NEXUS : Journal of Information & Computational Science
Vol. 1, No. 1, January (2026) 21 - 30
e-ISSN XXXX-XXXX
p-ISSN XXXX-XXXX
https://ejournal.unupasuruan.ac.id/index.php/NEXUS

Received : July 31, 2025
Accepted : October 18, 2025
Available Online : January 11, 2026

KLASIFIKASI TINGKAT KEMISKINAN DI INDONESIA MENGGUNAKAN METODE LOGISTIC REGRESSION BERBASIS GOOGLE COLAB

Sapto Hadi Riono[1], Lailatul Imama[2], Inda Islami[3] , Adelia Puspitasari[4]
[1,2,3,4] Ilmu Komputer, Universitas PGRI Wiranegara
email: [1]saptoenator@gmail.com

Abstract: Poverty remains a major problem in Indonesia, requiring data-driven analysis to support targeted policy formulation. The development of machine learning provides an opportunity for more systematic analysis and classification of poverty levels. This study aims to classify poverty levels in Indonesia using the Logistic Regression method based on socioeconomic indicators. The variables used include poverty level, average years of schooling, per capita expenditure, Human Development Index (HDI), life expectancy, poverty level, and access to clean water. The research stages begin with data processing, participant data (training and testing data), model training, and model performance evaluation using confusion matrices and classification reports. The entire drawing and modeling process was carried out using the Google Colab platform the support of the scikit-learn library. The test results show that the Logistic Regression model is capable of producing good classification performance. Based on the evaluation results, class 0 obtained a precision value of 0.92, recall of 1.00, and an F1 score of 0.96, while class 1 obtained a precision value of 1.00, recall of 0.50, and an F1 score of 0.67. Overall, the model produced an accuracy score of 92.23%, indicating that the Logistic Regression method is quite effective classifying poverty levels in Indonesia.

Keywords: Poverty Level; Data Classification; *Logistic Regression*; Google Colab

Abstrak: Kemiskinan tetap menjadi masalah utama di Indonesia yang memerlukan analisis berbasis data untuk mendukung formulasi kebijakan yang ditargetkan. Pengembangan metode pembelajaran mesin (machine learning) memberikan peluang untuk melakukan analisis dan klasifikasi tingkat kemiskinan secara lebih sistematis. Studi ini bertujuan untuk mengklasifikasikan tingkat kemiskinan di Indonesia menggunakan metode Regresi Logistik berdasarkan indikator sosioekonomi. Variabel yang digunakan meliputi tingkat kemiskinan, rata-rata tahun sekolah, pengeluaran per kapita, Indeks Pembangunan Manusia (IPM), harapan hidup, tingkat pengangguran, dan akses terhadap air bersih. Tahapan penelitian dimulai dengan proses data, partisi data (data pelatihan dan pengujian), pelatihan model, dan evaluasi kinerja model. menggunakan matriks kebingungan dan laporan klasifikasi. Seluruh proses pemrosesan dan pemodelan dilakukan menggunakan platform Google Colab dengan dukungan perpustakaan scikit-learn. Hasil pengujian menunjukkan bahwa model Regresi Logistik mampu menghasilkan kinerja klasifikasi yang baik. Berdasarkan hasil evaluasi, kelas 0 memperoleh nilai presisi 0.92, recall 1.00, dan skor F1 0.96, sementara kelas 1 memperoleh nilai presisi 1.00, recall 0.50, dan skor F1 0.67. Secara keseluruhan, model menghasilkan skor akurasi sebesar 92,23%, menunjukkan bahwa metode Regresi Logistik cukup efektif dalam mengklasifikasikan tingkat kemiskinan di Indonesia.

Kata kunci: Tingkat Kemiskinan; Klasifikasi Data*; Logistic Regression*; Google Colab

PENDAHULUAN

Perbedaan tingkat kesejahteraan antarwilayah, keterbatasan kualitas pendidikan, serta minimnya akses terhadap sumber daya ekonomi turut berkontribusi terhadap tingginya angka kemiskinan di sejumlah daerah [1]. Berbagai kebijakan telah dijalankan oleh pemerintah untuk menekan tingkat kemiskinan, namun keberhasilan program tersebut sangat ditentukan oleh ketepatan analisis data serta pemahaman kondisi wilayah yang menjadi sasaran kebijakan [2]. Oleh sebab itu, dibutuhkan pendekatan analisis yang berbasis data untuk mendukung pengambilan keputusan yang lebih akurat dan objektif.

