

Pemodelan dan Prediksi Tingkat Kemiskinan Provinsi Sumatera Barat Menggunakan Support Vector Machine

Melani Septina Putri¹, Satrio Junaidi^{*2}, Ainil Mardiyah³Email: ¹melaniseptinaputri02@gmail.com, ^{2*}satriojunaidy@gmail.com, ³M.ainil@yahoo.com^{1,2,3}Program Studi Sains Data, Fakultas Sains dan Teknologi, Universitas PGRI Sumatera Barat

Diterima: 11 Maret 2026 | Direvisi: 23 April 2026 | Disetujui: 26 April 2026

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer

Universitas Muhammadiyah Riau, Indonesia

Abstrak

Penelitian ini dilatarbelakangi oleh permasalahan penyebaran tingkat kemiskinan di Provinsi Sumatera Barat yang masih bervariasi antar wilayah. Kondisi ini mendorong perlunya model prediksi yang mampu menggambarkan tingkat kemiskinan secara lebih akurat di masa mendatang. Tujuan penelitian ini adalah membangun model prediksi menggunakan algoritma *Support Vector Machine* (SVM), mengevaluasi kinerja model, serta mengimplementasikan hasil prediksi dalam bentuk *dashboard* interaktif sebagai pendukung pengambilan keputusan pemerintah daerah. Penelitian menggunakan data sekunder dari Badan Pusat Statistik (BPS) Provinsi Sumatera Barat periode 2015–2024 yang mencakup 19 kabupaten/kota. Variabel dependen adalah persentase penduduk miskin (P_0), sedangkan variabel independen terdiri dari tujuh indikator sosial ekonomi. Metode yang digunakan mengacu pada tahapan CRISP-DM. Pada tahap persiapan data dilakukan penanganan *missing value* dengan imputasi median, penanganan outlier menggunakan *winsorizing*, standarisasi dengan *Z-Score*, serta penambahan variabel lag satu periode (P_0_lag1). Data dibagi menjadi data latih (2015–2022) dan data uji (2023–2024), dengan optimasi parameter menggunakan *GridSearchCV* dan *TimeSeriesSplit*. Hasil penelitian menunjukkan bahwa model *Support Vector Regression* (SVR) dengan kernel *radial basis function* (RBF) memberikan performa terbaik dengan parameter $C=1000$, $\epsilon=0.05$, dan $\gamma=0.001$. Model ini menghasilkan nilai MAE sebesar 0.32, RMSE sebesar 0.36, dan R^2 sebesar 0.98. Implementasi hasil prediksi dalam *dashboard* Streamlit untuk periode 2025–2030 menunjukkan tren penurunan tingkat kemiskinan di sebagian besar wilayah. Model ini dinilai efektif sebagai dasar perencanaan dan evaluasi kebijakan penanggulangan kemiskinan berbasis data.

Kata kunci: kemiskinan, support vector machine, CRISP-DM

Modeling and prediction of poverty levels in west sumatra province using support vector machines

Abstract

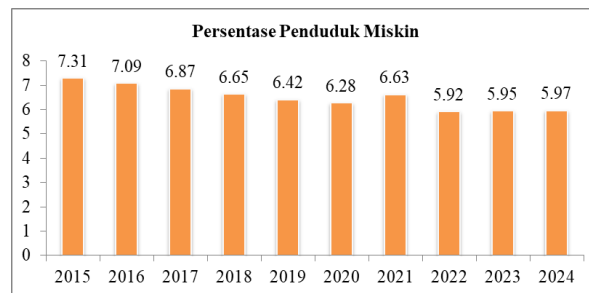
This research is motivated by the problem of poverty distribution in West Sumatra Province, which still varies between regions. The objectives of this study are to build a prediction model using the Support Vector Machine (SVM) algorithm, evaluate the model's performance, and implement the prediction results in the form of an interactive dashboard to support local government decision-making. The study uses secondary data from the Central Statistics Agency (BPS) of West Sumatra Province for the period 2015–2024, covering 19 districts/cities. The dependent variable is the percentage of poor people (P_0), while the independent variables consist of seven socio-economic indicators. The method used refers to the CRISP-DM stages. In the data preparation stage, missing values are handled using median imputation, outliers are handled using winsorizing, standardization is carried out using Z-Score, and the addition of a one-period lag variable (P_0_lag1). The data is divided into training data (2015–2022) and test data (2023–2024), with parameter optimization using GridSearchCV and TimeSeriesSplit. The results showed that the Support Vector Regression (SVR) model with a radial basis function (RBF) kernel provided the best performance with parameters $C=1000$, $\epsilon=0.05$, and $\gamma=0.001$. This model produced an MAE value of 0.32, RMSE of 0.36, and R^2 of 0.98. The implementation of the prediction results in the Streamlit dashboard for the 2025–2030 period showed a downward

trend in poverty levels in most regions. This model is considered effective as a basis for planning and evaluating data-based poverty alleviation policies.

Keywords: poverty, support vector machine, CRISP-DM

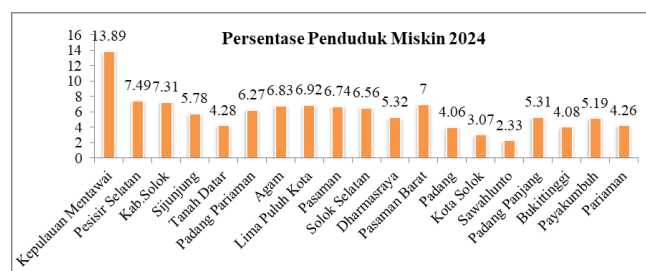
1. PENDAHULUAN

Kemiskinan merupakan permasalahan multidimensional yang masih menjadi tantangan utama dalam pembangunan, khususnya di negara berkembang seperti Indonesia. *Sustainable Development Goals* (SDGs) menetapkan tujuan pertama untuk menghapus kemiskinan dalam segala bentuk di seluruh dunia hingga tahun 2030. Seseorang dikatakan miskin apabila memiliki tingkat pendidikan yang rendah, ketidakmampuan untuk mendapatkan pekerjaan yang layak, pengeluaran per kapita yang tidak mencukupi untuk kebutuhan makanan, dan fasilitas perumahan yang tidak memenuhi standar hidup.



