PENERAPAN METODE *PRINCIPAL COMPONENT ANALYSIS-SUPPORT VECTOR MACHINE* PADA KLASIFIKASI STATUS *STUNTING* DI KALIMANTAN BARAT

MARIA MEILINDA INDRIANI
NIM H1091211060

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

PENERAPAN METODE *PRINCIPAL COMPONENT ANALYSIS-SUPPORT VECTOR MACHINE* PADA KLASIFIKASI STATUS *STUNTING* DI KALIMANTAN BARAT

MARIA MEILINDA INDRIANI
NIM H1091211060

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika



PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir	:	Penerapan Metode *Principal Component Analysis-Support Vector Machine* Pada Klasifikasi Status *Stunting* di Kalimantan Barat
Nama Mahasiswa	:	Maria Meilinda Indriani
NIM	:	H1091211060
Jurusan/Program Studi	:	Matematika/Statistika
Tanggal Lulus	:	31 Januari 2025
SK Pembimbing	:	No.2634/UN22.8/TD.06/2024/Tanggal 2 September 2024
SK Penguji	:	No.98/UN22.8/TD.06/2025/Tanggal 23 Januari 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Shantika Martha, M.Si.	Hendra Perdana, M.Sc.
NIP. 198403082008122003	NIP. 198810102019031020

Dosen Penguji

Ketua Penguji	Anggota Penguji
Nurfitri Imro'ah, M.Si.	Dr. Evy Sulistianingsih, M.Sc.
NIP. 198907182019032021	NIP. 198502172008122006

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Shantika Martha, M.Si.	Hendra Perdana, M.Sc.
NIP. 198403082008122003	NIP. 198810102019031020

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP. 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama	: Maria Meilinda Indriani
NIM	: H1091211060
Program Studi/ Jurusan	: Statistika/Matematika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, Januari 2025

Maria Meilinda Indriani
H1091211060

Penerapan Metode *Principal Component Analysis-Support Vector Machine* Pada Klasifikasi Status *Stunting* di Kalimantan Barat

Abstrak

Stunting merupakan permasalahan kesehatan masyarakat yang serius, ditandai dengan gangguan pertumbuhan anak akibat kekurangan gizi kronis dalam jangka panjang. Kondisi ini tidak hanya memengaruhi tinggi badan anak yang berada di bawah standar, tetapi juga memengaruhi perkembangan kognitif, kemampuan belajar, dan produktivitas di masa depan. Kalimantan Barat menjadi salah satu provinsi di Indonesia dengan prevalensi *stunting* yang relatif tinggi. Penelitian ini bertujuan untuk mengklasifikasikan status *stunting* di 14 kabupaten/kota di Kalimantan Barat menggunakan metode *Principal Component Analysis-Support Vector Machine* (PCA-SVM) serta mengevaluasi tingkat akurasi model klasifikasi yang dihasilkan. PCA digunakan untuk mereduksi dimensi data dan mengatasi multikolinearitas, serta menghasilkan komponen utama yang independen sebagai variabel *input* dalam model. Selanjutnya, SVM dengan *kernel Radial Basis Function* (RBF) diterapkan untuk membangun model klasifikasi yang optimal dan efektif. Penelitian ini menggunakan 14 variabel yang mencerminkan kondisi kesehatan anak, ibu, lingkungan, serta aspek sosial-ekonomi. Wilayah-wilayah tersebut dikategorikan ke dalam tiga kelompok status *stunting*, yaitu rendah, sedang, dan tinggi. Kinerja model dievaluasi berdasarkan tingkat akurasi. Hasil penelitian menunjukkan bahwa pendekatan PCA-SVM dengan *kernel* RBF berhasil mencapai akurasi sebesar 92,86% dalam mengklasifikasikan status *stunting*. Analisis ini memberikan wawasan penting mengenai pengelompokan status *stunting* di setiap wilayah, sehingga dapat dijadikan landasan untuk merumuskan kebijakan intervensi yang lebih terarah dan tepat sasaran. Dengan demikian, hasil penelitian ini diharapkan dapat mendukung upaya pemerintah daerah dalam menurunkan prevalensi *stunting* secara efektif dan berkelanjutan.

