APAKAH ANDA/ART LAINNYA <b>MERASA LAPAR TAPI TIDAK MAKAN</b> KARENA KURANGNYA UANG ATAU SUMBER DAYA LAINNYA UNTUK MENDAPATKAN MAKANAN?	Ya ..... 1 Tidak ..... 5 Tidak tahu ..... 8 Menolak menjawab ..... 9
1708. SELAMA SETAHUN TERAKHIR, APAKAH ANDA/ART LAINNYA <b>TIDAK MAKAN SEHARIAN</b> KARENA KURANGNYA UANG ATAU SUMBER DAYA LAINNYA?	Ya ..... 1 Tidak ..... 5 Tidak tahu ..... 8 Menolak menjawab ..... 9

Lampiran 2. Sebaran responden rumah tangga

Variabel	Kategori	Jumlah rumah tangga	Rawan pangan
Pendidikan KRT	Tidak tamat SD	4.757	27 %
	SD	8.054	16 %
	SMP	3.855	16 %
	SMA	5.976	13 %
	Perguruan tinggi	2.150	7%
Rumah tangga rentan	Tidak	23.738	16%
	Ya	1.054	23%
Sumber pendapatan sebagai penerima transfer	Tidak	22.809	16%
	Ya	1.983	24%
Kepemilikan aset tanah	Tidak	5.062	24%
	Ya	19.730	15%
Akses internet	Tidak	11.168	20%
	Ya	13.624	14%
Sakit namun tidak rawat jalan	Tidak	24.784	17%
	Ya	8	38%
Jumlah anggota rumah tangga yang memiliki tabungan	Tidak ada	9.122	26%
	1	8.297	13%
	2 atau lebih	7.373	3%
Jumlah anggota rumah tangga buta huruf	Tidak ada	21.301	15%
	Ada	3.491	27% <sup>s</sup>
Penerima Program Keluarga Harapan (PKH)	Tidak	21.257	15%
	Ya	3.535	25%
Penerima Kartu Sejahtera (KKS)	Tidak	22.009	15%
	Ya	2.783	27%
Penerima Bantuan Pemda	Tidak	24.372	17%
	Ya	420	24%
Penerima Bantuan Pangan Non Tunai (BPNT)	Tidak	18.516	14%
	Ya	6.276	25%
Penerima Program Bantuan Iuran (PBI) BPJS	Tidak	15.787	14%
	Ya	9.005	21%