Gambar 1. Tingkat Kemiskinan Provinsi Sumatera Barat (2015-2024)

Berdasarkan data Badan Pusat Statistik (BPS), tingkat kemiskinan (P_0) di Indonesia pada tahun 2024 sebesar 9.03% atau sekitar 25.21 juta jiwa. Di Provinsi Sumatera Barat menunjukkan tren penurunan kemiskinan dalam sepuluh tahun terakhir, meskipun masih terdapat ketimpangan antar kabupaten/kota.



Gambar 2. Tingkat Kemiskinan Kabupaten/Kota

Sembilan dari 19 kabupaten/kota Provinsi Sumatera Barat memiliki tingkat kemiskinan di atas rata-rata provinsi. Kabupaten yang memiliki tingkat kemiskinan tertinggi yang diukur berdasarkan persentase penduduk miskin (P_0) sebesar 13.89% adalah Kepulauan Mentawai. Ketimpangan ini menunjukkan bahwa pertumbuhan ekonomi masih berpusat di daerah perkotaan, khususnya di kota Sawahlunto, Kota Solok, Padang, Bukittinggi, Pariaman, Padang Panjang dan Payakumbuh. Kondisi ini menunjukkan perlunya pendekatan berdasarkan data untuk mendukung perencanaan kebijakan yang lebih tepat sasaran.

Salah satu upaya yang dapat dilakukan untuk mengentaskan kemiskinan adalah dengan melakukan prediksi terhadap penduduk miskin di Provinsi Sumatera Barat. Prediksi adalah upaya untuk meramalkan apa yang akan terjadi di masa depan berdasarkan data dari masa lalu [1]. Dalam konteks jangka waktu yang panjang, prediksi memainkan peran yang krusial karena dapat menentukan waktu terjadinya peristiwa tertentu, sehingga memungkinkan persiapan yang diperlukan dapat dilakukan [2]. Oleh karena itu, prediksi berfungsi sebagai alat strategis dalam proses perencanaan dan pengambilan keputusan, mengingat kemampuannya untuk memberikan proyeksi awal tentang kondisi masa depan yang potensial, sehingga memungkinkan formulasi langkah-langkah pencegahan yang lebih optimal.

Salah satu metode yang digunakan dalam memprediksi tingkat kemiskinan adalah dengan menggunakan pembelajaran mesin (*machine learning*). *Machine learning* adalah bagian dari *Artificial Intelligence* (AI) yang memiliki kemampuan untuk membuat sistem cerdas komputer tanpa penentuan aturan oleh manusia secara langsung. Namun, sistem ini dibangun untuk mengenali pola dari kumpulan data yang digunakan sebagai data latih yang menghasilkan model untuk memprediksi atau klasifikasi data [3]. Salah satu algoritmanya adalah *support vector machine* (SVM). SVM adalah salah satu algoritma yang banyak digunakan saat ini untuk klasifikasi dan regresi [4]. Dalam penelitian ini, SVM diimplementasikan dalam bentuk *support vector regression* (SVR) karena variabel yang diprediksi berupa data numerik kontinu. SVR memiliki kemampuan untuk menangani masalah linier maupun nonlinier. Dengan menerapkan fungsi kernel, SVR dapat memodelkan pola hubungan yang kompleks, sehingga dapat digunakan untuk menangani beragam variasi data [5].

Beberapa peneliti telah melakukan penelitian tentang prediksi kemiskinan, yaitu dilakukan oleh [6] dengan menggunakan *support vector machine*, *neural network*, dan *stochastic gradient descent* untuk memprediksi angka kemiskinan yang ada di

Kabupaten Ogan Komering Ulu. Hasil pengujian menunjukkan bahwa algoritma SVM memberikan hasil prediksi paling akurat dengan nilai *mean absolute percentage error* (MAPE) sebesar 4.5% serta mampu memprediksi angka kemiskinan untuk tahun 2025 dan 2026 sebesar 72.42 ribu jiwa dan 70.15 ribu jiwa. Penelitian yang dilakukan oleh [7] yang membandingkan prediksi kemiskinan di Indonesia menggunakan SVM dengan regresi linear dan menemukan SVM lebih baik dengan RMSE sebesar 3.141 dibandingkan dengan regresi linear. Penelitian lainnya yang dilakukan oleh [8] yang bertujuan untuk mengklasifikasikan kemiskinan di Papua menggunakan metode SVM dan *K-Nearest Neighbor* (K-NN) dan didapatkan hasil bahwa SVM lebih baik daripada K-NN dengan tingkat akurasi sebesar 93% dan nilai *root mean squared error* (RMSE) sebesar 0.26. Penelitian terkait tersebut menunjukkan bahwa algoritma SVM telah banyak digunakan dalam penelitian mengenai kemiskinan dan terbukti memiliki kinerja yang baik. Namun, masih terdapat celah penelitian (*research gap*). Penelitian sebelumnya melakukan prediksi jumlah penduduk miskin atau melakukan klasifikasi kemiskinan pada wilayah lain atau tingkat nasional, serta menggunakan variabel yang terbatas dan metode pengolahan data yang berbeda. Selain itu, sebagian penelitian belum berfokus pada prediksi tingkat kemiskinan di tingkat kabupaten/kota dan belum mengintegrasikan hasil prediksi ke dalam media yang mudah digunakan oleh pengambil keputusan.

