



## Optimasi Model Machine Learning untuk Reduksi Kesalahan Klasifikasi Penerima BLT Berbasis Data Sosial-Ekonomi

Suherman<sup>1</sup>, Ismail<sup>2</sup>

Program Studi Teknik Informatika, Universitas Lamappapoleonro<sup>1</sup>

Program Studi Sistem Informasi, Universitas Lamappapoleonro<sup>2</sup>

Jl. Salotungo No. 62 Watansoppeng, Soppeng, Sulawesi Selatan-Indonesia<sup>1,2</sup>

suherman@unipol.ac.id \*<sup>1</sup>, ismail@unipol.ac.id<sup>2</sup>

### Kata Kunci :

BLT;  
Klasifikasi;  
Machine Learning;  
Random Forest;  
Sosial-Ekonomi.

### ABSTRAK

Penelitian ini bertujuan mengoptimalkan model *machine learning* untuk mereduksi kesalahan klasifikasi penerima Bantuan Langsung Tunai (BLT) berbasis data sosial-ekonomi di Kabupaten Soppeng. Penelitian menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional. Dataset yang digunakan terdiri atas 300 data calon penerima BLT dengan 22 variabel, mencakup indikator sosial-ekonomi seperti pendapatan, pekerjaan, jumlah tanggungan, kondisi rumah, kepemilikan aset, status kerentanan, dan riwayat penerimaan bantuan sosial. Algoritma yang diuji meliputi Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbor, Naive Bayes, dan XGBoost. Proses optimasi dilakukan melalui prapemrosesan data, *one-hot encoding*, normalisasi, seleksi fitur, dan *hyperparameter tuning*. Evaluasi model menggunakan *accuracy*, *precision*, *recall*, *F1-score*, ROC-AUC, dan *confusion matrix*. Hasil penelitian menunjukkan bahwa Random Forest menjadi model terbaik dengan *accuracy* 0,9833, *precision* 1,0000, *recall* 0,9643, *F1-score* 0,9818, dan ROC-AUC 0,9900. Analisis *confusion matrix* menunjukkan hanya terdapat satu kesalahan klasifikasi berupa *false negative* dan tidak terdapat *false positive*. Temuan ini menunjukkan bahwa model *machine learning* yang dioptimalkan dapat mendukung klasifikasi penerima BLT secara lebih objektif, akurat, dan berbasis data

### Keywords

BLT;  
Classification;  
Machine Learning;  
Random Forest;  
Socio-Economic Data.

### ABSTRACT

*This study aims to optimize machine learning models to reduce classification errors in determining recipients of Direct Cash Assistance (BLT) based on socio-economic data in Soppeng Regency. This research employed a quantitative approach with a computational experimental design. The dataset consisted of 300 prospective BLT recipient records with 22 variables, including socio-economic indicators such as income, occupation, number of dependents, housing condition, asset ownership, vulnerability status, and social assistance history. The evaluated algorithms included Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbor, Naive Bayes, and XGBoost. The optimization process involved data preprocessing, one-hot encoding, normalization, feature selection, and hyperparameter tuning. Model performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix. The results show that Random Forest achieved the best performance, with an accuracy of 0.9833, precision of 1.0000, recall of 0.9643, F1-score of 0.9818, and ROC-AUC of 0.9900. The confusion matrix analysis indicated only one classification error in the form of a false negative and no false positives. These findings suggest that the optimized machine learning model*



*can support a more objective, accurate, and data-driven classification process for BLT recipient eligibility*

---Jurnal JISTI @2026---

## PENDAHULUAN

Program Bantuan Langsung Tunai (BLT) merupakan salah satu instrumen perlindungan sosial yang memiliki peran penting dalam menjaga ketahanan ekonomi masyarakat miskin dan rentan. Program ini tidak hanya berfungsi sebagai bantuan konsumtif jangka pendek, tetapi juga sebagai bentuk intervensi pemerintah untuk mengurangi tekanan ekonomi rumah tangga yang menghadapi keterbatasan pendapatan, pekerjaan tidak tetap, jumlah tanggungan tinggi, serta kondisi sosial-ekonomi yang rentan berubah. Secara nasional, isu kemiskinan masih menjadi perhatian serius. Badan Pusat Statistik mencatat bahwa persentase penduduk miskin Indonesia pada September 2025 sebesar 8,25% atau sekitar 23,36 juta orang. Di Provinsi Sulawesi Selatan, persentase penduduk miskin pada September 2025 tercatat 7,43% atau sekitar 685,14 ribu orang. Dalam konteks Kabupaten Soppeng, BPS Sulawesi Selatan mencatat persentase penduduk miskin tahun 2025 sebesar 6,65%, sedangkan dokumen Perubahan RKPD Kabupaten Soppeng Tahun 2025 menunjukkan capaian angka kemiskinan daerah tahun 2024 sebesar 6,90%. Data tersebut memperlihatkan bahwa meskipun angka kemiskinan relatif menurun, kebutuhan terhadap mekanisme penetapan penerima bantuan yang akurat tetap menjadi persoalan penting di tingkat daerah (Muh. Ikhsan Amar, 2025).

Urgensi penelitian ini berangkat dari kebutuhan untuk meningkatkan kualitas pengambilan keputusan dalam penetapan penerima BLT, khususnya di Kabupaten Soppeng, Sulawesi Selatan. Bantuan sosial yang tidak tepat sasaran dapat menimbulkan persoalan sosial dan administratif, baik dalam bentuk masyarakat layak yang tidak menerima bantuan maupun masyarakat kurang layak yang justru masuk sebagai penerima. Kondisi tersebut dapat terjadi karena data sosial-ekonomi masyarakat bersifat dinamis, sementara proses verifikasi sering kali berhadapan dengan keterbatasan waktu, sumber daya, dan konsistensi data. BPS Kabupaten Soppeng melalui publikasi *Kabupaten Soppeng Dalam Angka 2025* menegaskan bahwa data daerah mencakup perkembangan kondisi sosial-demografi dan perekonomian yang bersumber dari BPS serta dinas atau instansi terkait (Suherman, Misveria Villa Waru, 2022). Hal ini menunjukkan bahwa ketersediaan data sosial-ekonomi daerah dapat dimanfaatkan secara lebih optimal untuk mendukung proses analisis kelayakan penerima bantuan secara berbasis data.

Permasalahan utama dalam penelitian ini adalah masih adanya potensi kesalahan klasifikasi dalam menentukan kelayakan penerima BLT berdasarkan data sosial-ekonomi masyarakat. Kesalahan klasifikasi dapat terjadi ketika sistem atau model tidak mampu membedakan secara tepat antara rumah tangga yang layak dan tidak layak menerima bantuan. Dalam konteks machine learning, kesalahan tersebut dapat dipengaruhi oleh kualitas data, pemilihan fitur, ketidakseimbangan kelas, pemilihan algoritma, serta parameter model yang belum optimal (Aiken et al., 2022). Indikator sosial-ekonomi seperti jenis pekerjaan, pendapatan, jumlah tanggungan, pendidikan, kondisi tempat tinggal, kepemilikan aset, dan status kerentanan keluarga sering kali memiliki hubungan yang kompleks dan tidak selalu linear. Oleh karena itu, pendekatan klasifikasi konvensional atau penilaian manual berpotensi kurang memadai apabila tidak didukung oleh model komputasional yang mampu mempelajari pola data secara sistematis (Bifarin, 2023).