Seiring dengan perkembangan teknologi informasi, pemanfaatan metode machine learning dalam pengolahan data sosial ekonomi semakin banyak digunakan. Salah satu teknik yang umum diterapkan adalah metode klasifikasi, fungsinya adalah untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola dan karakteristik data tersebut [3]. *Logistic Regression* merupakan metode klasifikasi dengan pendekatan *supervised learning* yang mampu memodelkan hubungan antara variabel prediktor dan variabel target yang bersifat kategorikal. Sejumlah penelitian terdahulu menunjukkan bahwa metode ini cukup efektif dalam analisis data sosial ekonomi, termasuk dalam pengelompokan tingkat kemiskinan dan kesejahteraan masyarakat [4].

Berdasarkan latar belakang tersebut, penelitian ini membahas penerapan metode *Logistic Regression* dalam mengklasifikasikan tingkat kemiskinan di Indonesia menggunakan indikator sosial ekonomi yang tersedia. Tujuan penelitian ini adalah untuk mengimplementasikan metode *Logistic Regression* dalam proses klasifikasi serta mengevaluasi kinerja model berdasarkan nilai akurasi dan hasil evaluasi klasifikasi. Studi ini diharapkan dapat menjadi panduan untuk pemanfaatan pendekatan berbasis data dalam analisis kemiskinan di Indonesia [5].

METODOLOGI

Metode klasifikasi berbasis machine learning digunakan dalam penelitian ini sebagai pendekatan kuantitatif [6]. Metode yang diterapkan adalah *Logistic Regression*, yang termasuk dalam kategori *supervised learning* [7]. Pendekatan ini dipilih karena memiliki kemampuan untuk mengkategorikan data ke dalam kelas tertentu berdasarkan hubungan antara variabel bebas dan variabel terikat, yang merupakan teknik kategorikal [8].

Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang berkaitan dengan tingkat kemiskinan di Indonesia [9]. Dataset diperoleh dari platform Kaggle dengan judul "Klasifikasi Tingkat Kemiskinan di Indonesia", yang disusun berdasarkan data statistik resmi yang berasal dari data yang dikumpulkan oleh Badan Pusat Statistik (BPS) Republik Indonesia [10].

Dataset tersebut memuat berbagai indikator sosial dan ekonomi yang digunakan untuk menggambarkan kondisi kemiskinan, di antaranya Indeks Pembangunan Manusia (IPM), rata-rata durasi sekolah, pengeluaran per kapita, dan persentase penduduk miskin, serta beberapa variabel pendukung lainnya yang relevan. Data telah melalui tahap seleksi dan penyesuaian agar sesuai dengan kebutuhan penelitian dan metode yang digunakan. Lalu seluruh data kemudian diolah dan dianalisis menggunakan platform Google Colab dengan menerapkan metode *Logistic Regression* untuk melakukan klasifikasi tingkat kemiskinan di Indonesia.

Training Model Logistic Regression

Pada tahap pelatihan, metode *Logistic Regression* diterapkan menggunakan data latih yang telah ditentukan sebelumnya. Model dibangun untuk mengidentifikasi keterkaitan antara berbagai indikator sosial ekonomi dengan kategori tingkat kemiskinan. Melalui proses ini, model mempelajari pola data sehingga mampu memperkirakan peluang suatu data termasuk ke dalam kelas tertentu [11].

Pelatihan model bertujuan untuk menghasilkan model klasifikasi yang memiliki kemampuan prediksi yang baik. Penyesuaian parameter dilakukan selama proses pelatihan agar mendapatkan hasil yang optimal

Model Evaluation Based on Accuracy

Setelah proses pelatihan selesai, model diuji menggunakan data uji untuk mengevaluasi kinerjanya. Evaluasi model dilakukan dengan membandingkan hasil prediksi model terhadap label aktual pada data uji. Salah satu metrik yang digunakan adalah nilai akurasi, yang menunjukkan tingkat ketepatan model dalam mengklasifikasikan tingkat kemiskinan.

