

# Evaluasi Pengaruh RFE Terhadap Kinerja Random Forest dengan SVM pada Klasifikasi Kemiskinan Kabupaten/Kota Indonesia

Shafa Kirana Aralia, Mula Agung Barata, Ita Aristia Sa'ida\*

Sains dan Teknologi, Teknik Indormatika, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia

Email: <sup>1</sup>shafaaralia@gmail.com, <sup>2</sup>mula.ab@gmail.com, <sup>3</sup>\*itaaristia@unugiri.ac.id

Email Penulis Korespondensi: shafaaralia@gmail.com

Submitted 22-01-2026; Accepted 19-02-2026; Published 28-02-2026

## Abstrak

Kemiskinan merupakan persoalan sosial ekonomi yang masih menjadi perhatian di Indonesia, dengan perbedaan karakteristik pembangunan antar kabupaten/kota yang menyebabkan variasi indikator yang luas dan saling berkorelasi. Redundansi fitur serta keberadaan nilai ekstrem berpotensi menurunkan kemampuan generalisasi model klasifikasi dan mengurangi interpretabilitas hasil. Oleh karena itu, diperlukan pendekatan yang tidak hanya menghasilkan akurasi tinggi, tetapi juga mampu mengidentifikasi indikator yang paling relevan. Penelitian ini bertujuan untuk mengevaluasi pengaruh Recursive Feature Elimination (RFE) terhadap kinerja Support Vector Machine (SVM) dan Random Forest dalam mengklasifikasikan status kemiskinan kabupaten/kota di Indonesia. Dataset yang digunakan terdiri dari 514 observasi dengan dua kelas target, yaitu tidak miskin dan miskin. Tahap pra-proses meliputi pembersihan data dan penanganan outlier menggunakan metode IQR capping, kemudian data dibagi menjadi 80% data latih dan 20% data uji. Pengujian dilakukan pada empat skenario: SVM, SVM+RFE, Random Forest, dan Random Forest+RFE. Evaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Hasil menunjukkan bahwa RFE tidak mengubah akurasi SVM (0,971), namun meningkatkan kinerja Random Forest dari 0,981 menjadi 0,99 serta memperbaiki presisi kelas minoritas. Kombinasi Random Forest+RFE menjadi konfigurasi paling efektif dan efisien untuk klasifikasi kemiskinan wilayah.

**Kata Kunci:** Kemiskinan; Klasifikasi; Recursive Feature Elimination; Random Forest; Support Vector Machine

## Abstract

Poverty is a socio-economic issue that remains a concern in Indonesia, with differences in development characteristics between districts/cities causing wide variations in indicators that are intercorrelated. Feature redundancy and the existence of extreme values have the potential to reduce the generalization ability of classification models and reduce the interpretability of results. Therefore, an approach is needed that not only produces high accuracy but is also capable of identifying the most relevant indicators. Therefore, an approach is needed that not only produces high accuracy but is also capable of identifying the most relevant indicators. This study aims to evaluate the effect of Recursive Feature Elimination (RFE) on the performance of Support Vector Machine (SVM) and Random Forest in classifying the poverty status of districts/cities in Indonesia. The dataset used consists of 514 observations with two target classes, namely non-poor and poor. The preprocessing stage included data cleaning and outlier handling using the IQR capping method, then the data was divided into 80% training data and 20% test data. Testing was conducted on four scenarios: SVM, SVM+RFE, Random Forest, and Random Forest+RFE. Evaluation used a confusion matrix, accuracy, precision, recall, and F1-score. The results show that RFE does not change the accuracy of SVM (0.971), but improves the performance of Random Forest from 0.981 to 0.99 and improves the precision of the minority class. The Random Forest+RFE combination is the most effective and efficient configuration for regional poverty classification.

**Keywords:** Classification; Poverty; Random Forest; Recursive Feature Elimination; Support Vector Machine

## 1. PENDAHULUAN

Kemiskinan merupakan permasalahan multidimensi yang menjadi tantangan utama pembangunan global dan nasional. Meskipun berbagai program pengentasan telah dilakukan, ketimpangan sosial ekonomi antarwilayah masih relatif tinggi, terutama pada tingkat kabupaten/kota. Permasalahan utama dalam penelitian ini terletak pada kesulitan mengklasifikasikan tingkat kemiskinan secara akurat akibat banyaknya indikator sosial ekonomi yang saling berkorelasi, memiliki rentang nilai berbeda, serta berpotensi menimbulkan redundansi fitur. Pendekatan konvensional atau berbasis aturan sederhana sering kali tidak mampu menangkap pola kompleks antarvariabel, sehingga menghasilkan klasifikasi yang kurang optimal [1].

Di Indonesia, tingginya tingkat kemiskinan dipengaruhi oleh sejumlah faktor, termasuk pendidikan, kesehatan, dan pengangguran, yang mempengaruhi kondisi kemiskinan di Masyarakat [2]. Meskipun data Badan Pusat Statistik (BPS) menunjukkan tren penurunan kemiskinan secara nasional, ketimpangan ekonomi antarwilayah masih relatif tinggi [3]. Seseorang dikategorikan sebagai miskin jika tidak mampu memenuhi kebutuhan dasar yang wajar, termasuk pakaian, makanan, tempat tinggal, dan aspek sosial lainnya [4]. Kemiskinan yang tinggi memiliki dampak negatif, seperti antara lain terbatasnya akses anak-anak terhadap pendidikan yang berkualitas, kesulitan dalam pembiayaan layanan kesehatan, rendahnya kapasitas menabung, hambatan dalam memperoleh layanan publik, minimnya perlindungan sosial bagi keluarga, hingga meningkatnya arus urbanisasi [5]. Oleh karena itu, penting untuk memahami pengelompokan dan kategorisasi tingkat kemiskinan di tingkat kabupaten atau kota, mengingat kebijakan dan distribusi bantuan sering disesuaikan dengan kondisi lokal; pendekatan manual atau yang menggunakan aturan sederhana umumnya tidak cukup efektif dalam mengidentifikasi pola dan hubungan yang kompleks antara variabel-variabel sosio-ekonomi [6]. Oleh karena itu, diperlukan metode yang mampu mengkategorikan tingkat kemiskinan di berbagai provinsi di Indonesia [5]. Beberapa studi terbaru menunjukkan bahwa penggunaan algoritma klasifikasi dalam *data mining* dapat meningkatkan efektivitas analisis terkait kemiskinan.

*Data mining* merupakan suatu proses pengambilan serta menemukan pola atau informasi yang sebelumnya tidak terlihat dari kumpulan data besar menggunakan teknik statistik, kecerdasan buatan, dan pembelajaran mesin. Tujuan dari proses ini adalah untuk mengungkap atau memperoleh wawasan dan pengetahuan dari data yang tersedia [7]. Pendekatan ini mampu meningkatkan akurasi pengelompokan sosial melalui penerapan algoritma klasifikasi dan metode *ensemble*. Teknik *data mining* meliputi asosiasi, pengelompokan, klasifikasi, dan regresi, dengan klasifikasi sebagai metode yang paling umum digunakan [8]. Klasifikasi bertujuan membangun model yang mampu merepresentasikan dan membedakan kategori data sehingga dapat memprediksi kelas objek yang belum diketahui [3]. Di bidang analisis data, dua algoritma yang sering digunakan adalah SVM dan *Random Forest*. Kemajuan dalam teknologi data dan perkembangan di bidang *machine learning* telah mendorong penerapan metode inovatif dalam klasifikasi kemiskinan, termasuk melalui algoritma-algoritma ini. Misalnya, algoritma *Random Forest* telah digunakan untuk klasifikasi status kemiskinan rumah tangga di beberapa studi dan menunjukkan kinerja yang menjanjikan [9]. Sementara itu, SVM juga terbukti efektif dalam mengalokasikan objek ke dalam kelas berdasarkan kriteria yang telah ditentukan [10].