Kata Kunci: *Stunting*, Klasifikasi, PCA, SVM, *Kernel* RBF

Implementation of Principal Component Analysis-Support Vector Machine Method for Stunting Status Classification in West Kalimantan

Abstract

Stunting is a serious public health issue characterized by impaired growth in children due to chronic malnutrition over an extended period. This condition not only affects children's height, which falls below the standard, but also impacts their cognitive development, learning ability, and future productivity. West Kalimantan is one of the provinces in Indonesia with a relatively high prevalence of stunting. This study aims to classify stunting in 14 districts/cities in West Kalimantan using the Principal Component Analysis-Support Vector Machine (PCA-SVM) method and to evaluate the accuracy of the resulting classification model. PCA is used to reduce data dimensionality and address multicollinearity, producing independent principal components as input variables for the model. Subsequently, SVM with a Radial Basis Function (RBF) kernel is applied to develop an optimal and effective classification model. This study employs 14 variables reflecting child and maternal health conditions, environmental factors, and socio-economic aspects. These areas are categorized into three stunting status groups: low, moderate, and high. The model's performance is evaluated based on accuracy. The results show that the PCA-SVM approach with the RBF kernel achieved an accuracy of 92.86% in classifying stunting status. This analysis provides valuable insights into the grouping of stunting status in each region, which can serve as a basis for designing more targeted and precise intervention policies. Thus, this study is expected to support local government efforts to effectively and sustainably reduce stunting prevalence.

Keywords: *Stunting, Classification, PCA, SVM RBF Kernel*

PRAKATA

Puji syukur kehadirat Tuhan Yang Maha Esa, karena berkat dan rahmat-Nya, penulis dapat menyelesaikan penulisan tugas akhir dengan judul **"Penerapan Metode *Principal Component Analysis-Support Vector Machine* Pada Klasifikasi Status *Stunting* di Kalimantan Barat"**. Tugas akhir dibuat untuk memenuhi salah satu syarat untuk memenuhi salah satu syarat untuk memperoleh gelar Sarjana Statistika pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak.

Penulis menyadari bahwa selama proses penulisan tugas akhir ini tidak lepas dari bimbingan, arahan, dan dukungan dari berbagai pihak. Maka dari itu, penulis ingin mengucapkan terima kasih sebesar-besarnya kepada:

1. Keluarga tercinta, terutama kedua orang tua penulis yaitu Bapak Elias & Ibu Anastresia, Alm. Kakek Najah, Nenek Lenyak, serta Abang Stefanus Andry. Terima kasih telah memberikan kepercayaan, pengorbanan, doa, semangat, dan selalu mendukung segala keputusan dan pilihan dalam hidup penulis. Terima kasih selalu berjuang untuk kehidupan penulis.
2. Ibu Shantika Martha, M.Si. selaku dosen pembimbing pertama yang telah banyak memberikan bimbingan dan arahan selama penyusunan tugas akhir ini.
3. Bapak Hendra Perdana, M.Sc. selaku dosen pembimbing kedua dan dosen pembimbing akademik yang telah banyak memberikan bimbingan dan arahan selama masa perkuliahan serta penyusunan tugas akhir ini.
4. Ibu Nurfitri Imro'ah, M.Si. selaku dosen penguji pertama yang telah memberikan masukan dan saran kepada penulis.
5. Ibu Dr. Evy Sulistianingsih, M.Sc. selaku dosen penguji kedua yang telah memberikan masukan dan saran kepada penulis.
6. Seluruh dosen Jurusan Matematika FMIPA Universitas Tanjungpura yang telah memberikan ilmu dan pengalaman selama perkuliahan.
7. Teruntuk Viren Marcellya Clarenda Siboro, Utin Azwa Sayhani Sadikin, dan Vriska Dwi Riovani yang membersamai penulis dari awal perkuliahan

sampai saat ini. Terima kasih selalu memberikan semangat dalam keadaan apapun serta menjadi bagian penting dalam perjalanan penulis hingga saat ini.

8. Teruntuk Abang Basilius Legito yang selalu menemani dalam penulisan skripsi ini. Terima kasih sudah menjadi bagian penting dalam perjalanan penulis hingga saat ini.
9. Teruntuk Mba Klaudia Jasinta dan Kak Tri Nela Selanda yang selalu menemani dan membersamai di segala kondisi. Terima kasih sudah menjadi teman dalam berproses dan bertumbuh di kota ini serta selalu mendukung dan membantu penulis.
10. Teman-teman Statistika Universitas Tanjungpura Angkatan 2021 yang telah membersamai selama masa perkuliahan serta memberikan semangat dalam penulisan skripsi ini.
11. Semua pihak yang tidak bisa penulis sampaikan satu persatu atas dukungan dan bantuannya baik dari segi material maupun non material.
12. *Last but not least*, terima kasih untuk diri sendiri, karena telah berjuang sejauh ini. *I wanna thank me for just being me at all times*.