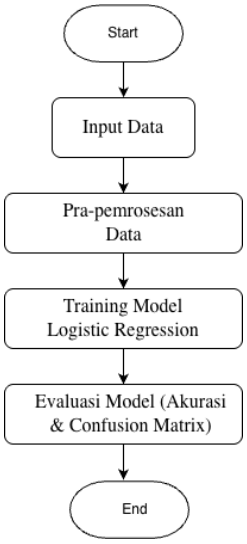
Selain akurasi, evaluasi juga dilakukan menggunakan *classification report* yang mencakup nilai *precision, recall, dan f1-score* untuk setiap kelas. Metrik tersebut digunakan untuk memberikan gambaran yang lebih rinci mengenai kemampuan model dalam melakukan klasifikasi pada masing-masing kelas.

Confusion matrix turut digunakan untuk melihat distribusi hasil prediksi model, baik prediksi yang benar maupun kesalahan klasifikasi. Dengan kombinasi metrik evaluasi tersebut, kinerja model *Logistic Regression* dapat dianalisis secara lebih komprehensif [12].

Evaluasi Model Berdasarkan Confusion Matrix

Selain menggunakan nilai akurasi, evaluasi kinerja model juga dilakukan dengan *confusion matrix*. Confusion Matrix dapat digunakan untuk menunjukkan perbandingan antara hasil prediksi model dan data aktual secara lebih rinci. Matriks ini juga dapat menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar dan jumlah data yang mengalami kesalahan klasifikasi dalam setiap kelas. Penggunaan *confusion matrix* memberikan gambaran yang lebih menyeluruh terhadap performa model *Logistic Regression* dalam mengklasifikasikan tingkat kemiskinan.

Flowchart penelitian ini ditunjukkan pada Gambar 1 menggambarkan Langkah – Langkah penelitian yang dilakukan. Proses diawali dengan tahap input data sebagai bahan utama penelitian. Selanjutnya, data melalui tahap pra-pemrosesan untuk meningkatkan kualitas dan kesiapan data sebelum dilakukan pemodelan. Data yang telah diproses kemudian digunakan pada tahap pelatihan model menggunakan metode *Logistic Regression* untuk membangun model klasifikasi. Tahap akhir penelitian adalah evaluasi kinerja model menggunakan metrik akurasi dan *confusion matrix* guna menilai kemampuan model dalam melakukan klasifikasi tingkat kemiskinan.



Gambar 1. Diagram Alur Metodologi Penelitian.

Untuk memperjelas karakteristik data pada tahap input, Tabel 1 menyajikan deskripsi masing-masing variabel yang digunakan beserta satuan atau tipe datanya. Penyajian tabel ini bertujuan untuk memberikan gambaran awal mengenai variabel penelitian sebelum data diproses lebih lanjut pada tahap pra-pemrosesan dan pemodelan.

Tabel 1. Deskripsi Variabel Penelitian

Nama Variabel	Deskripsi	Satuan / Tipe Data
Persentase Penduduk Miskin	Persentase penduduk yang berada di bawah garis kemiskinan	Persen (%)
Rata-rata Lama Sekolah	Rata-rata jumlah tahun pendidikan yang ditempuh penduduk	Tahun
Pengeluaran Per Kapita	Rata-rata pengeluaran penduduk per kapita	Rupiah
Indeks Pembangunan Manusia	Indeks yang menggambarkan kualitas hidup penduduk	Indeks
Umur Harapan Hidup	Perkiraan rata-rata usia hidup penduduk	Tahun
Tingkat Pengangguran	Persentase penduduk usia kerja yang belum bekerja	Persen (%)
Akses Air Bersih	Persentase rumah tangga memiliki akses ke air bersih	Persen (%)
Tingkat Kemiskinan	Kategori tingkat kemiskinan hasil klasifikasi	Kategorikal (0/1)

Variabel-variabel yang digunakan dalam penelitian ini sebagaimana ditunjukkan pada Tabel 1 selanjutnya diproses menggunakan metode *Logistic Regression*. Setiap variabel independen direpresentasikan sebagai $x_1, x_2, \ldots, x_n$dalam model matematis, sedangkan tingkat kemiskinan digunakan sebagai variabel dependen. Hubungan antara variabel independen dan variabel dependen dimodelkan menggunakan fungsi sigmoid untuk menghasilkan nilai probabilitas klasifikasi.

$$p(y = 1|\, x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_{2+\cdots} + \beta_n x_n}} \quad (1)$$

Dimana :

β_0 : konstanta (*intercept*) yang menjadi nilai awal model saat seluruh variabel bernilai nol