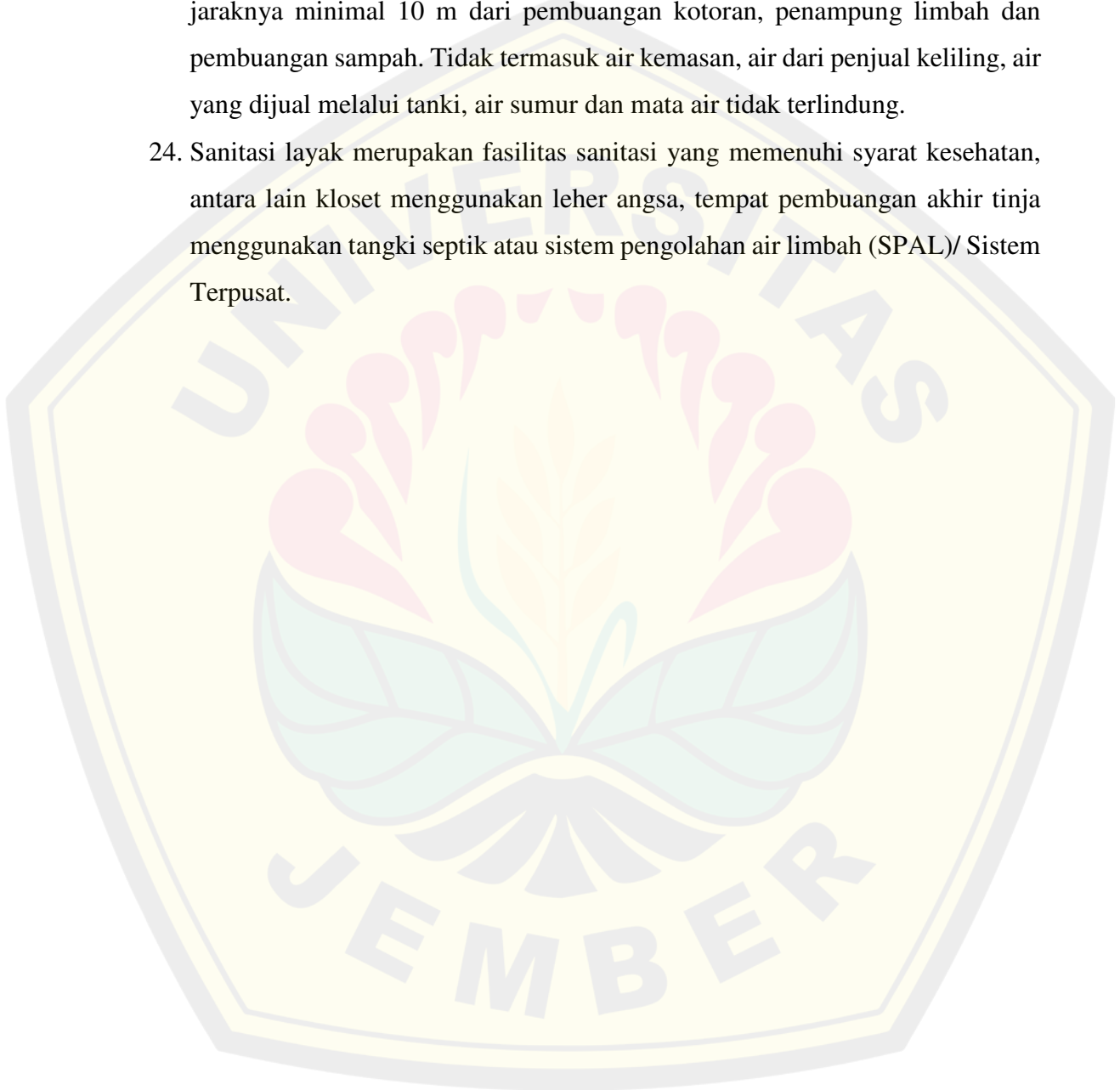
Variabel	Kategori	Jumlah rumah tangga	Rawan pangan
Kepemilikan	Tidak	21.605	17%
Jamkesda	Ya	3.187	16%
Kepemilikan KIP	Tidak	24.209	17%
	Ya	583	22%
Jenis atap rumah	Asbes	1.096	23%
	Seng	288	27%
	Beton	495	12%
	Genteng	22.831	16%
	Lainnya	82	27%
Jenis lantai rumah	Ubin/tegel/teraso	2.165	17%
	Semen/bata merah	133	35%
	Parket/vinil/karpet	53	23%
	Marmmer/keramik	15.695	13%
	Lainnya	6.746	25%
Jenis dinding rumah	Kayu/papan	1.885	22%
	Tembok/plesteran	22.100	15%
	Lainnya	807	41%
Sumber penerangan rumah	Non listrik	17	35%
	Listrik	24.775	17%
Luas Lantai	< 50	5.069	26%
	50 – 100	13.429	16%
	101 – 200	5.248	11%
	> 200	1.0465	7%
Bahan bakar masak	Tidak memasak	293	18%
	Kayu bakar	3.675	24%
	Minyak tanah	42	43%
	LPG 3Kg	19.656	16%
	LPG	634	5%
	12Kg/5.5Kg/Bluegaz	492	16%
Sumber air minum	Lainnya	492	16%
	Sumur bor/pompa	6.477	20%
	Leding	2.411	17%
	Air kemasan/isi ulang	7.831	12%
	Lainnya	8.073	18%
Air minum layak	Tidak	22.333	17%
	Ya	2.459	17%
Sanitasi layak	Tidak	5.931	26%
	Ya	18.861	14%

Penjelasan variabel yang digunakan :

1. Pendidikan kepala rumah tangga adalah pendidikan formal maupun non formal yang berhasil ditamatkan kepala rumah tangga yang terdiri dari tidak tamat SD, SD, SMP, SMA, dan Perguruan Tinggi.
2. Rumah tangga rentan/rawan adalah kepala rumah tangga perempuan yang memiliki anggota rumah tangga berusia 0 – 14 tahun.
3. Sumber pendapatan sebagai penerima transfer merupakan rumah tangga yang sumber terbesar pembiayaan rumah tangganya berasal dari kiriman uang/barang.
4. Kepemilikan aset tanah merupakan rumah tangga yang memiliki aset tanah yang dibuktikan dengan sertifikat tanah atau surat kepemilikan.
5. Akses internet adalah rumah tangga yang menggunakan internet termasuk sosial media dalam kurun waktu 3 bulan terakhir.
6. Sakit namun tidak rawat jalan adalah adanya anggota keluarga yang memiliki keluhan kesehatan hingga penyakit kronis dan sebagainya namun tidak rawat jalan.
7. Jumlah anggota rumah tangga yang memiliki tabungan merupakan jumlah anggota keluarga yang memiliki rekening tabungan dan dihitung sebagai satu penabung meski memiliki lebih dari satu tabungan.
8. Jumlah anggota rumah tangga buta huruf merupakan jumlah anggota keluarga berusia lebih dari lima tahun yang tidak bisa membaca dan menuliskan kalimat sederhana dalam bahasa sehari-hari (huruf latin/alfabet, huruf arab/hijaiyah, huruf jawa, huruf kanji, huruf cina dll).
9. Penerima Program Keluarga Harapan (PKH) adalah terdapat atau tidaknya rumah tangga mendapat program bantuan sosial bersyarat kepada keluarga miskin yang ditetapkan sebagai keluarga penerima manfaat PKH
10. Penerima Kartu Keluarga Sejahtera (KKS) adalah kartu penanda bagi Penyandang Masalah Kesejahteraan Sosial (PMKS) kurang mampu di antaranya penyandang disabilitas, lanjut usia yang belum memperoleh layanan atau bantuan sosial dan berada di dalam panti atau Lembaga Kesejahteraan Sosial (LKS)