Di antara algoritma klasifikasi yang paling sering digunakan, SVM dan *Random Forest* menempati posisi penting karena karakteristik komplementer keduanya. SVM efektif ketika pemisahan kelas dapat dibentuk melalui margin maksimum pada ruang fitur (termasuk ketika diperlukan kernel untuk pemetaan non-linear), sehingga sering dipilih pada masalah klasifikasi dengan struktur batas keputusan yang kompleks. Sementara itu, *Random Forest* sebagai metode *ensemble* berbasis pohon keputusan umumnya kuat terhadap non-linearitas dan interaksi fitur, serta cenderung stabil karena menggabungkan banyak pohon hasil bootstrap dan pemilihan fitur acak pada setiap split. Perbandingan dua algoritma ini menjadi menarik pada kasus kemiskinan kabupaten/kota karena data indikator umumnya mengandung korelasi antar variabel, rentang nilai yang bervariasi, serta potensi ketidakseimbangan kelas.

Pada penelitian Mukharyahya dkk. tentang perbandingan tingkat kemiskinan Kab/Kota di Indonesia dengan pendekatan klasifikasi menempatkan *Naive Bayes* dan SVM sebagai dua algoritma utama yang diuji berdampingan, sehingga memberikan gambaran yang jelas mengenai trade off antara model probabilistik yang sederhana dan model margin-based yang cenderung lebih kuat pada pemisahan kelas yang kompleks [11]. Studi tersebut juga menekankan isu ketidakseimbangan kelas dan menunjukkan bahwa strategi penyeimbangan (*oversampling*) dapat mempengaruhi metrik kinerja, khususnya pada sensitivitas/*recall* untuk kelas tertentu [11]. Walaupun temuan ini menguatkan relevansi SVM untuk klasifikasi kemiskinan, penelitian tersebut belum memposisikan seleksi fitur sebagai variabel eksperimen utama, padahal indikator sosial ekonomi sering bersifat redundan (berkorelasi tinggi) sehingga berpotensi mempengaruhi margin SVM maupun struktur split pada *Random Forest*.

Pada penelitian Khalik dan Arifin yang mengangkat klasifikasi Indeks Kedalaman Kemiskinan di Provinsi Sulawesi Selatan menggunakan algoritma *Random Forest* memfokuskan klasifikasi pada indikator kemiskinan dan menguji pendekatan *Random Forest* dalam kerangka evaluasi klasifikasi [3]. Penelitian ini penting karena mengindikasikan bahwa model berbasis *ensemble* seperti *Random Forest* layak digunakan pada masalah kemiskinan yang indikatornya dapat memiliki hubungan non-linear. Meskipun demikian, fokus penelitian tersebut masih terbatas pada pencapaian metrik klasifikasi, sementara analisis terhadap pemilihan subset fitur yang benar-benar esensial serta dampak pengurangan fitur terhadap kinerja model belum menjadi perhatian utama. Padahal, pada level kabupaten/kota, indikator kemiskinan dan pembangunan sering saling tumpang tindih, sehingga eliminasi fitur yang kurang informatif berpotensi meningkatkan generalisasi dan efisiensi model.

Konteks indikator pembangunan yang berkaitan dengan kemiskinan juga muncul pada penelitian Ressa dkk. dengan topik komparasi *Random Forest*, *AdaBoost*, dan SVM pada klasifikasi Indeks Pembangunan Manusia [12]. Walaupun target klasifikasinya adalah kategori IPM, penelitian ini relevan secara metodologis karena IPM dan indikator kemiskinan berada dalam spektrum yang sama, yakni indikator sosial ekonomi wilayah. Studi tersebut memperlihatkan praktik umum yang membandingkan beberapa algoritma kuat (*Random Forest*, *boosting*, SVM) untuk klasifikasi indikator pembangunan [12]. Dalam konteks tersebut, seleksi fitur menjadi aspek metodologis yang krusial, karena keberadaan fitur yang tidak relevan atau redundan pada data sosial ekonomi kabupaten/kota dapat meningkatkan risiko *overfitting*, menurunkan stabilitas model, dan mengurangi kejelasan interpretasi indikator utama. Salah satu teknik seleksi fitur yang banyak digunakan adalah RFE, yaitu metode *wrapper* yang secara iteratif mengeliminasi fitur berdasarkan bobot kepentingan yang dihasilkan model.

Pada penelitian Priyatno dan Widiyaningtyas yang menyusun sebuah *systematic literature review* mengenai algoritma *Recursive Feature Elimination* (RFE), RFE diposisikan sebagai teknik seleksi fitur yang bekerja secara iteratif dengan mengeliminasi fitur paling tidak relevan secara bertahap hingga diperoleh subset fitur yang lebih ringkas dan informatif. Studi tersebut menguatkan argumen bahwa RFE bukan hanya alat pengurang dimensi, tetapi juga instrumen untuk meningkatkan kualitas generalisasi model ketika fitur awal mengandung noise atau redundansi [13]. Selain itu, Pada penelitian Aqmar dkk. mengenai analisis kinerja model pembelajaran mesin pada klasifikasi data tidak seimbang, *Random Forest* dan *LightGBM* dievaluasi dalam sebuah pipeline terpadu yang menggabungkan seleksi fitur *Recursive Feature Elimination*, penalaan hiperparameter melalui Bayesian Optimization, serta strategi penanganan ketidakseimbangan kelas berupa *random undersampling*, *random oversampling*, dan *SMOTENC*, dengan tujuan utama mengklasifikasikan kerawanan pangan rumah tangga di Papua sekaligus menjelaskan determinan pentingnya secara transparan menggunakan SHAP. Secara konseptual, studi ini penting karena menegaskan dua hal yang pertama, RFE dapat diposisikan sebagai komponen inti pipeline dan yang kedua, efek RFE dapat bersifat bergantung pada model (model-dependence), sehingga perlu diuji secara eksplisit pada algoritma yang berbeda, bukan diasumsikan selalu meningkatkan kinerja [14]. Hal ini sangat relevan untuk kasus kemiskinan kabupaten/kota, karena distribusi kelas tingkat