Semoga Tuhan membalas semua kebaikan atas segala bantuan yang diberikan dalam penulisan tugas akhir ini. Penulis berharap tugas akhir ini dapat bermanfaat bagi semua pihak. Penulis menyadari bahwa tugas akhir ini jauh dari kata sempurna, oleh karena itu, kritik dan saran yang bersifat membangun sangat diharapkan oleh penulis untuk menjadikan tugas akhir ini lebih baik.

Pontianak, Januari 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN ..... Error! Bookmark not defined.
PERNYATAAN INTEGRITAS AKADEMIK ..... **iii**
Abstrak ..... **iv**
Abstract ..... **v**
PRAKATA ..... **vi**
DAFTAR ISI ..... **viii**
DAFTAR TABEL ..... **x**
DAFTAR GAMBAR ..... **xi**
DAFTAR LAMPIRAN ..... **xii**
DAFTAR SIMBOL ..... **xiii**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 4
1.4 Batasan Masalah 4
1.5 Tinjauan Pustaka 4
1.6 Metodologi Penelitian 5
BAB II LANDASAN TEORI **9**
2.1 *Machine Learning* 9
2.2 Klasifikasi 10
2.3 *Confusion Matrix* 11
2.4 *Outlier* 11
2.5 Standarisasi Data 12
2.6 Multikolinearitas 14
2.7 Matriks Varians-Kovarians 15
2.8 Nilai Eigen dan Vektor Eigen 16
2.9 *Stunting* 17
BAB III *PRINCIPAL COMPONENT ANALYSIS – SUPPORT VECTOR MACHINE* **20**

3.1 *Principal Component Analysis* 20
3.2 *Support Vector Machine* 22
3.3 *Soft Margin* 25
3.4 Fungsi *Kernel* (*Kernel Trick*) 25
BAB IV HASIL DAN PEMBAHASAN 29
4.1 Deskripsi Data 29
4.2 Kategorisasi Variabel Dependen 31
4.3 Deteksi *Outlier* dengan *Boxplot* 31
4.4 Standarisasi Data 32
4.5 Uji Multikolinearitas 32
4.6 Reduksi Dimensi Data dengan *Principal Component Analysis* (PCA) .. 33
4.7 Klasifikasi dengan SVM *Kernel* RBF 39
4.7.1 *Hyperparameter Tuning* 39
4.7.2 Pembentukan SVM *Kernel* RBF 40
BAB V KESIMPULAN DAN SARAN 44
5.1 Kesimpulan 44
5.2 Saran 45
DAFTAR PUSTAKA 46
LAMPIRAN 52

DAFTAR TABEL

Tabel 2.1 *Confusion Matrix* 11
Tabel 2.2 Pengelompokkan Status *Stunting* 18
Tabel 4.1 Variabel Penelitian 29
Tabel 4.2 Statistik Deskriptif Data *Stunting* 30
Tabel 4.3 Kategorisasi Variabel Dependen 31
Tabel 4.4 Nilai VIF Data *Stunting* 32
Tabel 4.5 Nilai Eigen 35
Tabel 4.6 Nilai *Component Matrix* 36
Tabel 4.7 Data Baru Hasil PCA 38
Tabel 4.8 Hasil Perhitungan Nilai pada SVM *Kernel* RBF 40
Tabel 4.9 *Confusion Matrix* 41
Tabel 4.10 Perbandingan Hasil Aktual dan Prediksi 42

DAFTAR GAMBAR

Gambar 1.1 *Flowchart* Penelitian 8
Gambar 2.1 Ilustrasi *Boxplot* 12
Gambar 3.1 Ilustrasi SVM dengan *Hyperplane* 22
Gambar 3.2 Pemetaan *Input Space* Berdimensi Dua ke Ruang Vektor yang Berdimensi Lebih Tinggi 26
Gambar 4.1 *Scree Plot* Nilai Eigen 36