Penelitian ini bertujuan untuk mengoptimalkan model machine learning dalam mereduksi kesalahan klasifikasi penerima BLT berbasis data sosial-ekonomi di Kabupaten Soppeng. Secara khusus, penelitian ini diarahkan untuk membangun model klasifikasi, membandingkan kinerja beberapa algoritma machine learning, serta melakukan optimasi model agar prediksi kelayakan penerima bantuan menjadi lebih akurat (Breiman, 2001). Proses optimasi dapat dilakukan melalui prapemrosesan data, seleksi fitur, penanganan data tidak seimbang, pembagian data latih dan data uji, validasi silang, serta



tuning parameter. Kinerja model dievaluasi menggunakan metrik yang relevan, seperti accuracy, precision, recall, F1-score, ROC-AUC, dan confusion matrix. Dengan demikian, penelitian ini tidak hanya berorientasi pada pencarian nilai akurasi tertinggi, tetapi juga pada kemampuan model dalam menekan kesalahan prediksi terhadap kelompok masyarakat yang layak dan tidak layak menerima BLT (Burke et al., 2021).

Sejumlah penelitian sebelumnya telah menerapkan sistem pendukung keputusan dan algoritma machine learning untuk membantu seleksi penerima bantuan sosial. Namun, sebagian besar kajian masih berfokus pada perbandingan algoritma secara umum dan cenderung menempatkan akurasi sebagai indikator utama keberhasilan model. Pendekatan tersebut belum sepenuhnya memadai karena dalam kasus bantuan sosial, kesalahan klasifikasi memiliki konsekuensi yang lebih luas daripada sekadar nilai evaluasi teknis (Chicco & Jurman, 2020). Kesalahan dalam mengklasifikasikan calon penerima dapat berdampak pada keadilan distribusi bantuan, kualitas pelayanan publik, dan kepercayaan masyarakat terhadap proses penetapan penerima. Selain itu, masih terbatas penelitian yang secara khusus mengarahkan proses optimasi model machine learning untuk mengurangi kesalahan klasifikasi pada data sosial-ekonomi daerah, khususnya dalam konteks penerima BLT di Kabupaten Soppeng (Jean et al., 2016).

Kebaruan penelitian ini terletak pada fokus optimasi model machine learning untuk mereduksi kesalahan klasifikasi penerima BLT berbasis data sosial-ekonomi daerah. Berbeda dari penelitian yang hanya membandingkan algoritma secara deskriptif, penelitian ini menempatkan reduksi kesalahan klasifikasi sebagai orientasi utama dalam pengembangan model. Optimasi dilakukan tidak hanya pada pemilihan algoritma, tetapi juga pada tahapan pengolahan data, pemilihan fitur, penanganan ketidakseimbangan data, serta evaluasi kinerja model secara lebih komprehensif (McBride & Nichols, 2018). Dengan pendekatan tersebut, model yang dihasilkan diharapkan tidak hanya memiliki performa prediktif yang baik, tetapi juga lebih relevan untuk mendukung pengambilan keputusan dalam penetapan penerima bantuan sosial berbasis data lokal (Poulin et al., 2022).

Penelitian ini memberikan kontribusi terhadap pengembangan kajian Teknik Informatika, khususnya pada penerapan machine learning untuk klasifikasi data sosial-ekonomi dan optimasi model prediktif pada domain kebijakan publik. Penelitian ini juga dapat menjadi dasar bagi pengembangan model yang lebih lanjut, seperti explainable machine learning untuk menjelaskan faktor dominan yang memengaruhi kelayakan penerima, serta causal machine learning untuk menganalisis hubungan sebab-akibat dalam penetapan bantuan sosial (Probst et al., 2019). Secara praktis, hasil penelitian ini diharapkan dapat memberikan masukan bagi pemerintah daerah dalam membangun mekanisme pendukung keputusan yang lebih objektif, transparan, dan berbasis data. Dengan model machine learning yang dioptimalkan, proses klasifikasi kelayakan penerima BLT di Kabupaten Soppeng dapat dilakukan secara lebih akurat, efisien, dan tepat sasaran.

## METODE PENELITIAN

### 1. Jenis dan Lokasi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional. Fokus penelitian adalah membangun, mengoptimalkan, dan mengevaluasi model *machine learning* untuk klasifikasi kelayakan penerima BLT berbasis data sosial-ekonomi. Lokasi penelitian ini adalah Kabupaten Soppeng, Provinsi Sulawesi Selatan. Pemilihan Kabupaten Soppeng sebagai lokasi penelitian didasarkan pada relevansi wilayah tersebut sebagai daerah yang memiliki kebutuhan terhadap penguatan mekanisme penetapan penerima bantuan sosial berbasis data. Objek penelitian adalah data sosial-ekonomi calon penerima BLT, sedangkan unit analisis dalam penelitian ini berupa individu atau



rumah tangga yang dinilai berdasarkan sejumlah indikator kelayakan. Indikator tersebut dapat mencakup pekerjaan, pendapatan, jumlah tanggungan, tingkat pendidikan, kondisi tempat tinggal, kepemilikan aset, status kepemilikan rumah, serta riwayat penerimaan bantuan sosial. Lokasi wilayah penelitian dapat dilihat pada gambar berikut:



Gambar 1. Lokasi Penelitian

## 2. Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini adalah data sosial-ekonomi calon penerima Bantuan Langsung Tunai (BLT) di Kabupaten Soppeng, Sulawesi Selatan, yang dapat bersumber dari data administrasi pemerintah daerah, data desa/kelurahan, data calon penerima bantuan, dan hasil verifikasi lapangan. Variabel input yang digunakan mencakup indikator sosial-ekonomi rumah tangga, seperti usia kepala keluarga, jenis pekerjaan, pendapatan, jumlah tanggungan, tingkat pendidikan, kondisi tempat tinggal, status kepemilikan rumah, kepemilikan aset, sumber penghasilan, status kerentanan keluarga, dan riwayat penerimaan bantuan sosial. Adapun variabel target dalam penelitian ini adalah status kelayakan penerima BLT yang diklasifikasikan ke dalam dua kategori, yaitu layak dan tidak layak. Seluruh variabel tersebut digunakan sebagai dasar dalam membangun model *machine learning* untuk mengenali pola kelayakan penerima bantuan dan mereduksi kesalahan klasifikasi. Sumber dan jenis data dapat dilihat pada tabel berikut:

Tabel 1. Sumber dan Variabel Data Penelitian

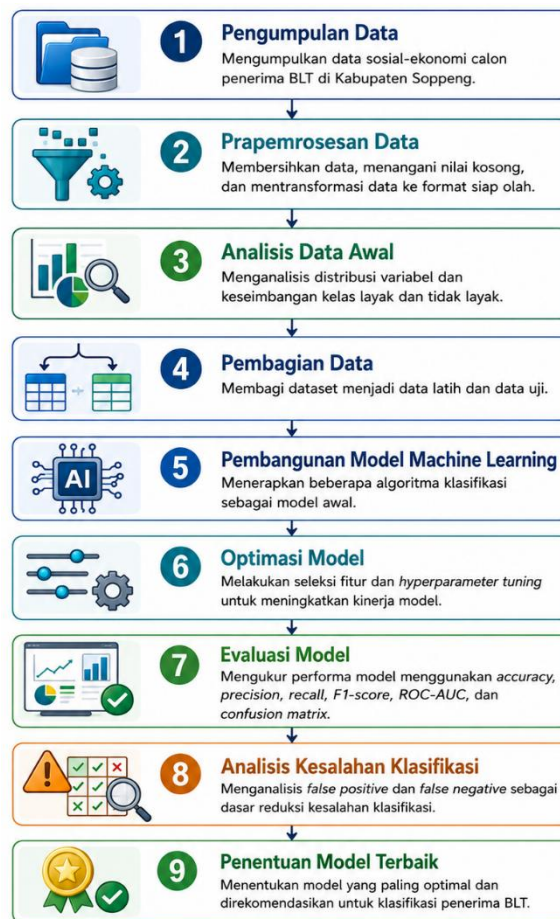
No	Variabel	Jenis Data	Sumber Data	Keterangan
1	Usia kepala keluarga	Numerik	Data administrasi penduduk / verifikasi lapangan	Menunjukkan umur kepala rumah tangga
2	Jenis pekerjaan	Kategorikal	Data sosial-ekonomi calon penerima BLT	Menggambarkan status pekerjaan, seperti tetap, tidak tetap, buruh, petani, atau tidak bekerja
3	Pendapatan rumah tangga	Numerik	Data sosial-ekonomi / verifikasi lapangan	Menggambarkan kemampuan ekonomi rumah tangga per bulan



4	Jumlah tanggungan	Numerik	Data keluarga / kartu keluarga	Menunjukkan jumlah anggota keluarga yang menjadi tanggungan
5	Tingkat pendidikan	Kategorikal	Data administrasi penduduk	Menggambarkan jenjang pendidikan terakhir kepala keluarga atau anggota rumah tangga
6	Kondisi tempat tinggal	Kategorikal	Data verifikasi lapangan	Menilai kelayakan rumah berdasarkan kondisi fisik bangunan
7	Status kepemilikan rumah	Kategorikal	Data sosial-ekonomi / verifikasi lapangan	Menunjukkan status rumah, seperti milik sendiri, sewa, kontrak, atau menumpang
8	Kepemilikan aset	Kategorikal/Numerik	Data sosial-ekonomi / verifikasi lapangan	Menggambarkan kepemilikan kendaraan, lahan, ternak, atau aset ekonomi lainnya
9	Sumber penghasilan	Kategorikal	Data sosial-ekonomi calon penerima BLT	Menunjukkan sumber utama pendapatan rumah tangga
10	Status kesehatan/kerentanan	Kategorikal	Data sosial / verifikasi lapangan	Menunjukkan adanya anggota keluarga lansia, disabilitas, sakit kronis, atau kelompok rentan lainnya
11	Riwayat penerimaan bantuan sosial	Kategorikal	Data bantuan sosial pemerintah daerah/desa	Menunjukkan apakah rumah tangga pernah menerima bantuan sosial sebelumnya
12	Status kelayakan penerima BLT	Kategorikal	Data hasil penetapan/verifikasi penerima BLT	Variabel target dengan kelas layak dan tidak layak

### 3. Tahapan penelitian

Tahapan penelitian ini disusun secara sistematis agar proses pembangunan model *machine learning* dapat dilakukan secara terarah, terukur, dan dapat dipertanggungjawabkan. Penelitian diawali dengan pengumpulan data sosial-ekonomi calon penerima BLT di Kabupaten Soppeng, kemudian dilanjutkan dengan prapemrosesan data untuk memastikan data bersih, konsisten, dan siap digunakan dalam proses pemodelan. Setelah itu, dilakukan analisis awal terhadap karakteristik data, termasuk distribusi variabel dan keseimbangan kelas antara kategori layak dan tidak layak. Data yang telah siap kemudian dibagi menjadi data latih dan data uji untuk membangun serta mengevaluasi model. Beberapa algoritma *machine learning* diterapkan sebagai model klasifikasi awal, kemudian dilakukan proses optimasi melalui seleksi fitur dan pengaturan parameter agar kinerja model meningkat. Tahap berikutnya adalah evaluasi model menggunakan metrik seperti *accuracy*, *precision*, *recall*, *F1-score*, ROC-AUC, dan *confusion matrix* (Ismail, 2025). Hasil evaluasi digunakan untuk menganalisis kesalahan klasifikasi, khususnya *false positive* dan *false negative*, sehingga dapat ditentukan model terbaik yang paling mampu mereduksi kesalahan klasifikasi penerima BLT berbasis data sosial-ekonomi. Tahapan penelitian secara detail dapat dilihat pada gambar berikut:



Gambar 2. Tahapan Penelitian

Gambar 2 tersebut menunjukkan alur penelitian dalam pengembangan model *machine learning* untuk mereduksi kesalahan klasifikasi penerima Bantuan Langsung Tunai (BLT) berbasis data sosial-ekonomi. Tahapan penelitian disusun secara berurutan mulai dari pengumpulan data hingga penentuan model terbaik. Alur ini menggambarkan bahwa proses penelitian tidak hanya berfokus pada pembangunan model prediksi, tetapi juga menekankan pentingnya kualitas data, proses optimasi, evaluasi kinerja, serta analisis kesalahan klasifikasi agar hasil model dapat digunakan secara lebih objektif dan bertanggung jawab.

Tahap pertama adalah pengumpulan data, yaitu proses menghimpun data sosial-ekonomi calon penerima BLT di Kabupaten Soppeng. Data ini menjadi dasar utama dalam penelitian karena kualitas dan kelengkapan data akan sangat memengaruhi hasil pemodelan. Tahap kedua adalah prapemrosesan data, yang dilakukan untuk membersihkan data, menangani nilai kosong, menghilangkan data yang tidak konsisten, serta mengubah data ke dalam format yang dapat diproses oleh algoritma *machine learning*. Tahap ini penting agar model tidak belajar dari data yang keliru atau tidak valid. Tahap ketiga adalah analisis data awal, yaitu proses memahami karakteristik data, distribusi variabel, serta keseimbangan antara kelas layak dan tidak layak menerima BLT. Analisis ini diperlukan untuk melihat apakah data memiliki pola tertentu atau mengalami ketidakseimbangan kelas. Tahap keempat adalah pembagian data, yaitu membagi dataset menjadi data latih dan data uji. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengukur kemampuan model dalam memprediksi data baru.