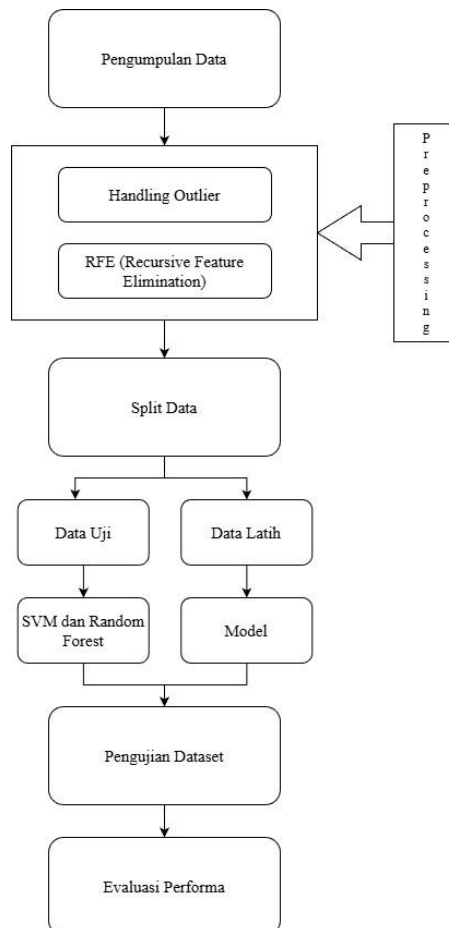
kemiskinan juga kerap tidak seimbang, dan indikator wilayah bisa menghasilkan struktur keputusan yang berbeda antara SVM dan *Random Forest*.

Berdasarkan paparan tersebut, dapat disimpulkan bahwa riset terkait klasifikasi kemiskinan dan indikator pembangunan pada dataset wilayah telah berkembang melalui komparasi algoritma SVM, *Random Forest*, dan metode *ensemble* lainnya [11] [12]. Namun, terdapat celah yang menonjol dan belum banyak penelitian yang secara spesifik menguji kinerja RFE sebagai variabel utama pada dua algoritma yang berbeda karakter (SVM vs *Random Forest*) untuk klasifikasi tingkat kemiskinan kabupaten/kota, termasuk analisis *trade-off* antara kenaikan atau penurunan metrik kinerja (*accuracy, precision, recall, F1*) pengurangan jumlah fitur, dan konsistensi indikator terpilih sebagai penjelas kemiskinan. Karena itu, penelitian ini bertujuan untuk mengevaluasi pengaruh penerapan *Recursive Feature Elimination* terhadap kinerja algoritma SVM dan *Random Forest* dalam klasifikasi tingkat kemiskinan kabupaten/kota di Indonesia. Penelitian ini diharapkan dapat menghasilkan model klasifikasi yang lebih efisien dan akurat serta mengidentifikasi indikator kemiskinan yang paling informatif sebagai dasar pengambilan kebijakan pembangunan wilayah.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian merupakan rangkaian langkah sistematis yang dirancang untuk menggambarkan proses penelitian mulai dari perencanaan awal hingga penyusunan kesimpulan akhir, guna memastikan bahwa tujuan penelitian secara sistematis dan akurat [15]. Penelitian ini berfokus pada penentuan metode yang paling sesuai untuk mengklasifikasikan tingkat kemiskinan kabupaten/kota di Indonesia. Penelitian ini menggunakan metode kuantitatif eksperimental-komparatif. Rangkaian tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

### 2.2 Pengumpulan Data

Data yang digunakan merupakan dataset yang diperoleh dari platform berbagi data yaitu Kaggle, yang merupakan situs populer penyedia dataset untuk keperluan analisis pembelajaran, yang dapat di akses pada <https://www.kaggle.com/datasets/ldausl/klasifikasi-tingkat-kemiskinan-di-indonesia>. Dataset pada penelitian ini berjumlah 514 data dari berbagai kabupaten/kota dan memuat sejumlah variabel sosio-ekonomi. Di antara fitur tersebut terdapat persentase penduduk miskin, pengeluaran per kapita, indeks pembangunan manusia, dan tingkat pengangguran terbuka. Selanjutnya, data diklasifikasikan kedalam dua kelas, yaitu tidak miskin (0) dan miskin (1) [11].

**Tabel 1.** Dataset Kemiskinan Indonesia

No	Kab/ Kota	PO	RLS	PPK	IPM	UHH	SL	AML	TPT	TPAK	PDRB
1	Simeulue	18.98	9.48	7148	66.41	65.28	71.56	87,45	5.71	71.15	1648096
2	Singkil	20.36	8.68	8776	69.22	67.43	69.56	78,58	8.36	62.85	1780419
3	Aceh Selatan	13.18	8.88	8180	67.44	64.4	62.55	79,65	6.46	60.85	4345784
4	Aceh Tenggara	14.45	9.67	8030	69.44	68.22	66.71	86,71	6.43	69.62	3487157
5	Aceh Timur	15.26	8.21	8577	67.83	68.74	66.75	83,16	7.13	59.48	8433526
...	...	...	...	...	...	...	...	...	...	...	...
513	Deiyai	40.59	3.25	4673	49.96	65.36	0	85,23	0,79	85.01	841296
514	Kota Jayapura	11.39	11.57	14937	80.11	70.52	85.31	97,1	11,67	63.75	22852202

Keterangan :

- PO : Persentase penduduk miskin menurut Kab/Kota  
RLS : Rata- rata lama sekolah penduduk 15+ (Tahun)  
PPK : Pengeluaran per kapita disesuaikan (Ribu rupiah/ orang/ tahun)  
IPM : Indeks Pembangunan manusia  
UHH : Umur harapan hidup  
SL : Persentase rumah tangga yang memiliki akses terhadap sanitasi layak  
AML : Persentase rumah tangga yang memiliki akses terhadap air minum layak  
TPT : Tingkat pengangguran terbuka  
TPAK : Tingkat partisipasi Angkatan kerja  
PDRB : PDRB atas dasar harga konstan menurut pengeluaran (rupiah)

### 2.3 Preprocessing

Tahap *preprocessing* pada penelitian ini dirancang untuk memastikan dataset berada dalam kondisi konsisten, bersih, dan stabil sebelum memasuki tahap seleksi fitur RFE dan pembentukan model klasifikasi (SVM dan *Random Forest*). Fokus utama *preprocessing* dalam penelitian ini adalah *handling outlier* menggunakan pendekatan *winsorizing/capping*, yaitu membatasi nilai ekstrem agar tidak mendominasi distribusi tanpa menghapus record. Pendekatan ini dipilih karena lebih aman untuk data kemiskinan ukuran sampel tetap terjaga dan proporsi kelas tidak berubah, namun efek leverage dari nilai ekstrem dapat ditekan. Strategi *winsorizing* sebagai teknik praproses untuk menangani *outlier* juga digunakan dalam studi prediksi kemiskinan berbasis data-driven sebelum pelatihan model, sehingga relevan sebagai dasar metodologis [16].