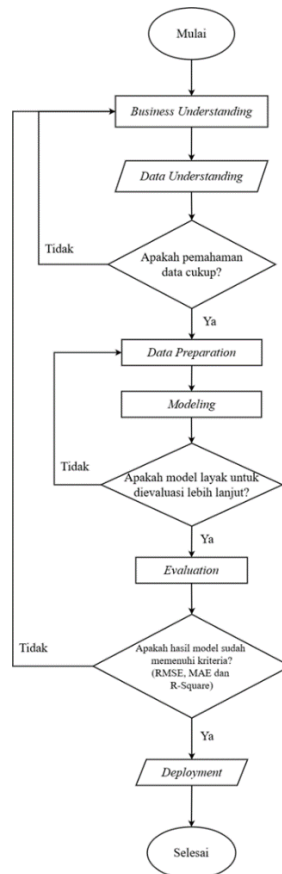
Berdasarkan celah tersebut, *novelty* penelitian ini terletak pada pemodelan dan prediksi tingkat kemiskinan Provinsi Sumatera Barat menggunakan algoritma SVM dengan pendekatan CRISP-DM, yang melibatkan tujuh variabel sosial-ekonomi, serta mengimplementasikan hasil prediksi ke dalam dashboard menggunakan streamlit sebagai dasar pendukung pengambilan keputusan pemerintah daerah.

Penelitian ini bertujuan untuk:

- a. Untuk membangun model prediksi tingkat kemiskinan Provinsi Sumatera Barat menggunakan SVM.
- b. Untuk menilai kinerja model prediksi SVM menggunakan metrik evaluasi, seperti MAE, RMSE, dan R^2 .
- c. Menghasilkan dashboard sederhana menggunakan streamlit yang menampilkan hasil prediksi serta informasi pendukung sehingga dapat membantu pemerintah daerah, khususnya Bappeda dalam proses analisis dan dasar pendukung pengambilan keputusan.

2. METODE PENELITIAN

Penelitian ini mengikuti tahapan *Cross-Industry Standard for Data Mining* (CRISP-DM) yang merupakan sebuah proses penelitian dari data mining. Metode CRISP-DM digunakan untuk menentukan apakah ada hubungan antara aktivitas data mining berdasarkan tujuan, konteks, dan minat pengguna, serta data yang paling penting [9]. Tahapan penelitian dimulai dari *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.



Gambar 3. Diagram Alir CRISP-DM

2.1. Business Understanding

Tujuan dari tahapan ini adalah untuk memberikan informasi tentang tingkat kemiskinan di Provinsi Sumatera Barat. Penelitian ini berfokus pada pemodelan dan prediksi tingkat kemiskinan sebagai dasar untuk pengambilan keputusan yang lebih tepat sasaran. Penelitian ini diharapkan dapat menghasilkan data prediktif yang akurat untuk membantu pemerintah daerah dalam pengambilan keputusan yang lebih strategis, berkelanjutan dan efisien.

2.2. Data Understanding

Pada tahap ini, identifikasi dan analisis karakteristik data yang digunakan dalam studi ini dilakukan. Sumber data diperoleh dari Badan Pusat Statistik (BPS), yang mencakup variabel-variabel sosioekonomi terkait fenomena kemiskinan. Data dikumpulkan dari 19 kabupaten/kota di Provinsi Sumatera Barat selama periode 2015 hingga 2024. Secara keseluruhan, dataset terdiri dari 191 baris dan 8 kolom, dengan rincian variabel sebagai berikut: persentase penduduk miskin, rata-rata lama sekolah, tingkat pengangguran terbuka, pengeluaran per kapita, akses air minum yang layak, harapan hidup, tingkat partisipasi angkatan kerja, dan tingkat melek huruf.

Pada tahap ini, eksplorasi terhadap dataset yang digunakan dalam penelitian juga dilakukan. Struktur, distribusi, dan karakteristik data dianalisis untuk memperoleh pemahaman yang lebih mendalam. Selain itu, kemungkinan adanya nilai yang hilang dan duplikasi diidentifikasi, dan outlier pada variabel numerik dideteksi menggunakan metode *Interquartile Range (IQR)*. Melalui metode IQR, sebaran nilai data relatif terhadap median ditentukan. Semakin besar nilai IQR, semakin tinggi tingkat variabilitas data dalam sebuah dataset. Nilai yang berada di luar batas bawah dan atas yang ditetapkan dapat dikategorikan sebagai *outlier* dalam dataset [10].

2.3. Data Preparation

Sebelum data digunakan untuk pemodelan, tahap *data preparation* melibatkan penanganan missing value dan duplikat, standarisasi serta pemilihan atribut yang paling relevan dengan tujuan penelitian [11]. Variabel yang dipilih sebagai variabel dependen adalah persentase penduduk miskin (P_0) dan variabel sosial-ekonomi lainnya digunakan sebagai variabel independen. Standarisasi dilakukan dengan menggunakan *Z-Score* untuk menyamakan skala setiap fitur. Tujuannya adalah untuk meningkatkan kinerja algoritma *support vector machine (SVM)*, terutama dalam menghitung jarak antar titik data. *Z-Score* adalah metode standarisasi yang digunakan dalam penelitian ini. *Z-Score* dikenal juga sebagai nilai standar atau nilai baku, dimana ukuran penyimpangan data dari nilai rata-rata (μ) diukur dalam satuan standar deviasi (σ) [12]. Berikut ini merupakan rumus untuk menghitung nilai *Z-Score* dari setiap variabel yang digunakan:

$$Z = \frac{x-\mu}{\sigma} \quad (1)$$

Kemudian, untuk memastikan bahwa kinerja model dapat diterapkan pada kedua subset data, data dibagi menjadi data latih 80% dan data uji 20%. Penanganan *outlier* dilakukan dengan menggunakan metode *winsorizing*. Metode *winsorizing* yang merupakan teknik untuk membatasi nilai-nilai ekstrem agar tetap berada dalam rentang tertentu tanpa menghapus data. Batas penyesuaian nilai ditentukan berdasarkan *interquartile range* (IQR), yaitu selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1) [13]. Pendekatan ini membantu menjaga keseluruhan struktur data sekaligus mengurangi pengaruh nilai ekstrem terhadap proses pemodelan dan hasil analisis. Berikut ini cara untuk menentukan batas bawah dan batas atas [13]:

$$\text{Batas bawah} = Q1 - 1.5 \times IQR \quad (2)$$

$$\text{Batas atas} = Q3 + 1.5 \times IQR \quad (3)$$

Selain itu, dalam tahap ini juga dilakukan proses *feature engineering* yang bertujuan untuk meningkatkan kemampuan model dalam melakukan prediksi atau klasifikasi. Proses ini berguna untuk menghasilkan fitur baru atau menyederhanakan hubungan nonlinier antar variabel menjadi bentuk yang lebih mudah dipelajari oleh model [14].