DAFTAR LAMPIRAN

Lampiran 1 Data *Stunting* di Kalimantan Barat Tahun 2023 52
Lampiran 2 Hasil Standarisasi Variabel yang Terdeteksi *Outlier*..... 53
Lampiran 3 *Output* Deteksi *Outlier* menggunakan *Software RStudio* 53
Lampiran 4 *Syntax Software RStudio* 57

DAFTAR SIMBOL

σ	:	Standar deviasi
R_i^2	:	Koefisien determinasi dari regresi variabel independen X_i
$\boldsymbol{\Sigma}$	:	Matriks varians-kovarians
$\boldsymbol{A_{n\times n}}$	:	Matriks berukuran $n \times n$
λ	:	Nilai eigen
$\boldsymbol{I}$	:	Matriks identitas
$\boldsymbol{v}$	:	Vektor eigen
p	:	Banyak variabel independen dalam suatu model regresi
m	:	Banyak komponen utama dalam PCA
$\boldsymbol{y_i}$	:	Matriks komponen utama
$\boldsymbol{w}$	:	Vektor bobot berdimensi l
$\|\boldsymbol{w}\|$	:	Normalisasi $\boldsymbol{w}$
$\boldsymbol{x_i}$	:	Data *input* ke-i, dengan $i = 1,2, \dots, l$
b	:	Nilai konstanta bias
$\tau(\boldsymbol{w})$	:	Fungsi objektif yang digunakan untuk meminimalkan bobot $\boldsymbol{w}$
α_i, α_j	:	Koefisien *lagrange multiplier* untuk data ke-i dan ke-j
α_i^*	:	Koefisien *multiplier* untuk pelanggaran margin (*soft margin)*
L	:	Fungsi objektif yang dimaksimalkan untuk menentukan nilai α_i yang optimal
ξ_i, ξ_i^*	:	Variabel *slack*
$\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j})$	:	Fungsi *kernel*
$\Phi(\boldsymbol{x_i})$	:	Fungsi yang memetakan $\boldsymbol{x_i}$ ke dalam suatu dimensi
γ	:	Nilai parameter konstanta *gamma*
C	:	Nilai parameter konstanta *cost*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Stunting merupakan salah satu permasalahan kesehatan masyarakat yang masih menjadi perhatian utama saat ini, baik di tingkat nasional maupun global. *Stunting* diartikan sebagai gangguan pertumbuhan yang ditandai dengan tinggi badan anak yang berada di bawah standar *World Health Organization* (WHO) untuk usianya. Penyebab utama *stunting* adalah kekurangan gizi kronis sejak dalam kandungan hingga dua tahun pertama kehidupan anak. Selain itu, faktor lingkungan seperti akses terhadap sanitasi yang buruk dan infeksi berulang semakin memperburuk kondisi ini. Dampaknya tidak hanya terbatas pada fisik anak, tetapi juga meluas pada perkembangan kognitif, kemampuan belajar, dan produktivitas individu di masa depan (Fauziah et al., 2024). Dalam skala yang lebih besar, tingginya prevalensi *stunting* berpotensi menghambat pertumbuhan ekonomi negara karena hilangnya sumber daya manusia yang optimal.

Stunting masih menjadi masalah gizi utama yang dihadapi Indonesia. Pada tahun 2022, prevalensi *stunting* nasional mencapai 21,6 persen, turun dari 24,4 persen pada tahun sebelumnya (Kementerian Kesehatan, 2023). Namun, angka ini masih berada di atas ambang batas yang ditetapkan WHO, yakni 20 persen. Kalimantan Barat merupakan salah satu provinsi dengan tingkat *stunting* yang cukup mengkhawatirkan, menempati peringkat ke-13 dengan prevalensi *stunting* tertinggi di Indonesia pada tahun 2023 (Survei Kesehatan Indonesia, 2023). Pada tahun 2022, prevalensi *stunting* di Kalimantan Barat mencapai 27,8 persen. Survei Kesehatan Indonesia (2023) menyatakan bahwa meskipun berhasil turun menjadi 20,6 persen pada tahun 2023, masih diperlukan strategi intervensi yang lebih spesifik dan berbasis data, dengan berfokus pada wilayah berisiko tinggi untuk menurunkan prevalensi *stunting* secara menyeluruh.