### 2.4 Seleksi Fitur Dengan Recursive Feature Elimination (RFE)

Metode pemilihan fitur yang secara berulang kali menghilangkan fitur yang paling tidak penting dari dataset berdasarkan kepentingannya [17]. Secara konsep, RFE merupakan metode seleksi fitur yang bekerja dengan cara melatih model pada seluruh fitur, mengukur kontribusi/tingkat kepentingan setiap fitur, menghapus fitur dengan kontribusi terendah, mengulangi proses hingga tersisa sejumlah fitur terbaik. Karakter “rekursif” merujuk pada eliminasi bertahap sampai fitur yang tersisa dianggap paling relevan terhadap target [18] [13]. Dalam implementasi eksperimen, RFE dipakai sebagai komponen untuk memilih fitur yang paling berpengaruh terhadap prediksi, sehingga fitur yang kurang informatif dapat dieliminasi secara iteratif [18].

$$r_j = |w_j| \quad (1)$$

Keterangan :

- $r_j$  : nilai kepentingan (ranking) fitur ke-  $j$   
 $w_j$  : bobot fitur ke-  $j$  yang dihasilkan oleh model  
| | : nilai absolut

### 2.5 Split Data

Tahap berikutnya adalah pemisahan dataset menjadi data latih dan data uji. Skema yang digunakan adalah 80% data latih dan 20% data uji dengan *random sampling* sehingga evaluasi dilakukan pada data yang tidak dilihat saat pelatihan [9]. Secara procedural data latih digunakan untuk membangun model (termasuk fitting RFE pada skenario RFE). data uji hanya digunakan untuk evaluasi akhir agar estimasi performa mencerminkan generalisasi.

### 2.6 Penerapan Model Klasifikasi

#### 2.6.1 Support Vector Machine (SVM)

*Support Vector Machine* adalah algoritma klasifikasi berbasis pendekatan diskriminatif yang dirancang untuk menentukan batas pemisah paling optimal antar kelas, yang dikenal sebagai *hyperplane*. Titik-titik data yang posisinya paling dekat

dengan *hyperplane* tersebut disebut *support vectors*, dan kontribusi pemisahan kelas direpresentasikan melalui bobot yang berasal dari subset sampel yang berperan membatasi kompleksitas permasalahan [19]. Dalam konteks klasifikasi tingkat kemiskinan, SVM berperan mempelajari batas keputusan berdasarkan fitur sosio-ekonomi. Jika pemisahan linier sulit, SVM dapat diperluas melalui fungsi kernel untuk memproyeksikan data ke ruang berdimensi lebih tinggi agar menjadi lebih mudah dipisahkan [20]. *Hyperplane* dinyatakan dalam bentuk persamaan:

$$\min \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right) \quad (2)$$

Keterangan :

- $w$  : vektor bobot (weight vector)
- $\|w\|$  : norma Euclidean dari vektor bobot
- $C$  : parameter regularisasi untuk mengontrol trade-off margin dan kesalahan
- $\xi_i$  : slack variable untuk kesalahan klasifikasi data ke-
- $n$  : jumlah data latih

### 2.6.2 Random Forest

*Random Forest* merupakan metode *ensemble* yang membangun banyak *decision tree* dan menggabungkan hasilnya melalui mekanisme pemungutan suara (voting). Intinya, model ini bekerja dengan dua sumber keacakan utama yaitu pengambilan sampel data secara acak (bootstrap), dan pemilihan subset fitur secara acak pada setiap proses pemecahan node [14] [21] [22].

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_T(x)\} \quad (3)$$

Keterangan :

- $\hat{y}$  : hasil prediksi
- $ht(x)$  : hasil prediksi pohon keputusan
- $T$  : jumlah pohon keputusan
- mode : nilai yang paling sering muncul (voting mayoritas)

### 2.7 Evaluasi

Evaluasi performa dilakukan setelah model dilatih pada data latih dan diuji pada data uji. Tahap analisis performa mengacu pada penggunaan *confusion matrix* untuk membaca pola benar/salah prediksi pada masing-masing kelas. Berdasarkan *confusion matrix*, dihitung *matrix* evaluasi utama yaitu *accuracy*, *precision*, *recall*, dan *F1-score* (*F-measure*) [22]. Evaluasi berbasis *matrix* tersebut digunakan untuk menilai kualitas klasifikasi secara menyeluruh, tidak hanya benar-salah total tetapi juga keseimbangan kemampuan model mendeteksi kelas target (misalnya kabupaten/kota dengan tingkat kemiskinan tertentu) [20].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Keterangan :

- TP (True Positive) : data positif yang diprediksi benar
- TN (True Negative) : data negatif yang diprediksi benar
- FP (False Positive) : data negatif yang diprediksi sebagai positif
- FN (False Negative) : data positif yang diprediksi sebagai negative

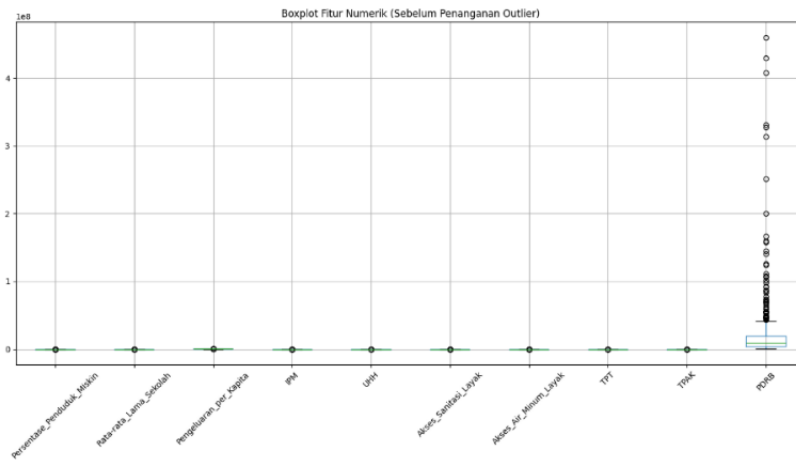
## 3. HASIL DAN PEMBAHASAN

### 3.1 Preprocessing

Pada tahap praproses, keberadaan *outlier* pada seluruh fitur numerik terlebih dahulu diperiksa menggunakan visualisasi *boxplot*. Pemeriksaan ini penting karena data sosial ekonomi antar kabupaten/kota umumnya memiliki heterogenitas tinggi, sehingga nilai ekstrem dapat muncul secara alami akibat perbedaan struktur ekonomi, kapasitas fiskal, dan kondisi pembangunan wilayah. Meskipun demikian, jika nilai ekstrem dibiarkan tanpa perlakuan, proses pembelajaran model

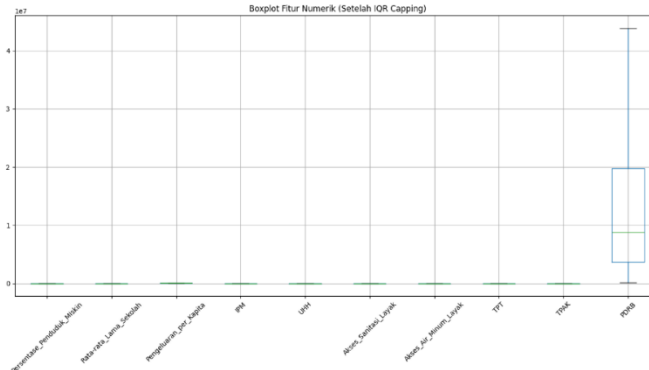


berisiko menjadi tidak stabil: distribusi fitur menjadi sangat timpang, parameter statistik untuk transformasi skala (misalnya saat standarisasi/normalisasi) dapat terdorong oleh sebagian kecil observasi, dan model dapat membentuk pola keputusan yang terlalu “mengikuti” titik ekstrem alih-alih mayoritas data. Kondisi sebelum penanganan *outlier* ditunjukkan pada Gambar 2.