2.4. Modeling

Pada tahap *modeling*, model prediksi dibangun untuk menentukan tingkat kemiskinan di setiap kabupaten/kota Provinsi Sumatera Barat. Prediksi dilakukan dengan menggunakan algoritma SVM yang diimplementasikan dalam bentuk SVR karena variabel target yang diprediksi berupa nilai numerik kontinu. SVR merupakan pengembangan dari SVM yang dirancang khusus untuk menyelesaikan permasalahan regresi dengan mencari fungsi regresi optimal yang memiliki galat sekecil mungkin dan margin maksimum [15].

Dalam penelitian ini, SVR menggunakan *cross-validation grid search* (*GridSearchCV*) yaitu metode pencarian *hyperparameter* secara sistematis dengan memanfaatkan *cross-validation* untuk memperoleh kombinasi parameter terbaik yang mampu meningkatkan kinerja model [16]. *GridSearchCV* secara otomatis menguji berbagai kombinasi parameter, seperti C, epsilon dan gamma yang kemudian memilih konfigurasi dengan performa paling optimal. Parameter C digunakan untuk mengatur tingkat penalti kesalahan dalam proses pelatihan model. Nilai C yang kecil menghasilkan margin yang lebih lebar, sedangkan nilai yang besar menekan kesalahan. Parameter epsilon menentukan batas toleransi kesalahan prediksi, dimana nilai epsilon yang kecil membuat model lebih sensitif terhadap kesalahan dan epsilon besar memberikan toleransi yang lebih luas. Sementara itu, parameter gamma mengontrol pengaruh setiap data latih terhadap model dengan pengaturan “*auto*” dan “*scale*”, dimana “*scale*” menghitung nilai berdasarkan jumlah fitur serta variasi data [17]. Pemilihan kernel rbf bertujuan untuk menangkap hubungan nonlinier antar variabel dalam data. Parameter terbaik ini akan digunakan untuk membangun model akhir. Model yang sudah dilatih digunakan untuk memprediksi tingkat kemiskinan pada data uji dan memperkirakan kondisi kemiskinan di tahun-tahun berikutnya, sehingga menghasilkan informasi yang dapat digunakan dasar pendukung pengambilan keputusan yang lebih tepat sasaran.

2.5. Evaluation

Pada fase evaluasi, kinerja model dianalisis untuk memastikan bahwa model SVM yang dikembangkan mampu memprediksi tingkat kemiskinan 23 65 56 118 20 secara akurat. Penilaian kinerja dilakukan menggunakan metrik *mean absolute error* (MAE) dan *root mean squared error* (RMSE) untuk mengukur tingkat kesalahan prediksi, dengan nilai yang lebih rendah menunjukkan kinerja model yang lebih baik. Selanjutnya, koefisien determinasi (R^2) digunakan untuk menilai sejauh mana model menjelaskan variasi dalam data target; semakin mendekati 1, semakin tinggi kemampuan prediksi model. Selain itu, perbandingan antara nilai aktual dan prediksi pada data uji divisualisasikan untuk mengamati secara langsung tingkat kesesuaian pola yang dihasilkan. Hasil dari proses evaluasi ini kemudian digunakan untuk memastikan bahwa model yang dikembangkan sesuai untuk digunakan pada fase implementasi, khususnya dalam proses prediksi dan sebagai dasar pengambilan keputusan.

2.6. Deployment

Deployment merupakan tahap akhir dalam penelitian ini. Tahap ini melibatkan penerapan model yang telah dibangun ke dalam bentuk sistem yang dapat digunakan. Pada tahap ini, model SVM yang bekerja dengan paling baik diterapkan pada aplikasi prediksi menggunakan *streamlit*. Dashboard ini dibuat untuk menampilkan prediksi tingkat kemiskinan di Provinsi Sumatera Barat supaya menjadi lebih mudah dan interaktif bagi *stakeholder* Badan Perencanaan Pembangunan Daerah. Tahap *deployment* ini memastikan bahwa model yang dikembangkan tidak hanya berhenti pada analisis dan evaluasi saja, namun juga dapat digunakan untuk mendukung proses pengambilan keputusan secara lebih tepat sasaran dan berlandaskan data.

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding

Tahap *business understanding* bertujuan untuk memastikan bahwa proses analisis data yang dilakukan selaras dengan permasalahan yang dihadapi serta tujuan penelitian yang ingin dicapai. Fokus utama pada tahap ini adalah memahami kebutuhan informasi yang ingin dihasilkan serta menentukan peran data dalam mendukung pengambilan keputusan terkait permasalahan kemiskinan di Provinsi Sumatera Barat. Permasalahan yang dihadapi adalah masih adanya perbedaan tingkat kemiskinan antar

kabupaten/kota, meskipun secara umum menunjukkan penurunan tingkat kemiskinan. Kondisi tersebut membutuhkan informasi prediktif yang mampu memberikan gambaran awal mengenai arah perubahan tingkat kemiskinan di masing-masing wilayah. Oleh karena itu, P_0 dipilih sebagai variabel target karena menjadi indikator yang digunakan dalam perencanaan pembangunan daerah. Tujuan dari penelitian ini adalah membangun model prediksi tingkat kemiskinan berdasarkan data sosial ekonomi yang dapat digunakan sebagai dasar pendukung pengambilan keputusan dalam perencanaan dan evaluasi kebijakan penanggulangan kemiskinan. Untuk mencapai tujuan tersebut, penelitian ini menerapkan pendekatan data mining dengan mengikuti tahapan CRISP-DM secara sistematis, sehingga hasil prediksi yang diperoleh diharapkan relevan dan dapat dimanfaatkan dalam proses perencanaan kebijakan di masa mendatang.