Faktor-faktor penyebab *stunting* sangat beragam dan sering kali saling berkaitan. Faktor utama meliputi asupan gizi yang tidak memadai, rendahnya kesehatan ibu selama hamil, akses terhadap air bersih dan sanitasi, pola asuh yang tidak tepat, serta kondisi ekonomi-sosial keluarga (UNICEF, 2023). Interaksi antar faktor-faktor ini menciptakan kompleksitas dalam analisis data yang sering kali menyebabkan multikolinearitas, sehingga dapat mengurangi keakuratan model analisis. Metode *Principal Component Analysis* (PCA) dapat digunakan untuk mengatasi masalah ini dengan melakukan pendekatan reduksi dimensi. Menurut Badri dan Sari (2021), PCA mampu menyederhanakan variabel yang saling berkorelasi menjadi komponen-komponen utama yang independen. Dengan demikian, PCA tidak hanya mengatasi multikolinearitas tetapi juga membantu mengidentifikasi faktor-faktor dominan yang berpengaruh.

Selain memahami faktor risiko, klasifikasi wilayah berdasarkan tingkat risiko *stunting* merupakan langkah penting dalam mendukung pengambilan keputusan yang tepat. Menurut laporan Sekretariat Wakil Presiden Republik Indonesia dan Tim Nasional Percepatan Penanggulangan Kemiskinan (2018), status daerah *stunting* di Indonesia dikelompokkan menjadi empat kategori, yaitu rendah, sedang, tinggi, dan sangat tinggi. Klasifikasi ini digunakan sebagai dasar dalam menentukan prioritas intervensi untuk percepatan pencegahan *stunting* yang dilakukan secara bertahap di seluruh kabupaten/kota. Melalui pendekatan kerja sama antar sektor, program ini dimulai pada tahun 2018 di 100 kabupaten/kota, diperluas pada tahun 2019, dan direncanakan mencakup seluruh wilayah secara bertahap hingga tahun 2024 dengan tujuan memastikan bahwa daerah dengan prevalensi tinggi mendapatkan penanganan intensif yang sesuai dengan kebutuhan mereka (Sekretariat Wakil Presiden Republik Indonesia & Tim Nasional Percepatan Penanggulangan Kemiskinan, 2018).

Klasifikasi *stunting* memiliki peran penting sebagai dasar dalam merancang program pencegahan *stunting* (Gaffar et al., 2024). Salah satu metode klasifikasi yang dapat digunakan adalah *Support Vector Machine* (SVM). Menurut Syafika dan Karisma (2023), SVM tidak hanya efektif untuk menangani data berdimensi tinggi,

tetapi juga dapat diterapkan pada data yang berdimensi rendah, serta tetap mempertahankan kinerjanya yang baik. SVM bekerja dengan mencari *hyperplane* optimal yang dapat memisahkan data ke dalam kelas-kelas yang berbeda. Proses ini dilakukan dengan memaksimalkan *margin* antar kelas yang dikenal dengan *support vector*. Hal ini memungkinkan SVM menghasilkan prediksi yang akurat meskipun data sulit dipisahkan secara linear. Dalam banyak kasus, data tidak selalu dapat dipisahkan secara linear di ruang dimensi rendah, sehingga perlu menggunakan fungsi *kernel*. Menurut Achmal, Cholissodin, dan Adikara (2022), *kernel Radial Basis Function* (RBF) merupakan fungsi *kernel* yang digunakan dalam mengatasi kasus non-linear, sekaligus dapat menangani kasus linear dengan hasil yang kurang optimal. Penentuan *hyperparameter* optimal untuk menghasilkan model terbaik merupakan tantangan umum dalam penggunaan model SVM. *Hyperparameter* adalah parameter yang perlu diatur sebelum proses pelatihan dimulai. Pemilihan *hyperparameter* optimal dapat dilakukan menggunakan metode *grid search*. *Grid search* adalah metode optimasi yang bekerja dengan mencoba setiap kombinasi nilai *hyperparameter* satu per satu sesuai dengan kombinasi yang telah ditentukan sebelumnya dan kemudian memilih *hyperparameter* terbaik berdasarkan hasil pemrosesan data (Nugraha & Sasongko, 2022).