**Gambar 2.** Data Sebelum Penanganan *Outlier*

Berdasarkan Gambar 2, *outlier* paling menonjol terlihat pada variabel dengan rentang nilai besar, khususnya PDRB, sementara beberapa indikator lain juga menunjukkan nilai yang berada jauh di luar whisker. Kondisi ini mengindikasikan adanya wilayah dengan karakteristik ekstrem yang berpotensi memengaruhi proses pembelajaran model. Untuk mengendalikan pengaruh outlier tanpa menghilangkan informasi observasi, penelitian ini menerapkan *IQR capping* (*winsorization* berbasis *IQR*) dengan menetapkan batas bawah dan atas berdasarkan Q1 dan Q3 serta rentang antarkuartil. Nilai yang berada di luar batas tersebut disesuaikan ke ambang terdekat, sehingga jumlah observasi tetap terjaga dan dominasi nilai ekstrem terhadap sebaran fitur dapat diminimalkan.



**Gambar 3.** Data Sesudah Penanganan *Outlier*

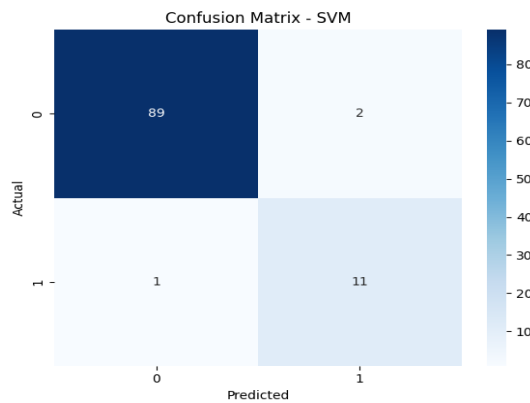
Setelah *IQR capping* diterapkan, distribusi fitur dievaluasi kembali menggunakan *boxplot* sebagaimana ditampilkan pada Gambar 3. Secara visual terlihat bahwa titik-titik ekstrem yang sebelumnya muncul jauh di luar whisker menjadi berkurang dan rentang sebaran terutama pada fitur berskala besar seperti PDRB lebih terkendali, sementara struktur inti distribusi (median dan rentang antarkuartil) tetap terjaga. Implikasinya, data menjadi lebih representatif untuk tahap pemodelan karena risiko distorsi akibat outlier menurun, sehingga proses pelatihan cenderung lebih stabil, baik pada model yang sensitif terhadap skala/jarak maupun pada model yang dipengaruhi oleh pemilihan ambang pemisah. Dengan demikian, penanganan *outlier* berbasis *IQR* pada penelitian ini berfungsi sebagai kompromi metodologis yang aman: tidak membuang informasi wilayah, tetapi tetap membatasi pengaruh nilai ekstrem agar tidak mendikte pola yang dipelajari model.

### 3.2 Pengujian Dataset

Pada tahap model klasifikasi, data hasil pra-proses digunakan untuk membangun model SVM dan *Random Forest*. Performa kedua model dievaluasi menggunakan *confusion matrix* serta *matrix accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai ketepatan prediksi sekaligus kemampuan mengenali kelas minoritas. Selain itu, diterapkan seleksi fitur RFE untuk menguji pengaruh reduksi fitur terhadap efisiensi dan kemampuan generalisasi model pada data uji.

#### 3.2.1 Support Vector Machine (SVM)

Untuk mengetahui distribusi kesalahan klasifikasi secara lebih rinci, dilakukan analisis menggunakan *confusion matrix*. Matriks ini menunjukkan jumlah prediksi benar dan salah yang dihasilkan oleh model *Support Vector Machine (SVM)* pada masing-masing kelas. Visualisasi hasil klasifikasi tersebut disajikan pada Gambar 4.



**Gambar 4.** *Confusion Matrix SVM*

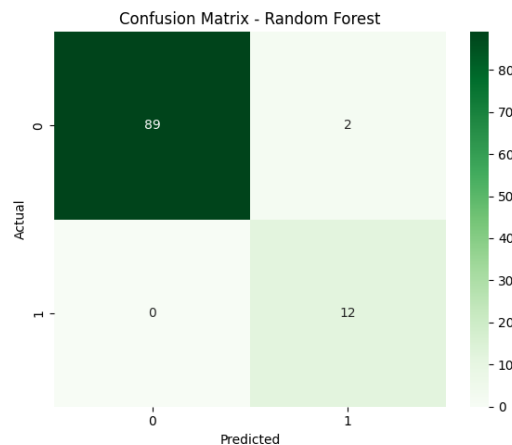
Berdasarkan hasil pengujian menggunakan *confusion matrix*, model SVM menunjukkan kinerja klasifikasi yang sangat baik pada 103 data uji dengan tingkat kesalahan yang minimal. Model mampu mengklasifikasikan kelas 0 (tidak miskin) secara benar sebanyak 89 data, dengan hanya 2 data yang salah diprediksi sebagai kelas 1. Pada kelas 1 (miskin), model berhasil memprediksi 11 data secara benar, dan hanya 1 data yang keliru diklasifikasikan sebagai kelas 0. Secara keseluruhan, terdapat 100 prediksi benar dari 103 data uji, yang mengindikasikan bahwa SVM memiliki kemampuan generalisasi yang baik, mampu mengenali kelas mayoritas secara akurat serta tetap efektif dalam mendeteksi kelas minoritas dengan tingkat kesalahan yang rendah. Untuk memberikan gambaran yang lebih terukur mengenai performa model, dilakukan evaluasi menggunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy*. Ringkasan hasil evaluasi kinerja model SVM disajikan pada Tabel 2.

**Tabel 2.** Laporan Klasifikasi Model SVM

	Precision	Recall	F1- Score	Support
0	0.99	0.98	0.98	91
1	0.85	0.92	0.88	12
Accuracy			0.97	103
Macro avg	0.92	0.95	0.93	103
Weighted avg	0.97	0.97	0.97	103

### 3.2.2 *Random Forest*

Untuk menganalisis distribusi kesalahan prediksi pada model *Random Forest*, digunakan *confusion matrix* sebagai alat evaluasi visual. Matriks ini menunjukkan jumlah prediksi benar dan salah untuk masing-masing kelas sehingga dapat memberikan gambaran detail mengenai performa klasifikasi model. Hasil visualisasi tersebut ditampilkan pada Gambar 5.