3.2. Data Understanding

Tahap *data understanding* bertujuan untuk memahami kualitas data serta karakteristik awal data sebelum dilakukan proses pengolahan lebih lanjut. Pada tahap ini, analisis dilakukan untuk mengidentifikasi kondisi data yang berpotensi mempengaruhi proses pemodelan, seperti *missing value*, duplikasi data, serta *outlier*.

```

=== Info Data ===
<<class 'pandas.core.frame.DataFrame'>
RangeIndex: 190 entries, 0 to 189
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Tahun           190 non-null   int64
1   Kabupaten/Kota  190 non-null   object
2   P0              190 non-null   float64
3   RLS             190 non-null   float64
4   TPT            171 non-null   float64
5   PPK            190 non-null   int64
6   AML            190 non-null   float64
7   UHH            190 non-null   float64
8   TPAK          171 non-null   float64
9   AMH            190 non-null   float64
dtypes: float64(7), int64(2), object(1)
memory usage: 15.0+ KB
    
```

Gambar 4. Deskripsi Data

Hasil load data pada Gambar 4 menunjukkan bahwa dataset memiliki 190 baris dan 10 kolom yang terdiri dari Tahun, Kabupaten/Kota, serta delapan variabel numerik yaitu Persentase Penduduk Miskin (P_0), Rata-Rata Lama Sekolah (RLS), Tingkat pengangguran Terbuka (TPT), Pengeluaran per Kapita (PPK), Akses Terhadap Air Minum Layak (AML), Umur Harapan Hidup (UHH), Tingkat Partisipasi Angkatan Kerja (TPAK), dan Angka Melek Huruf (AMH). Tipe data yang digunakan meliputi integer untuk variabel Tahun dan PPK, objek (*string*) untuk Kabupaten/Kota, serta float untuk variabel numerik lainnya. Tipe data ini menunjukkan bahwa dataset telah terformat dengan baik dan siap untuk dilakukan tahapan analisis lebih lanjut.

Tabel 1. Cek Missing Value

Variabel	Missing Value	Missing Value (%)
P_0	0	0
RLS	0	0
TPT	19	10
PPK	0	0
AML	0	0
UHH	0	0
TPAK	19	10
AMH	0	0

Hasil pemeriksaan pada Tabel 1 menunjukkan bahwa data masih memiliki missing value pada variabel TPT dan TPAK sebanyak 19 data atau 10% dari total data. Kondisi ini menunjukkan data cukup lengkap, namun tetap memerlukan penanganan agar tidak mengganggu proses analisis. Selanjutnya dilakukan pemeriksaan duplikasi data untuk memastikan tidak ada baris data yang tercatat lebih dari satu kali.

```

=== Cek Duplikasi ===
Jumlah data duplikat: 0
    
```

Gambar 5. Cek Duplikat Data

Berdasarkan Gambar 5, jumlah data duplikat adalah 0, yang artinya seluruh data dapat digunakan dalam penelitian ini. Selain itu, dilakukan analisis terhadap keberadaan nilai ekstrem (*outlier*) menggunakan metode IQR.

```

=== Jumlah Outlier per Kolom (IQR) ===
P0: 10 outlier | batas [1.396, 10.906]
RLS: 0 outlier | batas [3.890, 14.830]
TPT: 13 outlier | batas [1.665, 8.745]
PPK: 4 outlier | batas [6052.625, 15385.625]
AML: 0 outlier | batas [31.601, 129.771]
UHH: 0 outlier | batas [60.788, 79.947]
TPAK: 7 outlier | batas [58.875, 78.875]
AMH: 7 outlier | batas [98.964, 100.394]
    
```

Gambar 6. Cek *Outlier*

Hasil analisis pada Gambar 6 menunjukkan adanya *outlier* pada variabel P₀ sebanyak 10 data, TPT sebanyak 13 data, PPK sebanyak 4 data, TPAK sebanyak 7 data, dan AMH sebanyak 7 data. Keberadaan *outlier* ini menunjukkan adanya perbedaan kondisi sosial ekonomi yang cukup besar antarwilayah maupun antarperiode tertentu.

3.3. Data Preparation

Tahap *data preparation* dilakukan untuk mempersiapkan data agar bisa digunakan pada proses pemodelan dan mencegah terjadinya kebocoran data (*data leakage*). Proses ini dimulai dengan *feature engineering*, yaitu membentuk variabel lag satu periode pada persentase penduduk miskin (P₀) untuk setiap kabupaten/kota. Pembentukan variabel ini bertujuan agar model dapat mempertimbangkan pengaruh tingkat kemiskinan pada tahun sebelumnya terhadap kondisi pada tahun berjalan.

```

=== Contoh P0 & P0_lag1 (sebelum drop NaN) ===
Kabupaten/Kota Tahun P0 P0_lag1
0 Agam 2015 7.58 NaN
1 Agam 2016 7.83 7.58
2 Agam 2017 7.59 7.83
3 Agam 2018 6.76 7.59
4 Agam 2019 6.75 6.76
5 Agam 2020 6.75 6.75
6 Agam 2021 6.85 6.75
7 Agam 2022 6.22 6.85
8 Agam 2023 6.60 6.22
9 Agam 2024 6.83 6.60
10 Bukittinggi 2015 5.36 NaN
11 Bukittinggi 2016 5.48 5.36

=== Setelah Drop NaN akibat Lag ===
Jumlah baris sebelum: 190
Jumlah baris sesudah: 171
Baris yang hilang karena lag: 19
    
```

Gambar 7. *Feature Engineering*

Setelah fitur lag terbentuk, baris data pada tahun pertama pengamatan tidak memiliki nilai lag sehingga tidak digunakan dalam analisis selanjutnya. Hal ini mengakibatkan jumlah dataset dari 190 baris berkurang menjadi 171 baris.