Berdasarkan uraian di atas, penggunaan metode PCA-SVM dengan *kernel* RBF ditetapkan dalam penelitian ini untuk mengklasifikasi tingkat *stunting* di Kalimantan Barat. Metode ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai distribusi risiko *stunting* di setiap wilayah. Informasi ini dapat menjadi dasar untuk merancang intervensi yang lebih terarah dan efektif sehingga prevalensi *stunting* dapat terus ditekan hingga mencapai target nasional, memberikan kontribusi nyata dalam meningkatkan kualitas kesehatan masyarakat, dan mengurangi beban *stunting* di wilayah yang masih menjadi tantangan utama.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana metode *Principal Component Analysis-Support Vector Machine* dapat digunakan untuk mengklasifikasikan status *stunting* di kabupaten/kota di Kalimantan Barat?
2. Sejauh mana tingkat akurasi model klasifikasi yang dihasilkan dalam menganalisis dan memetakan status *stunting* di Kalimantan Barat?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan penelitian ini adalah sebagai berikut:

1. Mengklasifikasikan status *stunting* di kabupaten/kota di Kalimantan Barat menggunakan metode *Principal Component Analysis-Support Vector Machine*.
2. Mengevaluasi tingkat akurasi model klasifikasi yang dihasilkan dalam menganalisis dan memetakan status *stunting* di Kalimantan Barat.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini difokuskan pada analisis data *stunting* di Kalimantan Barat tahun 2023. Data yang digunakan merupakan data indikator *stunting* yang diperoleh dari Badan Pusat Statistik Provinsi Kalimantan Barat dan Dinas Kesehatan Provinsi Kalimantan Barat. Indikator-indikator tersebut mencakup aspek kesehatan anak, kesehatan ibu, kesehatan lingkungan, serta kondisi ekonomi masyarakat. Dalam penelitian ini, status daerah *stunting* akan diklasifikasikan ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi.

1.5 Tinjauan Pustaka

Penelitian Alamsyah dan Fadila (2021) memanfaatkan PCA untuk mengekstraksi dan mengurangi jumlah fitur pada data prediksi hepatitis. Data yang telah direduksi kemudian diklasifikasikan menggunakan SVM, dengan pengukuran kinerja dilakukan melalui *Confusion Matrix*. Hasil penelitian ini menunjukkan akurasi prediksi hepatitis sebesar 93,55%, lebih tinggi dibandingkan dengan hasil klasifikasi SVM tanpa penerapan PCA.

Penelitian Utami dan Arianti (2020) menerapkan metode PCA-SVM untuk mengklasifikasikan tingkat kesejahteraan masyarakat di Kabupaten Brebes tahun 2018. Penelitian ini menggunakan 13 indikator yang diperoleh dari hasil Survei Sosial Ekonomi Nasional (SUSENAS) oleh Badan Pusat Statistik. Hasil analisis menunjukkan bahwa metode PCA-SVM lebih efektif dibandingkan penggunaan SVM tanpa PCA.

Penelitian Ikhwana, Nusrang, dan Sudarmin (2021) menggunakan metode PCA-SVM untuk mengklasifikasikan Indeks Kepuasan Masyarakat di Kabupaten Jeneponto. Hasil penelitian menunjukkan bahwa metode PCA-SVM menghasilkan kinerja klasifikasi yang unggul dibandingkan metode SVM konvensional, dengan akurasi mencapai 100%.

Abdillah dan Prianto (2017) mengaplikasikan PCA-SVM untuk mendeteksi penyakit diabetes. Penelitian ini bertujuan menentukan model klasifikasi optimal serta mengevaluasi kinerja PCA-SVM. Hasilnya menunjukkan bahwa PCA-SVM mencapai akurasi sebesar 77,79%, sensitivitas sebesar 73,26%, spesifisitas sebesar 81,30%, dan waktu komputasi sekitar 9,21 detik yang menjadikannya model yang baik untuk mendeteksi diabetes.

1.6 Metodologi Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari publikasi Badan Pusat Statistik Provinsi Kalimantan Barat dan Dinas Kesehatan Provinsi Kalimantan Barat tahun 2023. Metode yang digunakan dalam penelitian ini adalah metode *Principal Component Analysis–Support Vector Machine* (PCA-SVM). Langkah-langkah penelitian diawali dengan pengumpulan data yang terdiri atas variabel dependen, yaitu balita *stunting* dan 13 variabel independen, yaitu balita gizi buruk, cakupan bayi berat badan lahir rendah (BBLR), cakupan imunisasi dasar lengkap pada bayi, bayi usia kurang dari enam bulan diberikan ASI eksklusif, persalinan pelayanan kesehatan, cakupan konsumsi tablet tambah darah pada ibu hamil, cakupan pelayanan kesehatan ada ibu hamil K-6, wanita berusia 15-49 tahun yang berstatus kawin dan menggunakan KB, rumah tangga yang memiliki sumber air minum layak, rumah tangga yang memiliki akses sanitasi layak, penduduk

miskin, dan jumlah keluarga penerima manfaat (KPM) bantuan sosial pangan (BANSOS Pangan).