**Gambar 5.** *Confusion Matrix Random Forest*

Berdasarkan hasil pengujian *confusion matrix*, *Random Forest* pada 103 data uji, model menunjukkan kinerja yang sangat baik dengan tingkat kesalahan yang relatif kecil. Model berhasil mengklasifikasikan kelas 0 dengan benar sebanyak

89 data, namun terdapat 2 data kelas 0 yang keliru diprediksi sebagai kelas 1. Pada kelas 1, seluruh sampel berhasil dikenali dengan tepat, yaitu 12 data diprediksi benar sebagai kelas 1 dan tidak ada kasus kelas 1 yang salah diprediksi menjadi kelas 0. Secara keseluruhan, terdapat 101 prediksi benar dari 103 data, yang menegaskan bahwa *Random Forest* sangat efektif, terutama karena mampu mendeteksi kelas 1 secara sempurna (tanpa kehilangan kasus), meskipun masih terdapat sedikit *false positive* pada kelas 0. Untuk memperoleh evaluasi kuantitatif terhadap performa model, digunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy*. Ringkasan hasil evaluasi kinerja model *Random Forest* disajikan pada Tabel 3

**Tabel 3.** Laporan Klasifikasi Model *Random Forest*

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
0	1.00	0.98	0.99	91
1	0.86	1.00	0.92	12
<i>Accuracy</i>			0.98	103
<i>Macro avg</i>	0.93	0.99	0.96	103
<i>Weighted avg</i>	0.98	0.98	0.98	103

### 3.3 Recursive Feature Elimination (RFE)

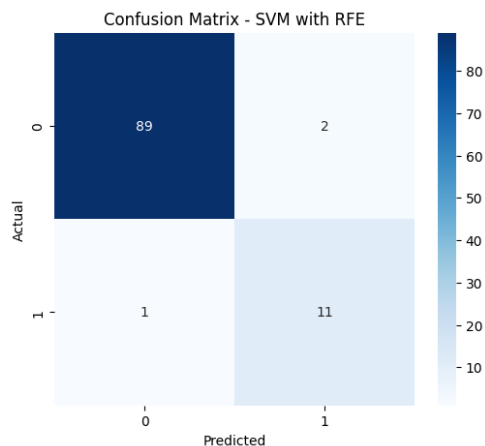
Seleksi fitur dalam penelitian ini dilakukan menggunakan metode RFE untuk mengidentifikasi subset fitur yang paling relevan dalam membedakan kelas target. Metode RFE bekerja secara iteratif dengan melatih model menggunakan seluruh fitur, kemudian mengeliminasi fitur dengan kontribusi terendah hingga diperoleh jumlah fitur yang ditetapkan. Pendekatan ini bertujuan mengurangi redundansi antarfitur, menyederhanakan kompleksitas model, serta meningkatkan interpretabilitas tanpa mengorbankan kinerja klasifikasi. Pada penelitian ini, RFE mempertahankan lima fitur terpilih dan diterapkan pada dua estimator, yaitu *Support Vector Machine* (linear) dan *Random Forest*, untuk mengamati karakteristik seleksi fitur pada model berbasis *hyperplane* dan pohon keputusan.

**Tabel 4.** Hasil RFE Eliminasi

Atribut	Coefficient
PO	1.8482
PPK	-0.3567
TPT	0.2960
TPAK	0.2238
PDRB	-0.2454

### 3.4 Support Vector Machine dengan RFE

Berdasarkan hasil pengujian menggunakan *confusion matrix*, model SVM dengan penerapan RFE menunjukkan kinerja klasifikasi yang sangat baik pada 103 data uji dengan tingkat kesalahan yang minimal. Model berhasil mengklasifikasikan kelas 0 (tidak miskin) secara benar sebanyak 89 data, dengan 2 data yang keliru diprediksi sebagai kelas 1. Pada kelas 1 (miskin), model mampu memprediksi 11 data secara benar, dan hanya terdapat 1 data yang salah diklasifikasikan sebagai kelas 0. Secara keseluruhan, terdapat 100 prediksi benar dari 103 data uji, yang menunjukkan bahwa penerapan RFE pada SVM tidak mengubah kinerja klasifikasi, namun tetap mempertahankan kemampuan generalisasi model serta efektivitas dalam mendeteksi kelas minoritas dengan tingkat kesalahan yang rendah.



**Gambar 6.** Confusion Matrix SVM dengan RFE

Untuk mengevaluasi pengaruh penerapan metode *Recursive Feature Elimination* (RFE) terhadap kinerja model *Support Vector Machine* (SVM), dilakukan pengujian menggunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy*. Ringkasan hasil evaluasi model SVM setelah penerapan RFE disajikan pada Tabel 5.



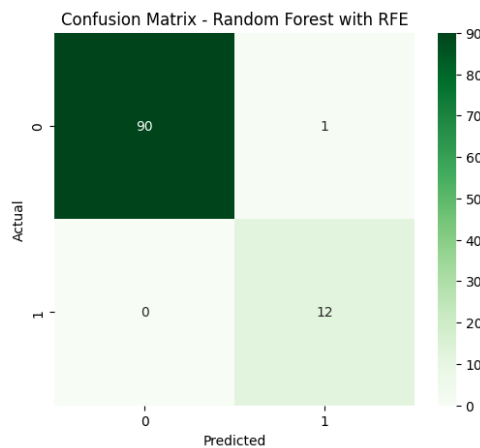
**Tabel 5.** Laporan Klasifikasi Model SVM dengan RFE

	<i>Precision</i>	<i>Recall</i>	<i>F1- Score</i>	<i>Support</i>
0	0.99	0.98	0.98	91
1	0.85	0.92	0.88	12
<i>Accuracy</i>			0.97	103
<i>Macro avg</i>	0.92	0.95	0.93	103
<i>Weighted avg</i>	0.97	0.97	0.97	103

Berdasarkan Tabel 5, model SVM dengan penerapan metode *Recursive Feature Elimination* memperoleh nilai *accuracy* sebesar 0,97 atau 97%, yang menunjukkan performa klasifikasi sangat baik secara keseluruhan. Pada kelas 0, diperoleh nilai *precision* sebesar 0,99 dan *recall* sebesar 0,98, yang menunjukkan bahwa model sangat akurat dalam mengidentifikasi kelas mayoritas. Sementara itu, pada kelas 1 diperoleh nilai *precision* sebesar 0,85 dan *recall* sebesar 0,92, yang menunjukkan bahwa model cukup efektif dalam mendeteksi kelas minoritas meskipun jumlah datanya lebih sedikit dibandingkan kelas 0. Nilai *macro average F1-score* sebesar 0,93 menunjukkan bahwa secara umum model memiliki performa yang baik pada kedua kelas tanpa mempertimbangkan ketidakseimbangan jumlah data. Sementara itu, nilai *weighted average F1-score* sebesar 0,97 menunjukkan bahwa performa keseluruhan model sangat stabil karena mempertimbangkan proporsi jumlah data pada masing-masing kelas. Hasil ini menunjukkan bahwa penerapan metode *Recursive Feature Elimination* tidak menurunkan performa model SVM, serta mampu mempertahankan tingkat akurasi yang tinggi dengan jumlah fitur yang lebih efisien. Hal ini mengindikasikan bahwa seleksi fitur membantu model dalam memfokuskan pembelajaran pada indikator yang lebih informatif.

### 3.5 *Random Forest* dengan RFE

Berdasarkan hasil pengujian menggunakan *confusion matrix*, model *Random Forest* dengan RFE menunjukkan kinerja klasifikasi yang sangat unggul pada 103 data uji dengan tingkat kesalahan yang sangat rendah. Model berhasil mengklasifikasikan kelas 0 (tidak miskin) secara benar sebanyak 90 data, dengan hanya 1 data yang keliru diprediksi sebagai kelas 1. Pada kelas 1 (miskin), seluruh 12 data berhasil diprediksi dengan benar sebagai kelas 1, tanpa adanya kesalahan klasifikasi (*false negative*). Secara keseluruhan, model menghasilkan 102 prediksi benar dari 103 data uji, yang menunjukkan bahwa kombinasi *Random Forest* dan RFE memiliki kemampuan generalisasi yang sangat baik serta sangat efektif dalam mendeteksi kelas minoritas secara sempurna dengan tingkat kesalahan yang minimal.