Tabel 2. Split Data

Data Latih	Data Uji
133 data	38 data

Pada penelitian ini digunakan data latih sekitar 80% (tahun 2015 - 2022) dan data uji sekitar 20% (tahun 2023 – 2024). Hasil pembagian data pada Tabel 2 menunjukkan bahwa data latih terdiri dari 133 baris, sedangkan data uji terdiri dari 38 baris dengan jumlah fitur sebanyak delapan variabel independen. Pembagian data berdasarkan urutan waktu ini dilakukan agar model hanya mempelajari pola dari periode sebelumnya, sehingga hasil prediksi dapat dievaluasi secara lebih objektif dan model tidak memperoleh informasi dari periode masa depan yang dapat menyebabkan bias dalam proses evaluasi.

Tahap selanjutnya adalah *preprocessing* untuk menangani *missing value*, *outlier*, serta melakukan standarisasi agar dapat digunakan pada tahap pemodelan. Nilai hilang pada beberapa variabel ditangani menggunakan metode imputasi median karena metode ini lebih stabil terhadap keberadaan nilai *outlier*.

```
Missing value TRAIN setelah imputasi:
RLS      0
TPT      0
PPK      0
AML      0
UHH      0
TPAK     0
AMH      0
P0_lag1  0
dtype: int64

Missing value TEST setelah imputasi:
RLS      0
TPT      0
PPK      0
AML      0
UHH      0
TPAK     0
AMH      0
P0_lag1  0
dtype: int64
```

Gambar 8. Mengatasi Missing Value

Hasil pada Gambar 8 tersebut menunjukkan bahwa setelah dilakukan imputasi menggunakan metode median, seluruh variabel pada data latih dan data uji tidak lagi memiliki *missing value*. Selanjutnya, nilai *outlier* ditangani menggunakan metode *winsorizing* dengan menggunakan IQR untuk mengurangi pengaruh nilai yang terlalu menyimpang tanpa menghilangkan data.

```
Ringkasan Min & Max TRAIN setelah winsor (IQR):
           Min      Max
RLS      6.520    11.920
TPT      2.565     8.125
PPK     6213.000 14889.000
AML      35.830   100.000
UHH      64.360   74.820
TPAK     61.150   77.080
AMH      98.900   100.000
P0_lag1  2.010    11.130

=== CEK ULANG OUTLIER (TRAIN) SETELAH WINSOR IQR ===
RLS: 0 outlier
TPT: 0 outlier
PPK: 0 outlier
AML: 0 outlier
UHH: 0 outlier
TPAK: 0 outlier
AMH: 0 outlier
P0_lag1: 0 outlier

Total outlier tersisa: 0
 Tidak terdapat outlier setelah winsorizing.
```

Gambar 9. Mengatasi Outlier

Pada Gambar 9 menunjukkan nilai minimum dan maksimum untuk setiap variabel telah berada dalam batas yang wajar setelah dilakukan proses *winsorizing* menggunakan IQR. Setelah itu, dilakukan standarisasi data menggunakan metode *Z-Score* melalui *StandarScaler* untuk menyamakan skala antar variabel numerik. Standarisasi ini diperlukan agar perbedaan rentang nilai antar variabel tidak memengaruhi kinerja algoritma prediksi.

```
Cek mean & std (TRAIN scaled) -> harus ~0 dan ~1
      mean  std
RLS   -0.0  1.0
TPT    0.0  1.0
PPK    0.0  1.0
AML   -0.0  1.0
UHH    0.0  1.0
TPAK   0.0  1.0
AMH   -0.0  1.0
P0_lag1 -0.0  1.0

5 baris pertama TRAIN setelah scaling:
      RLS      TPT      PPK      AML      UHH      TPAK      AMH \
1 -0.614945 -0.073582 -0.777755 -0.361080  0.528229 -0.003859  0.473016
2 -0.473665 -0.758880 -0.631472 -0.116376  0.575909 -0.827134 -0.433686
3 -0.271836 -0.251252 -0.578134 -0.764194  0.671269 -0.109112  0.019665
4 -0.164194 -0.360029 -0.424458  0.343446  0.795970  0.187902  0.333524
5 -0.090191 -0.483310 -0.492583  0.626366  0.869323  0.389756 -0.084954

      P0_lag1
1  0.539666
2  0.665139
3  0.544685
4  0.128115
5  0.123096
```

Gambar 10. Standarisasi

Standarisasi yang dilakukan menunjukkan setiap variabel pada data latih memiliki nilai rata-rata mendekati 0 dan standar deviasi mendekati 1. Kondisi ini menunjukkan bahwa skala antar variabel menjadi seragam sehingga model dapat mempelajari pola data secara lebih seimbang. Data yang telah melalui seluruh tahapan persiapan kemudian disimpan sebagai dataset bersih dan digunakan pada tahap pemodelan, evaluasi, serta proses prediksi selanjutnya.

3.4. Modeling

Pemodelan dimulai dengan pelatihan model awal untuk memastikan bahwa pipeline dan algoritma dapat berjalan dengan baik. Selanjutnya, dilakukan optimasi parameter menggunakan *GridsearchCV* dengan validasi *TimeSeriesSplit* sebanyak lima pembagian waktu, sehingga urutan waktu data tetap terjaga selama proses validasi.

Tabel 3. Parameter Terbaik

C	Epsilon	Gamma
1000	0.05	0.001

Hasil optimasi menunjukkan bahwa kombinasi parameter terbaik yang diperoleh adalah nilai C sebesar 1000, epsilon sebesar 0.05, dan gamma sebesar 0.001 dengan kernel rbf yang digunakan untuk menangkap hubungan nonlinier antar variabel. Nilai parameter terbaik pada Tabel 3 tersebut memberikan gambaran mengenai bagaimana model bekerja dalam mempelajari pola data. Nilai C sebesar 1000 menunjukkan bahwa model berusaha dalam menekan kesalahan prediksi pada data latih. Artinya, model berusaha agar selisih antara nilai prediksi dan nilai aktual sekecil mungkin sehingga hasil yang diperoleh lebih mendekati kondisi sebenarnya. Nilai epsilon sebesar 0.05 menunjukkan batas toleransi kesalahan yang diperbolehkan relatif kecil agar model tetap memperhatikan detail variasi data dan tidak terlalu banyak mengabaikan penyimpangan kecil. Sementara itu, nilai gamma sebesar 0.001 pada kernel rbf menunjukkan bahwa model tidak terlalu sensitif terhadap perubahan kecil pada data. Model dengan parameter tersebut kemudian diterapkan sebagai model akhir dan disimpan untuk digunakan pada tahap evaluasi dan prediksi tingkat kemiskinan.