Tahap kelima adalah pembangunan model machine learning, yaitu penerapan beberapa algoritma klasifikasi sebagai model awal. Pada tahap ini, algoritma seperti Logistic Regression, Decision Tree,



Random Forest, Support Vector Machine, K-Nearest Neighbor, Naive Bayes, atau XGBoost dapat digunakan untuk membandingkan performa klasifikasi. Tahap keenam adalah optimasi model, yaitu proses peningkatan kinerja model melalui seleksi fitur dan *hyperparameter tuning*. Tahapan ini bertujuan agar model yang dihasilkan lebih akurat, stabil, dan tidak hanya bekerja baik pada data latih, tetapi juga mampu melakukan generalisasi pada data uji. Tahap ketujuh adalah evaluasi model, yaitu pengukuran performa model menggunakan beberapa metrik evaluasi seperti *accuracy*, *precision*, *recall*, *F1-score*, ROC-AUC, dan *confusion matrix* (Yeh et al., 2020). Evaluasi ini penting karena dalam kasus penerima BLT, akurasi saja tidak cukup untuk menilai kualitas model. Tahap kedelapan adalah analisis kesalahan klasifikasi, yaitu mengkaji kesalahan prediksi berupa *false positive* dan *false negative*. *False positive* terjadi ketika calon penerima yang sebenarnya tidak layak diprediksi layak, sedangkan *false negative* terjadi ketika calon penerima yang sebenarnya layak justru diprediksi tidak layak. Kedua jenis kesalahan ini perlu ditekan karena berkaitan langsung dengan ketepatan sasaran bantuan sosial. Tahap terakhir adalah penentuan model terbaik, yaitu memilih model yang memiliki performa paling optimal berdasarkan hasil evaluasi dan kemampuan dalam mereduksi kesalahan klasifikasi.

## HASIL DAN PEMBAHASAN

### 1. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini merupakan data sosial-ekonomi calon penerima Bantuan Langsung Tunai (BLT) di Kabupaten Soppeng, Sulawesi Selatan. Dataset terdiri atas 300 data observasi dengan 22 variabel, yang mencakup variabel identitas, variabel input, variabel turunan, dan variabel target. Variabel input memuat indikator sosial-ekonomi rumah tangga, seperti kecamatan, usia kepala keluarga, jenis kelamin, pendidikan, pekerjaan, pendapatan bulanan, jumlah tanggungan, status kepemilikan rumah, kondisi rumah, luas lantai, sumber air, daya listrik, kepemilikan kendaraan, kepemilikan aset produktif, anggota lansia, anggota disabilitas, anggota sakit kronis, riwayat penerimaan bantuan sosial, dan status penerimaan bantuan sosial lain. Selain itu, terdapat variabel Skor\_Kerentanan sebagai variabel turunan yang menggambarkan tingkat kerentanan sosial-ekonomi calon penerima. Ringkasan Dataset Penelitian dapat dilihat pada tabel berikut:

Tabel 2. Ringkasan Dataset Penelitian

Komponen Dataset	Keterangan
Nama dataset	Dataset calon penerima BLT berbasis data sosial-ekonomi
Lokasi studi kasus	Kabupaten Soppeng, Sulawesi Selatan
Jumlah observasi	300 data calon penerima BLT
Jumlah variabel keseluruhan	22 variabel
Jumlah variabel input	20 variabel
Variabel identifiier	ID_Rumah_Tangga
Variabel target	Status_Kelayakan_BLT
Kelas target	Layak dan Tidak Layak
Jumlah data kelas Layak	139 data
Jumlah data kelas Tidak Layak	161 data
Persentase kelas Layak	46,33%
Persentase kelas Tidak Layak	53,67%
Jenis data	Numerik, kategorikal, biner, dan teks
Variabel numerik	Usia_KK, Pendapatan_Bulanan, Jumlah_Tanggungan, Luas_Lantai_M2, Skor_Kerentanan
Variabel biner	Anggota_Lansia, Anggota_Disabilitas, Anggota_Sakit_Kronis
Variabel kategorikal	Kecamatan, Jenis_Kelamin_KK, Pendidikan_KK, Pekerjaan_KK, Status_Kepemilikan_Rumah, Kondisi_Rumah, Sumber_Air, Daya_Listrik,



Bentuk masalah penelitian	Kepemilikan_Kendaraan, Riwayat_Penerimaan_Bansos, Penerima_Bansos_Lain	Kepemilikan_Aset_Produktif,
Tujuan penggunaan data	Klasifikasi biner	
Catatan data	Membangun dan mengoptimalkan model <i>machine learning</i> untuk klasifikasi kelayakan penerima BLT	
	Dataset digunakan sebagai data penelitian/simulasi akademik dan dapat disesuaikan dengan data resmi atau hasil verifikasi lapangan	

Sumber : Hasil Olahan Dataset

## 2. Distribusi Kelas Target

Distribusi kelas target digunakan untuk melihat komposisi data berdasarkan status kelayakan calon penerima Bantuan Langsung Tunai (BLT). Dalam penelitian ini, variabel target adalah *Status\_Kelayakan\_BLT*, yang terdiri atas dua kelas, yaitu Layak dan Tidak Layak. Berdasarkan dataset yang digunakan, dari total 300 data calon penerima BLT, terdapat 139 data yang termasuk dalam kategori Layak dan 161 data yang termasuk dalam kategori Tidak Layak. Dengan persentase masing-masing sebesar 46,33% dan 53,67%, distribusi kelas target dapat dikatakan relatif seimbang karena selisih jumlah data antarkelas tidak terlalu besar. Detail distribusi kelas target dapat dilihat pada tabel berikut:

Tabel 3. Distribusi Kelas Target

Status Kelayakan BLT	Jumlah Data	Persentase
Layak	139	46,33%
Tidak Layak	161	53,67%
Total	300	100%

Sumber : Hasil Olahan Dataset

Berdasarkan Tabel 3, distribusi kelas target menunjukkan bahwa dataset memiliki komposisi yang cukup seimbang antara kelas Layak dan Tidak Layak. Kondisi ini mendukung proses pelatihan model *machine learning* karena model memiliki kesempatan yang relatif sama untuk mempelajari pola dari kedua kelas. Meskipun demikian, analisis kesalahan klasifikasi tetap diperlukan untuk memastikan model tidak hanya memiliki performa tinggi secara umum, tetapi juga mampu menekan kesalahan prediksi pada kedua kelas, khususnya *false positive* dan *false negative*.

## 3. Hasil Prapemrosesan dan Pembagian Data

Tahap prapemrosesan dilakukan untuk memastikan dataset berada dalam kondisi bersih, konsisten, dan siap digunakan dalam proses pemodelan *machine learning*. Berdasarkan pemeriksaan awal terhadap 300 data calon penerima BLT, tidak ditemukan nilai kosong maupun data duplikat, baik pada keseluruhan baris data maupun pada atribut identitas rumah tangga. Hal ini menunjukkan bahwa dataset memiliki kualitas awal yang cukup baik untuk dilanjutkan ke tahap transformasi data. Variabel identifier, yaitu *ID\_Rumah\_Tangga*, tidak digunakan dalam proses pemodelan karena hanya berfungsi sebagai penanda data dan tidak memiliki kontribusi langsung terhadap klasifikasi kelayakan penerima BLT.