**Gambar 7.** *Confusion matrix Random Forest* dengan RFE

Untuk mengevaluasi pengaruh penerapan metode *Recursive Feature Elimination* (RFE) terhadap kinerja model *Random Forest*, dilakukan pengujian menggunakan *matrix precision*, *recall*, *F1-score*, dan *accuracy*. Hasil evaluasi model *Random Forest* setelah penerapan RFE ditunjukkan pada Tabel 6.

**Tabel 6.** Laporan Klasifikasi Model *Random Forest* dengan RFE

	<i>Precision</i>	<i>Recall</i>	<i>F1- Score</i>	<i>Support</i>
0	1.00	0.99	0.99	91
1	0.92	1.00	0.96	12
<i>Accuracy</i>			0.99	103
<i>Macro avg</i>	0.96	0.99	0.98	103
<i>Weighted avg</i>	0.99	0.99	0.99	103

Berdasarkan Tabel 6, model *Random Forest* dengan penerapan metode *Recursive Feature Elimination* memperoleh nilai *accuracy* sebesar 0,99 atau 99%, yang menunjukkan performa klasifikasi yang sangat tinggi. Pada kelas 0, diperoleh nilai *precision* sebesar 1,00 dan *recall* sebesar 0,99, yang menunjukkan bahwa hampir seluruh data kelas mayoritas berhasil diklasifikasikan dengan sangat akurat. Pada kelas 1, model memperoleh nilai *precision* sebesar 0,92 dan *recall* sebesar 1,00, yang menunjukkan bahwa seluruh data kelas minoritas berhasil terdeteksi dengan baik tanpa adanya kesalahan prediksi pada kelas tersebut. Nilai *F1-score* sebesar 0,96 pada kelas 1 menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Nilai *macro average F1-score* sebesar 0,98 menunjukkan bahwa model memiliki performa yang sangat baik secara seimbang pada kedua kelas, sementara nilai *weighted average F1-score* sebesar 0,99 menunjukkan stabilitas kinerja model secara keseluruhan dengan mempertimbangkan proporsi jumlah data masing-masing kelas. Hasil ini mengindikasikan bahwa penerapan metode *Recursive Feature Elimination* mampu meningkatkan performa model *Random Forest* secara signifikan, baik dalam hal akurasi maupun kemampuan mendeteksi kelas minoritas, sehingga model menjadi lebih efisien dan lebih andal dalam mengklasifikasikan tingkat kemiskinan kabupaten/kota.

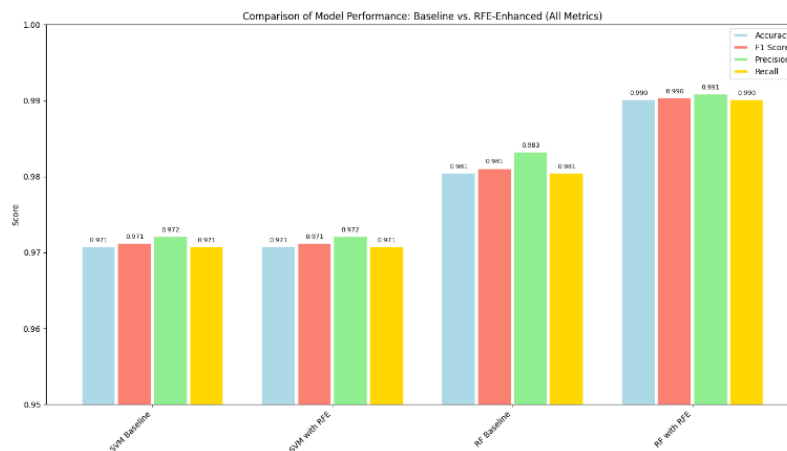
### 3.6 Evaluasi

Berdasarkan hasil RFE Eliminasi, evaluasi dilakukan untuk membandingkan kinerja model pada empat skenario, yaitu SVM, SVM + RFE, *Random Forest*, dan *Random Forest* + RFE menggunakan *matrix accuracy*, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan bahwa pada algoritma SVM, penerapan RFE tidak mengubah performa, di mana baik SVM maupun SVM + RFE memiliki nilai yang identik, yang dapat dilihat pada tabel 5. Hasil ini mengindikasikan bahwa seleksi fitur melalui RFE pada SVM lebih berfungsi untuk mereduksi dimensi dan meningkatkan efisiensi model, namun tidak memberikan peningkatan kinerja karena fitur-fitur yang dieliminasi tidak memberikan kontribusi tambahan yang signifikan terhadap pemisahan kelas.

Tabel 5. Perbandingan Hasil

Model	Accuracy	Precision	Recall	F1 Score
SVM	0,971	0,972	0,971	0,971
SVM + RFE	0,971	0,972	0,971	0,971
RF	0,981	0,983	0,981	0,981
RF + RFE	0,99	0,991	0,99	0,99

Berdasarkan Tabel 2, terlihat bahwa penerapan metode RFE memberikan dampak yang berbeda pada setiap algoritma. Pada algoritma SVM, nilai *accuracy*, *precision*, *recall*, dan *F1-score* pada model baseline dan model dengan RFE menunjukkan hasil yang identik. Hal ini mengindikasikan bahwa penerapan RFE pada SVM lebih berperan dalam mereduksi dimensi fitur dan meningkatkan efisiensi komputasi, namun belum memberikan peningkatan kinerja klasifikasi yang signifikan. Sebaliknya, pada algoritma *Random Forest*, penerapan RFE menunjukkan peningkatan kinerja yang jelas. Model *Random Forest* + RFE memperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang lebih tinggi dibandingkan model baseline. Hasil ini menunjukkan bahwa seleksi fitur menggunakan RFE mampu mengurangi redundansi fitur serta meningkatkan kemampuan model dalam mempelajari pola data kemiskinan secara lebih optimal. Perbandingan kinerja model secara visual ditunjukkan pada Gambar 8.



Gambar 8. Perbandingan Kinerja Algoritma Baseline

## 4. KESIMPULAN

Penelitian ini mengevaluasi pengaruh seleksi fitur menggunakan *Recursive Feature Elimination* terhadap kinerja klasifikasi tingkat kemiskinan kabupaten atau kota di Indonesia dengan membandingkan algoritma *Support Vector Machine* dan *Random Forest* dengan dataset berjumlah 514 observasi dengan dua kelas target, tidak miskin dan miskin,

setelah melalui tahap pembersihan data, penanganan nilai ekstrem menggunakan *IQR capping*, serta pembagian data 80 persen latih dan 20 persen uji. Eksperimen dilakukan pada empat skenario, yaitu SVM, SVM dengan RFE, *Random Forest*, dan *Random Forest* dengan RFE, di mana RFE mempertahankan lima indikator paling informatif yaitu persentase penduduk miskin, pengeluaran per kapita disesuaikan, tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, dan PDRB. Hasil pengujian menunjukkan bahwa pada SVM dengan RFE tidak mengubah performa sehingga *matrix* tetap sama dengan akurasi 0,971 dan *F1-score* 0,971, sedangkan pada *Random Forest* dengan RFE memberikan peningkatan yang jelas dari akurasi 0,981 dan *F1-score* 0,981 menjadi akurasi 0,99 dan *F1-score* 0,99 serta memperbaiki presisi pada kelas miskin tanpa mengurangi kemampuan deteksinya, sehingga model terbaik pada penelitian ini adalah *Random Forest* dengan RFE karena menghasilkan performa paling tinggi sekaligus lebih efisien melalui reduksi fitur. Dengan demikian, dapat disimpulkan bahwa RFE bermanfaat sebagai strategi penyederhanaan pada SVM dan sebagai penguat generalisasi pada *Random Forest*, serta layak dipertimbangkan untuk pemodelan klasifikasi kemiskinan yang menuntut akurasi tinggi dan indikator yang benar-benar relevan, dengan penguatan riset lanjutan melalui validasi silang, variasi jumlah fitur terpilih, dan analisis interpretabilitas agar rekomendasi kebijakan yang dihasilkan semakin robust.