11. Penerima Bantuan Pemerintah Daerah adalah terdapat atau tidaknya rumah tangga tersebut mendapatkan bantuan dari pemerintah daerah setempat.
12. Penerima Bantuan Pangan Non Tunai (BPNT) adalah terdapat atau tidaknya rumah tangga tersebut mendapatkan bantuan sosial pangan dalam bentuk non tunai dari pemerintah yang diberikan kepada Keluarga Penerima Manfaat (KPM) setiap bulannya melalui mekanisme perbankan.
13. Penerima Program Bantuan Iuran (PBI) BPJS adalah terdapat atau tidaknya rumah tangga tersebut mendapatkan bantuan sosial berupa bantuan iuran jaminan kesehatan BPJS Kesehatan.
14. Kepemilikan Jaminan Kesehatan Daerah adalah terdapat atau tidaknya anggota rumah tangga tersebut memiliki jaminan kesehatan daerah.
15. Kepemilikan KIP adalah terdapat atau tidaknya anggota rumah tangga yang menerima bantuan pendidikan berupa uang tunai, perluasan akses, dan kesempatan belajar dari pemerintah yang diberikan kepada peserta didik dan mahasiswa dari keluarga miskin atau rentan miskin untuk membiayai pendidikan.
16. Jenis atap rumah adalah jenis atap rumah yang digunakan oleh rumah tangga seperti beton, genteng, asbes, seng, bambu/kayu/sirap/, jeram/ijuk/daun/rumba ataupun jenis atap lainnya.
17. Jenis lantai rumah adalah jenis lantai rumah yang digunakan oleh rumah tangga seperti marmer/keramik/granit, tegel/traso, semen, kayu, tanah ataupun jenis lantai rumah lainnya.
18. Jenis dinding rumah adalah jenis dinding pada rumah yang ditempati seperti kayu/papan, tembok/plesteran/anyaman, dan jenis dinding rumah lainnya.
19. Sumber penerangan rumah adalah sumber utama penerangan rumah tangga berupa listrik, baik listrik PLN dengan meteran, listrik PLN tanpa meteran, maupun listrik non-PLN
20. Luas Lantai merupakan jumlah luas lantai dari bangunan tempat tinggal yang dihuni oleh rumah tangga.
21. Bahan bakar masak adalah jenis bahan bakar utama yang digunakan rumah tangga tersebut untuk memasak.

22. Sumber air minum adalah rumah tangga yang sumber utama air minumnya berasal dari ledeng air, air terlindungi, air hujan, dan air minum kemasan.
23. Air minum layak adalah air minum yang terlindungi meliputi air ledeng (keran), keran umum, *hydrant* umum, terminal air, penampungan air hujan, atau mata air dan sumur terlindungi, sumur bor atau sumur pompa, yang jaraknya minimal 10 m dari pembuangan kotoran, penampung limbah dan pembuangan sampah. Tidak termasuk air kemasan, air dari penjual keliling, air yang dijual melalui tanki, air sumur dan mata air tidak terlindungi.
24. Sanitasi layak merupakan fasilitas sanitasi yang memenuhi syarat kesehatan, antara lain kloset menggunakan leher angsa, tempat pembuangan akhir tinja menggunakan tangki septik atau sistem pengolahan air limbah (SPAL)/ Sistem Terpusat.



Lampiran 3. Nilai SHAP pada setiap variabel Provinsi Jawa Timur

No	Variabel	Nilai Variabel Penting SHAP
1	Pendidikan_KRT	0,04048
2	KRT_Rawan	0,00487
3	Jumlah_Penabung	0,07944
	Jumlah_Buta_Huruf	0,01483
4	Penerima_Transfer	0,01014
5	Aset_Tanah	0,01964
6	Akses_Internet	0,03340
7	Sakit_Namun_Tidak_Rawat_Jalan	0,00003
8	PKH	0,01253
9	KKS	0,00946
10	BPNT	0,02079
11	BPJS	0,02505
12	Jamkesda	0,01468
13	KIP	0,00259
14	Jenis_Atap	0,01347
15	Jenis_Lantai	0,04043
16	Jenis_Dinding	0,01693
17	Listrik	0,00005
18	Bahan_Bakar_Masak	0,02287
19	Sumber_Air_Minum	0,03978
20	Sanitasi_Layak	0,02109
21	Bantuan_Pemda	0,00213
22	Luas_Lantai	0,08928
23	Air_Minum_Layak	0,01651
24	Air_Layak	0,02749

Lampiran 4 Matriks korelasi nilai SHAP setiap variabel pada Provinsi Jawa Timur