3.5. Evaluation

Pada tahap evaluation, kinerja model prediksi dievaluasi untuk menilai kemampuan model dalam memprediksi tingkat kemiskinan secara akurat pada data yang belum pernah dilihat sebelumnya. Evaluasi dilakukan menggunakan data uji tahun 2023-2024 dengan membandingkan nilai prediksi P_0 terhadap nilai aktual. Untuk memperoleh gambaran kinerja model yang lebih menyeluruh, digunakan beberapa metrik evaluasi, yaitu MAE, RMSE, dan R^2 .

Tabel 4. Metrik Evaluasi

MAE	RMSE	R^2
0.32	0.36	0.98

Berdasarkan hasil evaluasi, tingkat akurasi model dinyatakan sangat baik. Nilai MAE sebesar 0.32 diartikan sebagai perbedaan absolut rata-rata antara nilai aktual dan nilai prediksi yang relatif kecil. Nilai RMSE sebesar 0.36 dipahami sebagai indikator bahwa prediksi yang dihasilkan oleh model relatif stabil. Selain itu, nilai R^2 sebesar 0.98 menunjukkan bahwa sebagian besar variasi dalam data tingkat kemiskinan dapat dijelaskan oleh model, sehingga memungkinkan pola perubahan P_0 dapat dipelajari secara efektif.

Berdasarkan hasil pada Tabel 4, analisis menunjukkan bahwa model prediksi yang dikembangkan telah mencapai tingkat kinerja yang sangat tinggi. Model ini mampu menghasilkan prediksi tingkat kemiskinan dengan tingkat akurasi yang mendekati nilai aktual, didukung oleh tingkat kesalahan yang sangat rendah dan kemampuan yang kuat untuk menjelaskan variasi data. Dengan karakteristik ini, model dianggap sangat layak untuk diimplementasikan dan dapat berfungsi sebagai dasar yang andal untuk mendukung proses pengambilan keputusan dalam konteks perencanaan kebijakan pengurangan kemiskinan.

3.6. Deployment

Pada tahap *deployment*, algoritma SVR yang telah melalui proses pelatihan dan evaluasi diterapkan untuk menghasilkan prediksi tingkat kemiskinan pada tahun 2025 hingga 2030 di seluruh kabupaten/kota di Provinsi Sumatera Barat berdasarkan pola historis data dan perkembangan variabel sosial ekonomi.

Tabel 5. Prediksi Tingkat Kemiskinan Provinsi Sumatera Barat

Tahun	Prediksi Provinsi
2025	5.68
2026	5.46
2027	5.25
2028	5.05
2029	4.86
2030	4.69

Hasil prediksi menunjukkan penurunan tingkat kemiskinan dari tahun ke tahun, yaitu dari 5.68% pada tahun 2025 menjadi 4.69% pada tahun 2030. Penurunan bertahap ini mengindikasikan bahwa secara umum kondisi sosial ekonomi provinsi diperkirakan mengalami perbaikan dalam lima tahun ke depan berdasarkan pola historis yang dipelajari oleh model. Namun, terdapat beberapa kabupaten/kota yang mengalami kenaikan tingkat kemiskinan yaitu Bukittinggi, Kota Solok, Padang, dan Payakumbuh yang memerlukan perhatian lebih lanjut. Sementara itu, kabupaten/kota yang diperkirakan mengalami penurunan tingkat kemiskinan yaitu Agam, Dharmasraya, Kabupaten Solok, Kepulauan Mentawai, Lima Puluh Kota, Padang Panjang, Padang Pariaman, Pariaman, Pasaman, Pasaman Barat, Pesisir Selatan, Sawahlunto, Sijunjung, Solok Selatan, dan Tanah Datar. Hasil prediksi tingkat kemiskinan untuk kabupaten/kota di Sumatera Barat dapat dilihat pada dashboard web prediksi menggunakan streamlit. Dashboard ini menampilkan grafik tren perubahan tingkat kemiskinan, tabel prediksi tahunan, serta ringkasan kinerja model berupa nilai MAE, RMSE, dan R^2 . Pengguna dapat memilih kabupaten/kota tertentu untuk melihat kondisi dan perkiraan kemiskinan secara lebih spesifik. Tampilan dashboard prediksi dapat dilihat diakses melalui link <https://prediksi-tingkat-kemiskinan-sumbar.streamlit.app/>.

Dashboard ini juga dilengkapi dengan visualisasi variabel yang berpengaruh terhadap prediksi, sehingga dapat membantu pengguna memahami faktor-faktor utama yang memengaruhi perubahan tingkat kemiskinan. Dengan demikian, tahap deployment memastikan bahwa model yang dikembangkan tidak hanya bersifat analitis, tetapi juga dapat dimanfaatkan secara langsung sebagai dasar pengambilan keputusan berlandaskan data dalam perencanaan kebijakan penanggulangan kemiskinan.

4. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan mengenai prediksi tingkat kemiskinan di Provinsi Sumatera Barat menggunakan algoritma SVM, maka dapat disimpulkan:

- a. Model prediksi tingkat kemiskinan yang dibangun menggunakan SVM dengan pendekatan SVR mampu mempelajari hubungan antara variabel sosial ekonomi dan P_0 dengan sangat baik. Hal ini ditunjukkan oleh nilai evaluasi model yang tinggi, khususnya nilai R^2 sebesar 0.98 serta nilai MAE sebesar 0.32 dan RMSE sebesar 0.36 yang relatif rendah, sehingga model memiliki tingkat akurasi yang baik dalam memprediksi tingkat kemiskinan pada data uji.
- b. Penggunaan *feature engineering* berupa variabel lag satu periode (P_0 _lag1) terbukti membantu model dalam menangkap pola perubahan tingkat kemiskinan dari waktu ke waktu. Pendekatan ini sesuai dengan karakteristik kemiskinan yang bersifat dinamis dan dipengaruhi oleh kondisi pada periode sebelumnya, sehingga menghasilkan prediksi yang lebih stabil dan realistis.
- c. Hasil prediksi tingkat kemiskinan untuk periode tahun 2025 hingga 2030 menunjukkan kecenderungan penurunan persentase penduduk miskin di kabupaten/kota Provinsi Sumatera Barat. Informasi ini dapat dimanfaatkan sebagai gambaran awal dalam mendukung perencanaan dan pendukung pengambilan keputusan yang lebih terarah dan berlandaskan data.
- d. Implementasi model prediksi ke dalam dashboard interaktif menggunakan streamlit memudahkan penyajian dan pemahaman hasil analisis. Dashboard ini memungkinkan pengguna untuk melihat tren kemiskinan, hasil prediksi tahunan, serta kinerja model secara visual dan informatif, sehingga hasil penelitian tidak hanya bersifat akademis tetapi juga dapat digunakan secara praktis oleh pemangku kepentingan.

DAFTAR PUSTAKA

- [1] H. Mukhtar, R. Gunawan, A. Hariyanto, and W. Mulyana, "Peramalan Kedatangan Wisatawan ke Suatu Negara Menggunakan Metode Support Vector," *J. Comput. Sci. Inf. Technol.*, vol. 3, no. 3, pp. 274–282, 2022, doi: 10.37859/coscitech.v3i3.4211.
- [2] Syaharuddin, E. Pujiana, I. P. Sari, V. M. Mardika, and M. Putri, "Analisis Algoritma Back Propagation Dalam Prediksi Angka Kemiskinan di Indonesia," *Pendek. J. Pendidik. Berkarakter*, vol. 3, no. 1, pp. 11–17, 2020.
- [3] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python," *Bina Insa. ICT J.*, vol. 7, no. 2, pp. 156–165, 2020.
- [4] M. R. Qisthiano, "Klasifikasi Terhadap Prediksi Kelulusan Mahasiswa Dengan Menggunakan Metode Support Vector Machine (SVM)," in *Proceeding STEKOM 2022*, 2022, pp. 204–208.
- [5] Herlangga, M. W. Pangestika, and S. P. A. Alkadri, "Prediksi Harga Mobil Bekas Menggunakan Algoritma Support Vector Regression," *J. Comput. Sci. Inf. Technol.*, vol. 6, no. 3, pp. 399–404, 2025, doi: 10.37859/coscitech.v6i3.10545.
- [6] D. Kurniawan, S. Apriani, and Pujianto, "Prediksi Angka Kemiskinan Di Kabupaten Ogan Komering Ulu Menggunakan Metode Data Mining Berbasis SVM, Neural Network & SGD," *Semin. Nas. Teknol. Inf. dan Bisnis*, pp. 133–138, 2025.
- [7] A. Karim, "Perbandingan Prediksi Kemiskinan di Indonesia Menggunakan Support Vector Machine (SVM) dengan Regresi Linear," *J. Sains Mat. dan Stat.*, vol. 6, no. 1, pp. 107–112, 2020, doi: 10.24014/jsms.v6i1.9259.

- [8] Fauziah, M. A. Tiro, and Ruliana, "Comparison of k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) Methods for Classification of Poverty Data in Papua," *J. Math. Appl. Sci.*, vol. 2, no. 2, pp. 83–91, 2022, doi: 10.35877/mathscience741.
- [9] F. H. Zufallah, "Implementasi Algoritma KNN Dalam Mengukur Ketepatan Kelulusan Mahasiswa UIN Syarif Hidayatullah Jakarta," Universitas Islam Negeri Syarif Hidayatullah, 2022.
- [10] R. Efendi, A. Junaidi, and A. M. Rizki, "Penentuan Pusat Kluster Secara Otomatis Pada Algoritma Density Peaks Clustering Berbasis Metode Inter Quartile Range," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4997.
- [11] A. E. Putri, Y. MZ, and J. E. Bororing, "Implementasi K-Means Clustering dan Model CRISP-DM Untuk Pengelompokan Daerah Rawan Tindak Pidana Narkoba Di DIY," *J. Inform. Teknol. dan Sains*, vol. 7, no. 2, pp. 606–615, 2025.
- [12] Hendriyana, I. M. K. Karo, and S. Dewi, "Analisis Perbandingan Algoritma Support Vector Machine, Naive Bayes, dan Regresi Logistik Untuk Memprediksi Donor Darah," *J. Teknol. Terpadu*, vol. 8, no. 2, pp. 121–126, 2022, [Online]. Available: <https://journal.nurulfikri.ac.id/index.php/jtt>
- [13] E. Aiziyah, A. H. Lukmana, and A. M. Abadi, "Prediksi Rate of Penetration pada Pengeboran Minyak Bumi dengan Elman Recurrent Neural Network," *Multimed. Artif. Intell. Netw. Database J.*, vol. 10, no. 2, pp. 145–161, 2025, doi: 10.26760/mindjournal.v10i2.145-161 |.
- [14] C. M. Sitorus, A. Rizal, and M. Jajuli, "Prediksi Risiko Perjalanan Transportasi Online Dari Data Telematik Menggunakan Algoritma Support Vector Machine," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 254–265, 2020, doi: 10.28932/jutisi.v6i2.2672.
- [15] E. Rahmawati, "Penerapan Metode Support Vector Regression (SVR) Menggunakan Kernel Linear, Polinomial, dan Radial Dengan Grid Search Optimization," Universitas Lampung, 2023.
- [16] Z. Maisat, E. Darmawan, and A. Fauzan, "Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM Implementation of GridSearchCV Hyperparameter Optimization in Heart Attack Prediction System Using SVM," *J. Ilm. Sist. Inf.*, vol. 13, no. 1, pp. 8–15, 2023, doi: 10.26594/teknologi.v13i1.3098.
- [17] W. Wulandari, "Metode SVR Untuk Memprediksi Jumlah Pemakaian Air PDAM (Studi Kasus : Sub Unit Muntilan)," Universitas Muhammadiyah Magelang, 2025.