## REFERENCES

- [1] World Bank, "Poverty Overview: Data dan tren kemiskinan global, termasuk definisi garis kemiskinan ekstrem." Accessed: Jan. 19, 2025. [Online]. Available: [https://www.worldbank.org/en/topic/poverty/overview?utm\\_source=chatgpt.com](https://www.worldbank.org/en/topic/poverty/overview?utm_source=chatgpt.com)
- [2] A. Heryati and T. Setiawan Saputra, "Optimizing Socioeconomic Features for Poverty Prediction in South Sumatera," *TIERS Inf. Technol. J.*, vol. 6, no. 1, pp. 16–32, 2025.
- [3] M. F. M. Khalik and F. Arifin, "Klasifikasi Indeks Kedalaman Kemiskinan Provinsi Sulawesi Selatan Berbasis Decision Tree, K-Nearest Neighbor, Naive Bayes, Neural Network, dan Random Forest," *J. Edukasi dan Penelit. Inform.*, vol. 9, no. 2, p. 282, 2023, doi: 10.26418/jp.v9i2.67492.
- [4] M. Y. Sofian, R. Dwijaya, and S. Rachmalija, "Kebijakan Anti Kemiskinan Program Pemerintah Dalam Penanggulangan Kemiskinan Di Indonesia," Vol. 2, No. 10, Pp. 3209–3218, 2022.
- [5] N. Putu, N. Hendayanti, and M. Nurhidayati, "Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi di Indonesia," vol. 12, no. 2, 2020.
- [6] A. Of, D. Mining, F. To, and P. Data, "Penerapan *Data mining* Dan Forecasting Terhadap Data Kemiskinan Di Indonesia Application of *Data mining* and Forecasting To Poverty Data in," no. September 2019, pp. 375–383, 2024.
- [7] M. S. Hartawan, M. Erkamim, and S. R. Yahya, "Application of Supervised Learning Algorithm for Classification of Family Hope Program Penerapan Algoritma Supervised Learning untuk Klasifikasi Program Keluarga Harapan," vol. 3, no. October, pp. 83–91, 2023.
- [8] E. F. Laili *et al.*, "Komparasi Algoritma Decision Tree Dan Support Vector Machine ( Svm ) Dalam," Vol. 8, No. 1, Pp. 67–76, 2025.
- [9] D. Daniel, "Poverty Prediction using Random Forest based Machine Learning Technique," vol. 10, no. 04, pp. 153–157, 2021.
- [10] L. Nuzula, A. Prahutama, A. R. Hakim, F. Sains, and U. Diponegoro, "Klasifikasi Status Kemiskinan Rumah Tangga Dengan Metode Support Vector Machines ( Svm ) Dan Classification And Regression Trees ( Cart ) Menggunakan Gui R ( Studi Kasus Di Kabupaten Wonosobo Tahun 2018 )," Vol. 9, Pp. 525–534, 2020.
- [11] V. No, Z. A. Mukharyahya, Y. P. Astuti, and O. N. Cahyani, "Edumatic : Jurnal Pendidikan Informatika Perbandingan Naive Bayes dan Support Vector Machine dalam Klasifikasi Tingkat Kemiskinan di Indonesia," vol. 9, no. 1, pp. 119–128, 2025.
- [12] C. Human and D. Index, "Perbandingan Metode Ensemble Learning : Random Forest , Support Vector Machine , AdaBoost pada Klasifikasi Indeks Pembangunan Manusia ( IPM ) Comparison of Ensemble Learning Method : Random Forest , Support Vector," vol. 12, pp. 206–218, 2023.
- [13] A. M. Priyatno, T. Widiyaningtyas, I. Engineering, and U. N. Malang, "A Systematic Literature Review : Recursive Feature," Vol. 9, No. 2, Pp. 196–207, 2024, Doi: 10.33480/Jitk.V9i2.5015.Introduction.
- [14] N. Aqmar, H. Wijayanto, and F. M. Afendi, "Performance Analysis of Machine Learning Models using RFE Feature Selection and Bayesian Optimization in Imbalanced Data Classification with Shap-Based Explanations," vol. 12, no. 3, pp. 539–554, 2025, doi: 10.15294/sji.v12i3.31459.
- [15] A. Amir, F. Fachruddin, F. Idris, M. Safriani, R. Saefuddin, and I. Sakti, "Perbandingan Model Decision Tree dan Random Forest," vol. 12, no. 5, pp. 695–704, 2025, doi: 10.30865/jurikom.v12i5.8672.
- [16] A. Informatics and A. Info, "Pendekatan Data - Driven untuk Pengembangan Model Prediksi Tingkat Kemiskinan di Provinsi Indonesia," vol. 8, no. 1, pp. 84–92, 2025.
- [17] A. R. Alsaber and J. Pan, "Handling Complex Missing Data Using Random Forest Approach for an Air Quality Monitoring Dataset : A Case Study of Kuwait Environmental Data ( 2012 to 2018 )," 2021.
- [18] B. Satria, T. Azhima, Y. Siswa, and W. J. Pranoto, "Optimasi Random Forest dengan Genetic Algorithm dan *Recursive Feature Elimination* pada High Dimensional Data Stunting Samarinda," vol. 8, pp. 1778–1789, 2024, doi: 10.30865/mib.v8i3.7883.
- [19] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sistemasi*, vol. 10, no. 1, p. 163, 2021, doi: 10.32520/stmsi.v10i1.1129.
- [20] I. Kurniawan *et al.*, "Implementasi Algoritma Random Forest Untuk Menentukan Implementation Of Random Forest Algorithm For Determining Recipients Of Raskin," Vol. 10, No. 2, Pp. 421–428, 2023, Doi: 10.25126/Jtiik.202396225.
- [21] S. Adi, S. Mola, Y. C. Luttu, and D. N. Rumlaklak, "Perbandingan Metode Machine Learning dalam Analisis Sentimen Komentar Pengguna Aplikasi InDriver pada Dataset Tidak Seimbang," vol. 03, 2024, doi: 10.21456/vol14iss3pp247-255.
- [22] S. D. Amalia, M. A. Barata, and P. E. Yuwita, "Optimization of Random Forest Algorithm with Backward Elimination Method in Classification of Academic Stress Levels," vol. 9, no. 3, 2025.