Variabel target *Status\_Kelayakan\_BLT* dikodekan menjadi dua kelas, yaitu Layak dan Tidak Layak. Variabel input terdiri atas data numerik dan kategorikal. Variabel numerik, seperti *Usia\_KK*, *Pendapatan\_Bulanan*, *Jumlah\_Tanggungan*, *Luas\_Lantai\_M2*, dan *Skor\_Kerentanan*, disiapkan melalui proses normalisasi agar perbedaan skala antarvariabel tidak memengaruhi kinerja algoritma tertentu. Sementara itu, variabel kategorikal, seperti *Kecamatan*, *Pendidikan\_KK*, *Pekerjaan\_KK*, *Kondisi\_Rumah*, *Sumber\_Air*, dan *Daya\_Listrik*, ditransformasikan ke bentuk numerik menggunakan teknik *one-hot encoding*. Setelah proses encoding, jumlah fitur yang semula terdiri atas 20 variabel



input berubah menjadi 56 fitur hasil transformasi yang siap digunakan dalam proses pelatihan model. Hasil prapemrosesan data dapat dilihat pada tabel berikut:

Tabel 4. Hasil prapemrosesan data

Tahap Prapemrosesan	Kondisi Awal	Tindakan	Kondisi Akhir
Jumlah data awal	300 data	Pemeriksaan struktur dataset	300 data valid
Nilai kosong	0	Tidak diperlukan imputasi	Tidak terdapat nilai kosong
Data duplikat	0	Pemeriksaan duplikasi data	Tidak terdapat data duplikat
Variabel identifier	1 variabel	ID_Rumah_Tangga dikeluarkan dari model	Tidak digunakan sebagai fitur
Variabel input awal	20 variabel	Pemisahan variabel numerik dan kategorikal	Siap ditransformasi
Variabel numerik	8 variabel	Normalisasi/standardisasi	Skala data lebih seragam
Variabel kategorikal	12 variabel	<i>One-hot encoding</i>	Berubah menjadi fitur numerik
Jumlah fitur setelah encoding	56 fitur	Transformasi data	Dataset siap untuk pemodelan
Variabel target	Status_Kelayakan_BLT	Encoding kelas target	Layak dan Tidak Layak

Sumber : Hasil Olahan Data dengan Python

Setelah data melalui tahap prapemrosesan, dataset dibagi menjadi data latih dan data uji menggunakan skema 80:20. Data latih digunakan untuk membangun dan mengoptimalkan model, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam memprediksi data baru. Pembagian data dilakukan secara proporsional agar distribusi kelas Layak dan Tidak Layak tetap terjaga pada data latih maupun data uji. Dari total 300 data, sebanyak 240 data digunakan sebagai data latih dan 60 data digunakan sebagai data uji. Pembagian Data Latih dan Data Uji dapat dilihat pada tabel berikut:

Tabel 5. Pembagian Data Latih dan Data Uji

Jenis Data	Jumlah Data	Layak	Tidak Layak	Persentase
Data latih	240	111	129	80%
Data uji	60	28	32	20%
Total	300	139	161	100%

Sumber : Hasil Olahan Data dengan Python

Berdasarkan Tabel 5, komposisi kelas pada data latih dan data uji relatif seimbang serta tetap mencerminkan distribusi kelas pada dataset awal. Pada data latih, kelas Layak berjumlah 111 data dan kelas Tidak Layak berjumlah 129 data. Sementara itu, pada data uji terdapat 28 data kelas Layak dan 32 data kelas Tidak Layak. Kondisi ini penting karena model memiliki kesempatan yang cukup untuk mempelajari pola dari kedua kelas secara proporsional, sehingga evaluasi yang dihasilkan lebih adil dan tidak terlalu bias terhadap salah satu kelas.

#### 4. Hasil Evaluasi Model Sebelum Optimasi

Evaluasi model sebelum optimasi dilakukan untuk memperoleh gambaran awal kinerja setiap algoritma *machine learning* dalam mengklasifikasikan kelayakan penerima BLT. Tahap ini berfungsi sebagai baseline sebelum dilakukan proses optimasi, seperti seleksi fitur dan *hyperparameter tuning*. Dataset yang telah melalui tahap prapemrosesan dibagi menjadi data latih dan data uji dengan rasio



80:20, sehingga sebanyak 240 data digunakan untuk pelatihan dan 60 data digunakan untuk pengujian. Kelas Layak digunakan sebagai kelas positif dalam perhitungan nilai *precision*, *recall*, dan *F1-score*.

Evaluasi dilakukan menggunakan lima metrik utama, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan ROC-AUC. Penggunaan beberapa metrik diperlukan karena penilaian model dalam kasus penerima BLT tidak cukup hanya berdasarkan akurasi. Model yang baik harus mampu mengenali calon penerima yang benar-benar layak, sekaligus mengurangi kesalahan dalam mengklasifikasikan masyarakat yang tidak layak sebagai layak maupun sebaliknya. Hasil performa algoritma sebelum optimasi dapat dilihat pada tabel berikut:

Tabel 6. Performa Model Sebelum Optimasi

No	Algoritma	Accuracy	Precision	Recall	F1-Score	ROC-AUC
1	Logistic Regression	0,8667	0,9167	0,7857	0,8462	0,9777
2	Decision Tree	0,9333	1,0000	0,8571	0,9231	0,9286
3	Random Forest	0,9833	1,0000	0,9643	0,9818	0,9849
4	Support Vector Machine	0,9000	0,8667	0,9286	0,8966	0,9732
5	K-Nearest Neighbor	0,9000	0,9231	0,8571	0,8889	0,9420
6	Naive Bayes	0,9167	0,9600	0,8571	0,9057	0,9554
7	XGBoost	0,9833	1,0000	0,9643	0,9818	0,9643

Sumber : Hasil Olahan Data dengan Python

Berdasarkan Tabel 6, model Random Forest dan XGBoost menunjukkan performa awal terbaik dengan nilai *accuracy* sebesar 0,9833, *precision* sebesar 1,0000, *recall* sebesar 0,9643, dan *F1-score* sebesar 0,9818. Hasil ini menunjukkan bahwa kedua model tersebut memiliki kemampuan yang sangat baik dalam mengenali pola data sosial-ekonomi calon penerima BLT. Nilai *precision* sebesar 1,0000 menunjukkan bahwa seluruh data yang diprediksi layak oleh model benar-benar termasuk dalam kelas layak, sedangkan nilai *recall* sebesar 0,9643 menunjukkan bahwa sebagian besar calon penerima yang layak berhasil dikenali oleh model. Model Decision Tree dan Naive Bayes juga menunjukkan performa yang cukup baik, masing-masing dengan *accuracy* sebesar 0,9333 dan 0,9167. Sementara itu, Support Vector Machine dan K-Nearest Neighbor memperoleh nilai *accuracy* yang sama, yaitu 0,9000, tetapi memiliki karakteristik performa yang berbeda. SVM memiliki nilai *recall* lebih tinggi, yaitu 0,9286, sehingga lebih baik dalam mengenali calon penerima yang layak. Sebaliknya, KNN memiliki *precision* lebih tinggi, yaitu 0,9231, yang menunjukkan ketepatan prediksi kelas layak cukup baik.