Setelah dilakukan analisis statistik deskriptif untuk memahami karakteristik data, dilakukan kategorisasi pada variabel dependen. Selanjutnya, deteksi *outlier* dilakukan menggunakan *boxplot* untuk mengidentifikasi dan memahami data yang memiliki nilai ekstrem yang dapat memengaruhi hasil analisis. Proses ini berguna untuk memastikan bahwa data yang digunakan lebih akurat, sehingga hasil analisis lebih valid. Kemudian dilakukan proses standarisasi data pada variabel independen dengan menggunakan metode *Robust Scalling*. Metode ini dipilih karena kemampuannya dalam menangani data dengan *outlier*, dengan standarisasi yang didasarkan pada median dan rentang interkuartil (IQR), sehingga nilai ekstrem tidak terlalu memengaruhi hasil standarisasi.

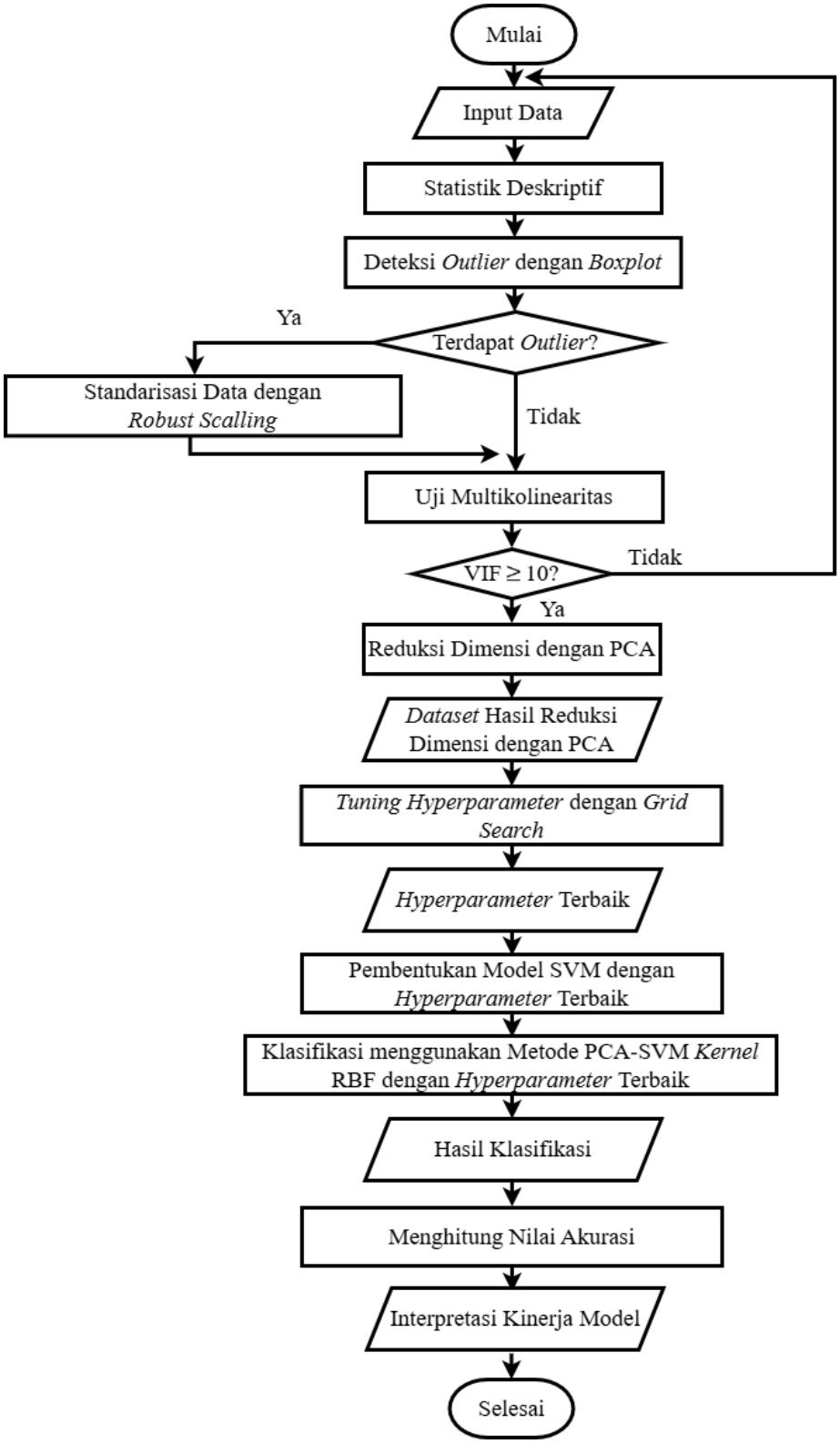
Setelah itu, dilakukan uji multikolinearitas menggunakan nilai *Variance Inflation Factor* (VIF) untuk mengidentifikasi adanya korelasi tinggi antar variabel independen. Uji ini bertujuan untuk memastikan bahwa variabel independen tidak memiliki hubungan linear yang kuat satu sama lain, yang dapat memengaruhi hasil analisis. Jika ditemukan variabel dengan nilai VIF ≥ 10, dilakukan pengurangan dimensi data menggunakan metode *Principal Component Analysis* (PCA) untuk memastikan data yang diolah lebih optimal.