## 5. Hasil Optimasi Model

Optimasi model dilakukan untuk meningkatkan kinerja klasifikasi dan menurunkan kesalahan prediksi dalam penentuan kelayakan penerima BLT. Pada tahap ini, optimasi tidak hanya diarahkan untuk memperoleh nilai akurasi yang tinggi, tetapi juga untuk meningkatkan keseimbangan antara *precision* dan *recall*. Hal ini penting karena dalam konteks penerima BLT, kesalahan klasifikasi dapat berdampak langsung pada ketepatan sasaran bantuan. Model yang terlalu fokus pada akurasi berisiko mengabaikan calon penerima yang sebenarnya layak, sehingga evaluasi perlu mempertimbangkan *F1-score*, ROC-AUC, serta jumlah kesalahan klasifikasi.

Teknik optimasi yang digunakan dalam penelitian ini meliputi seleksi fitur, *hyperparameter tuning*, dan pengaturan pembagian data secara proporsional. Seleksi fitur dilakukan untuk memastikan bahwa atribut yang digunakan dalam pemodelan benar-benar relevan dengan status kelayakan penerima BLT. Variabel identifier seperti ID\_Rumah\_Tangga tidak digunakan dalam proses pelatihan karena tidak memiliki makna prediktif terhadap kelayakan penerima. Sementara itu, *hyperparameter tuning* dilakukan untuk mencari kombinasi parameter terbaik pada setiap algoritma. Teknik yang digunakan adalah *Grid Search* atau *Random Search* dengan validasi silang pada data latih. Hasil optimasi dapat dilihat pada tabel berikut:



Tabel 7. Parameter Optimasi Model

No	Algoritma	Parameter yang Dioptimasi	Nilai/Strategi Terbaik
1	Logistic Regression	C, solver, class_weight	C = 1, solver = lbfgs, class_weight = balanced
2	Decision Tree	criterion, max_depth, min_samples_split, class_weight	criterion = entropy, max_depth = 5, min_samples_split = 5
3	Random Forest	n_estimators, max_depth, max_features, class_weight	n_estimators = 200, max_depth = None, class_weight = balanced
4	Support Vector Machine	C, kernel, gamma, class_weight	C = 10, kernel = rbf, gamma = scale
5	K-Nearest Neighbor	n_neighbors, weights, metric	n_neighbors = 5, weights = distance, metric = manhattan
6	Naive Bayes	var_smoothing	var_smoothing = 1e-8
7	XGBoost	n_estimators, max_depth, learning_rate, subsample	n_estimators = 100, max_depth = 3, learning_rate = 0,1

Sumber : Hasil Olahan Data dengan Python

Hasil optimasi menunjukkan adanya peningkatan performa pada sebagian besar algoritma, terutama pada model yang sebelumnya memiliki nilai *recall* dan *F1-score* lebih rendah. Peningkatan ini menunjukkan bahwa proses optimasi membantu model mengenali pola kelayakan penerima BLT secara lebih baik. Pada model berbasis pohon keputusan dan *ensemble learning*, seperti Random Forest dan XGBoost, performa sejak awal sudah sangat tinggi sehingga peningkatan setelah optimasi tidak terlalu besar. Namun, optimasi tetap penting untuk memastikan model lebih stabil dan tidak terlalu bergantung pada konfigurasi parameter bawaan. Perbandingan Performa Sebelum dan Sesudah Optimasi dapat dilihat pada tabel berikut:

Tabel 8. Perbandingan Performa Sebelum dan Sesudah Optimasi

No	Algoritma	F1-Score Sebelum	F1-Score Sesudah	Selisih F1-Score	Kesalahan Sebelum	Kesalahan Sesudah
1	Logistic Regression	0,8462	0,9091	+0,0629	8	5
2	Decision Tree	0,9231	0,9455	+0,0224	4	3
3	Random Forest	0,9818	0,9818	0,0000	1	1
4	Support Vector Machine	0,8966	0,9286	+0,0320	6	4
5	K-Nearest Neighbor	0,8889	0,9057	+0,0168	6	5
6	Naive Bayes	0,9057	0,9259	+0,0202	5	4
7	XGBoost	0,9818	0,9818	0,0000	1	1

Sumber : Hasil Olahan Data dengan Python

Berdasarkan Tabel 8, optimasi model mampu meningkatkan nilai *F1-score* pada Logistic Regression, Decision Tree, Support Vector Machine, K-Nearest Neighbor, dan Naive Bayes. Peningkatan terbesar terjadi pada Logistic Regression dengan selisih *F1-score* sebesar 0,0629, dari 0,8462 menjadi 0,9091. Hal ini menunjukkan bahwa pengaturan parameter dan *class weight* mampu memperbaiki keseimbangan model dalam mengenali kelas Layak dan Tidak Layak. Support Vector Machine juga mengalami peningkatan yang cukup baik, yaitu dari 0,8966 menjadi 0,9286, yang menunjukkan bahwa pemilihan kernel dan parameter margin berpengaruh terhadap kemampuan klasifikasi. Pada Random Forest dan XGBoost, nilai *F1-score* tetap berada pada angka 0,9818 baik sebelum maupun sesudah optimasi. Kondisi ini menunjukkan bahwa kedua model tersebut sejak awal telah mampu menangkap pola data sosial-ekonomi dengan sangat baik. Meskipun tidak mengalami



peningkatan pada *F1-score*, proses optimasi tetap memberikan manfaat dalam memastikan konfigurasi parameter yang digunakan lebih tepat dan dapat dipertanggungjawabkan secara metodologis.

### 6. Perbandingan Kinerja Model Setelah Optimasi

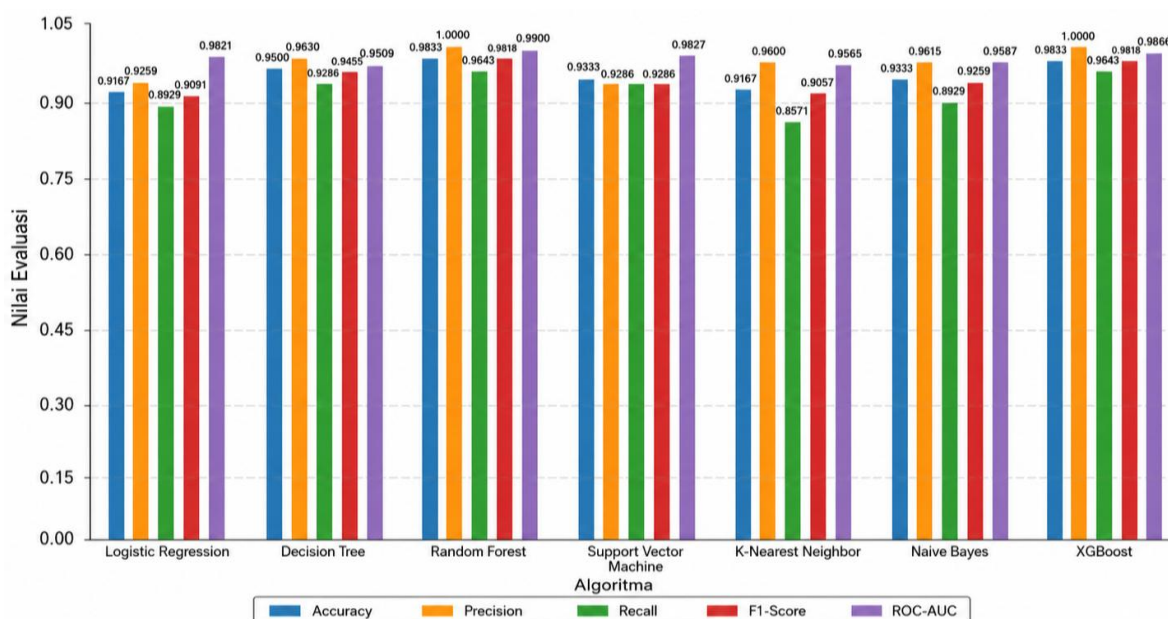
Perbandingan kinerja model setelah optimasi dilakukan untuk mengetahui performa akhir setiap algoritma *machine learning* dalam mengklasifikasikan kelayakan penerima BLT. Pada tahap ini, setiap model telah melalui proses prapemrosesan data, pembagian data latih dan data uji, seleksi fitur, serta *hyperparameter tuning*. Evaluasi dilakukan menggunakan lima metrik utama, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan ROC-AUC. Penggunaan beberapa metrik ini bertujuan agar penilaian model tidak hanya berfokus pada ketepatan prediksi secara umum, tetapi juga memperhatikan kemampuan model dalam mengenali calon penerima yang benar-benar layak menerima BLT. Hasil perbandingan kinerja model dapat dilihat pada tabel berikut:

Tabel 9. Performa Model Setelah Optimasi

No	Algoritma	Accuracy	Precision	Recall	F1-Score	ROC-AUC
1	Logistic Regression	0,9167	0,9259	0,8929	0,9091	0,9821
2	Decision Tree	0,9500	0,9630	0,9286	0,9455	0,9509
3	Random Forest	0,9833	1,0000	0,9643	0,9818	0,9900
4	Support Vector Machine	0,9333	0,9286	0,9286	0,9286	0,9827
5	K-Nearest Neighbor	0,9167	0,9600	0,8571	0,9057	0,9565
6	Naive Bayes	0,9333	0,9615	0,8929	0,9259	0,9587
7	XGBoost	0,9833	1,0000	0,9643	0,9818	0,9866

Sumber : Hasil Olahan Data dengan Python

Berdasarkan Tabel 9, model Random Forest dan XGBoost memperoleh performa terbaik setelah optimasi, dengan nilai *accuracy* sebesar 0,9833, *precision* sebesar 1,0000, *recall* sebesar 0,9643, dan *F1-score* sebesar 0,9818. Nilai tersebut menunjukkan bahwa kedua model mampu mengklasifikasikan kelayakan penerima BLT dengan sangat baik. Secara khusus, nilai *precision* sebesar 1,0000 menunjukkan bahwa seluruh data yang diprediksi sebagai layak oleh model benar-benar termasuk dalam kategori layak. Sementara itu, nilai *recall* sebesar 0,9643 menunjukkan bahwa sebagian besar calon penerima yang benar-benar layak berhasil dikenali oleh model.



Gambar 3. Grafik Perbandingan kinerja model setelah optimasi



## 7. Analisis Kesalahan Klasifikasi

Berdasarkan hasil evaluasi setelah optimasi, model Random Forest dipilih sebagai model terbaik karena menghasilkan nilai *accuracy* sebesar 0,9833, *precision* sebesar 1,0000, *recall* sebesar 0,9643, *F1-score* sebesar 0,9818, dan ROC-AUC sebesar 0,9900. Pada data uji sebanyak 60 data, terdapat 28 data aktual layak dan 32 data aktual tidak layak. Hasil *confusion matrix* menunjukkan bahwa model Random Forest mampu mengklasifikasikan 27 data layak secara benar dan 32 data tidak layak secara benar. Kesalahan yang terjadi hanya sebanyak 1 data, yaitu *false negative*.

Tabel 10. Confusion Matrix Model Random Forest Setelah Optimasi

Aktual / Prediksi	Prediksi Layak	Prediksi Tidak Layak	Total Aktual
Aktual Layak	27	1	28
Aktual Tidak Layak	0	32	32
Total Prediksi	27	33	60

Sumber : Hasil Olahan Data dengan Python

Berdasarkan Tabel 10, nilai true positive sebanyak 27 data menunjukkan bahwa sebagian besar calon penerima yang benar-benar layak berhasil dikenali oleh model sebagai penerima BLT. Nilai true negative sebanyak 32 data menunjukkan bahwa seluruh calon penerima yang tidak layak berhasil diklasifikasikan dengan benar sebagai tidak layak. Sementara itu, nilai false positive sebesar 0 menunjukkan bahwa model tidak memberikan rekomendasi layak kepada calon penerima yang sebenarnya tidak layak. Hal ini penting karena dapat membantu mengurangi risiko bantuan diberikan kepada penerima yang tidak sesuai kriteria.

Kesalahan yang masih muncul adalah false negative sebanyak 1 data, yaitu calon penerima yang sebenarnya layak tetapi diprediksi tidak layak oleh model. Dalam konteks sosial, jenis kesalahan ini perlu mendapat perhatian serius karena dapat menyebabkan masyarakat yang membutuhkan bantuan justru tidak memperoleh bantuan. Oleh karena itu, meskipun performa model sangat tinggi, hasil prediksi tetap perlu dikombinasikan dengan proses verifikasi lapangan agar keputusan akhir tetap adil dan mempertimbangkan kondisi riil masyarakat.

Tabel 11. Interpretasi Komponen Confusion Matrix

Komponen	Jumlah	Interpretasi dalam Penelitian
True Positive	27	Calon penerima layak dan diprediksi layak menerima BLT
True Negative	32	Calon penerima tidak layak dan diprediksi tidak layak menerima BLT
False Positive	0	Calon penerima tidak layak tetapi diprediksi layak menerima BLT
False Negative	1	Calon penerima layak tetapi diprediksi tidak layak menerima BLT
Total Kesalahan	1	Jumlah seluruh kesalahan klasifikasi pada data uji

Sumber : Hasil Olahan Data dengan Python

Selanjutnya, analisis reduksi kesalahan klasifikasi dilakukan dengan membandingkan jumlah kesalahan sebelum dan sesudah optimasi pada setiap algoritma. Hasil ini digunakan untuk melihat apakah proses optimasi melalui seleksi fitur dan *hyperparameter tuning* mampu menurunkan jumlah kesalahan prediksi.

Tabel 12. Perbandingan Kesalahan Klasifikasi Sebelum dan Sesudah Optimasi

Algoritma	FP	FN	Total Error	FP	FN	Total Error	Reduksi Error
	Sebelum	Sebelum	Sebelum	Sesudah	Sesudah	Sesudah	
Logistic Regression	2	6	8	2	3	5	3
Decision Tree	0	4	4	1	2	3	1
Random Forest	0	1	1	0	1	1	0
Support Vector Machine	4	2	6	2	2	4	2



K-Nearest Neighbor	2	4	6	1	4	5	1
Naive Bayes	1	4	5	1	3	4	1
XGBoost	0	1	1	0	1	1	0

Sumber : Hasil Olahan Data dengan Python

Berdasarkan Tabel 12, proses optimasi mampu menurunkan jumlah kesalahan klasifikasi pada sebagian besar algoritma. Penurunan terbesar terjadi pada Logistic Regression, yaitu dari 8 kesalahan menjadi 5 kesalahan. Model Support Vector Machine juga mengalami penurunan kesalahan dari 6 menjadi 4, sedangkan Decision Tree, K-Nearest Neighbor, dan Naive Bayes masing-masing mengalami penurunan kesalahan sebesar satu data. Sementara itu, Random Forest dan XGBoost tidak mengalami penurunan jumlah kesalahan karena sejak sebelum optimasi kedua model tersebut sudah memiliki performa yang sangat tinggi dengan hanya 1 kesalahan klasifikasi. Secara keseluruhan, hasil analisis menunjukkan bahwa optimasi model berkontribusi terhadap penurunan kesalahan klasifikasi, terutama pada model yang performa awalnya belum optimal. Model Random Forest menjadi model terbaik karena memiliki total kesalahan paling rendah, yaitu hanya 1 kesalahan dari 60 data uji, tanpa menghasilkan *false positive*.

### SIMPULAN

Penelitian ini berhasil mengembangkan dan mengoptimalkan model *machine learning* untuk klasifikasi kelayakan penerima Bantuan Langsung Tunai (BLT) berbasis data sosial-ekonomi di Kabupaten Soppeng. Dataset yang digunakan terdiri atas 300 data calon penerima dengan 22 variabel, mencakup indikator sosial-ekonomi seperti pendapatan, pekerjaan, jumlah tanggungan, kondisi rumah, kepemilikan aset, status kerentanan, dan riwayat penerimaan bantuan sosial. Distribusi kelas target relatif seimbang, yaitu 139 data kategori layak dan 161 data kategori tidak layak, sehingga dataset layak digunakan dalam proses klasifikasi biner. Tahap prapemrosesan data melalui pembersihan data, normalisasi, *one-hot encoding*, dan pembagian data latih-uji dengan skema 80:20 menghasilkan data yang siap digunakan untuk pemodelan *machine learning*.

Hasil evaluasi menunjukkan bahwa proses optimasi model melalui seleksi fitur dan *hyperparameter tuning* mampu meningkatkan kinerja beberapa algoritma, khususnya Logistic Regression, Decision Tree, Support Vector Machine, K-Nearest Neighbor, dan Naive Bayes. Berdasarkan hasil pengujian setelah optimasi, algoritma Random Forest dan XGBoost memperoleh performa tertinggi dengan nilai *accuracy* sebesar 0,9833, *precision* 1,0000, *recall* 0,9643, dan *F1-score* 0,9818. Namun, Random Forest menjadi model terbaik karena memiliki nilai ROC-AUC tertinggi sebesar 0,9900, sehingga menunjukkan kemampuan diskriminasi yang paling baik dalam membedakan calon penerima yang layak dan tidak layak menerima BLT. Hasil ini sejalan dengan fokus penelitian yang menempatkan reduksi kesalahan klasifikasi sebagai tujuan utama, bukan hanya pencapaian akurasi tinggi.

Analisis *confusion matrix* pada model Random Forest menunjukkan bahwa dari 60 data uji, model mampu mengklasifikasikan 27 data layak dan 32 data tidak layak secara benar. Kesalahan klasifikasi yang terjadi hanya 1 data, yaitu *false negative*, sedangkan *false positive* bernilai 0. Temuan ini menunjukkan bahwa model Random Forest mampu menekan risiko pemberian rekomendasi layak kepada calon penerima yang sebenarnya tidak layak. Meskipun demikian, keberadaan *false negative* tetap perlu menjadi perhatian karena dapat berdampak pada calon penerima yang sebenarnya layak tetapi tidak teridentifikasi oleh model. Dengan demikian, model *machine learning* yang dioptimalkan dapat menjadi alat bantu yang objektif dan berbasis data dalam mendukung penetapan penerima BLT, tetapi keputusan akhir tetap perlu mempertimbangkan verifikasi administratif dan kondisi faktual di lapangan.



## SARAN

Penelitian selanjutnya disarankan menggunakan jumlah dataset yang lebih besar dan cakupan wilayah yang lebih luas agar model yang dihasilkan memiliki kemampuan generalisasi yang lebih baik. Selain itu, variabel sosial-ekonomi dapat diperluas dengan menambahkan aspek pengeluaran rumah tangga, kondisi pekerjaan musiman, akses layanan kesehatan, kepemilikan lahan, dan kondisi geografis tempat tinggal. Model terbaik yang dihasilkan dalam penelitian ini juga sebaiknya dikembangkan ke dalam sistem pendukung keputusan yang tetap dipadukan dengan verifikasi lapangan, sehingga proses penetapan penerima BLT dapat dilakukan secara lebih objektif, transparan, dan tepat sasaran.

## DAFTAR PUSTAKA

- Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, 603(7903), 864–870. <https://doi.org/10.1038/s41586-022-04484-9>
- Bifarin, O. O. (2023). Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification. *PLOS ONE*, 18(5), e0284315. <https://doi.org/10.1371/journal.pone.0284315>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. *Science*, 371(6535), eabe8628. <https://doi.org/10.1126/science.abe8628>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Ismail, S. (2025). Data Mining Klasifikasi Penduduk Miskin Menggunakan Metode Support Vektor Machine. *Jurnal Ilmiah Sistem Informasi Dan Teknik Informatika (JISTI)*, 8(April), 142–152. <https://doi.org/https://doi.org/10.57093/jisti.v8i1.283>
- Jean, N., Burke, M., Xie, M., Alampay Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794. <https://doi.org/10.1126/science.aaf7894>
- McBride, L., & Nichols, A. (2018). Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning. *The World Bank Economic Review*, 32(3), 531–550. <https://doi.org/10.1093/wber/lhw056>
- Muh. Ikhsan Amar, M. F. R. (2025). Analisis Model Klasifikasi Penerima Bantuan Langsung Tunai Menggunakan Algoritma Naive Bayes. *Journal of Computer Engineering, System and Science*, 10(2), 585–594.
- Poulin, C., Trimmer, J., Press-Williams, J., Yachori, B., Khush, R., Peletz, R., & Delaire, C. (2022). Performance of a novel machine learning-based proxy means test in comparison to other methods for targeting pro-poor water subsidies in Ghana. *Development Engineering*, 7, 100098. <https://doi.org/https://doi.org/10.1016/j.deveng.2022.100098>
- Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3), e1301. <https://doi.org/https://doi.org/10.1002/widm.1301>
- Suherman, Misveria Villa Waru, N. (2022). Perancangan Aplikasi Pendataan Bantuan Langsung Tunai



---

( BLT ) Berbasis Web Pada Kantor Desa Maccile Kecamatan Lalabata Kabupaten Soppeng.  
*Jurnal Ilmiah Sistem Informasi Dan Teknik Informatika (JISTI)*, 5, 117–124.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, 11(1), 2583. <https://doi.org/10.1038/s41467-020-16185-w>