PCA kemudian digunakan untuk mentransformasikan data menjadi komponen utama yang tidak berkorelasi dengan cara mereduksi dimensi data tanpa kehilangan informasi penting. Proses ini dilakukan dengan mencari kombinasi linear dari variabel-variabel independen yang memiliki variansi terbesar, sehingga komponen utama yang dihasilkan dapat mewakili struktur data secara keseluruhan. Dengan melakukan reduksi dimensi, PCA tidak hanya mengatasi masalah multikolinearitas tetapi juga meningkatkan efisiensi pemrosesan data, terutama pada analisis dengan jumlah variabel yang banyak. Hasil reduksi melalui PCA digunakan sebagai *input* untuk membangun model klasifikasi menggunakan metode *Support Vector Machine* (SVM) dengan *kernel Radial Basis Function* (RBF). *Kernel* RBF dipilih karena fleksibilitasnya dalam menangani data yang

memiliki pola hubungan non-linear, sekaligus mampu menangani kasus linear yang memiliki hasil yang kurang optimal. Selain itu, *kernel* RBF efektif dalam bekerja dengan *dataset* yang berukuran kecil hingga sedang.

Proses pembentukan model diawali dengan *tuning* parameter *gamma* dan *cost* menggunakan teknik 10-*fold cross-validation*. Teknik ini dipilih karena dapat meminimalkan *overfitting* dengan cara membagi data menjadi 10 bagian (*folds*), dimana 9 *folds* digunakan untuk melatih model, sementara 1 *fold* digunakan untuk menguji model secara bergantian. Pendekatan ini menghasilkan evaluasi yang lebih akurat dan terukur. Proses *tuning* parameter ini dilakukan menggunakan *package* e1071 pada perangkat lunak *RStudio* yang menyediakan fungsi *tune* untuk melakukan pencarian *grid* terhadap parameter *gamma* dan *cost*. Fungsi ini secara otomatis menerapkan *cross-validation* untuk mengevaluasi berbagai kombinasi parameter yang diuji. *Tuning* parameter dilakukan untuk menemukan kombinasi yang menghasilkan kinerja terbaik, sehingga model dapat mencapai keseimbangan antara bias dan variansi, sekaligus meningkatkan akurasi prediksi pada data.

Setelah proses *tuning* selesai, model klasifikasi dibangun dan *Confusion Matrix* digunakan untuk mengevaluasi kinerja model, dimana nilai akurasi digunakan sebagai metrik utama untuk menilai kemampuan model dalam mengklasifikasi status *stunting* di Kalimantan Barat ke dalam kategori rendah, sedang, dan tinggi. Tahapan-tahapan penelitian ini dilakukan secara berurutan, sebagaimana digambarkan dalam diagram alir pada Gambar 1.1. Penelitian ini bertujuan untuk membangun model klasifikasi yang akurat dan efektif, sehingga dapat menjadi alat bantu dalam pengambilan keputusan yang lebih terarah dan tepat sasaran untuk intervensi kesehatan.



Gambar 1.1 *Flowchart* Penelitian