$\beta_0, \beta_1, \ldots, \beta_n$: koefisien yang menunjukkan pengaruh masing-masing variabel independen

$x_1, x_2, \ldots, x_n$: variabel independen atau indikator sosial ekonomi

p : probabilitas suatu data termasuk ke dalam kelas tertentu (misalnya tingkat kemiskinan)

HASIL DAN PEMBAHASAN

Dipaparkan hasil penelitian yang diperoleh dari penerapan metode *Logistic Regression* dalam proses klasifikasi tingkat kemiskinan di Indonesia berdasarkan indikator sosial ekonomi yang digunakan. Implementasi model dilakukan dengan memanfaatkan platform Google Colab, sementara *source code* digunakan sebagai pendukung untuk menjelaskan tahapan pemodelan yang dilakukan.

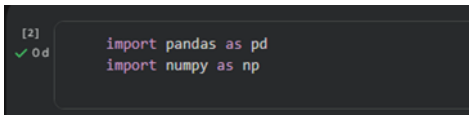
Selanjutnya, hasil evaluasi dari model klasifikasi dianalisis dan dibahas secara konseptual untuk memberikan gambaran mengenai kinerja serta kemampuan model *Logistic Regression* dalam mengelompokkan tingkat kemiskinan.

Hasil Klasifikasi Tingkat Kemiskinan Menggunakan Logistic Regression

Hasil klasifikasi tingkat kemiskinan diperoleh setelah model *Logistic Regression* melalui proses pelatihan menggunakan data latih dan dilakukan pengujian menggunakan data uji. Seluruh proses pemodelan diimplementasikan pada platform Google Colab dengan memanfaatkan pustaka *scikit-learn*. Persentase penduduk miskin, rata-rata lama sekolah, pengeluaran per kapita, umur harapan hidup, tingkat pengangguran, akses terhadap air bersih, dan Indeks Pembangunan Manusia adalah variabel independen yang digunakan dalam penelitian ini.

Berdasarkan hasil pengujian, model *Logistic Regression* menunjukkan kemampuan yang cukup baik dalam mengklasifikasikan tingkat kemiskinan ke dalam kategori yang telah ditentukan. Hal ini tercermin dari nilai akurasi yang diperoleh melalui perbandingan antara hasil prediksi model dan data aktual pada data uji. Nilai akurasi tersebut menunjukkan bahwa sebagian besar data dapat diklasifikasikan dengan tepat. Temuan ini mengindikasikan bahwa indikator sosial ekonomi yang digunakan memiliki keterkaitan yang kuat dengan tingkat kemiskinan dan relevan untuk digunakan dalam proses klasifikasi.

Hasil Output



Pada tahap awal, dilakukan pemanggilan pustaka yang diperlukan untuk proses pengolahan data. Pustaka pandas digunakan untuk membaca serta mengelola data dalam bentuk DataFrame, sehingga memudahkan proses manipulasi dan analisis data. Sementara itu, pustaka numpy dimanfaatkan untuk mendukung operasi numerik dan pengolahan data berbasis array yang dibutuhkan dalam proses komputasi.



Pada tahap ini, pustaka pandas digunakan untuk memuat dataset tingkat kemiskinan Indonesia ke dalam lingkungan Google Colab. Dataset dibaca menggunakan fungsi read_csv dengan pemisah data berupa tanda titik koma (;), kemudian disimpan dalam bentuk *DataFrame*. Data yang telah dimuat selanjutnya digunakan sebagai input pada tahap pra-pemrosesan dan pemodelan menggunakan metode *Logistic Regression.*

```
[22]
✓ 0d
df = df.dropna(subset=['Klasifikasi Kemiskinan'])
```

Pada tahap pra-pemrosesan data, dilakukan pembersihan dataset dengan menghapus baris data yang memiliki nilai kosong pada variabel Klasifikasi Kemiskinan. Langkah ini bertujuan untuk memastikan bahwa seluruh data yang digunakan dalam proses pelatihan dan pengujian model memiliki label kelas yang lengkap, sehingga kualitas dan keandalan hasil klasifikasi dapat terjaga.

```
[23]
✓ 0d
print(df.columns.tolist())

['Provinsi', 'Kab/Kota', 'Persentase Penduduk Miskin (P0)
```

Hasil peninjauan struktur dataset menunjukkan bahwa data terdiri dari beberapa variabel yang merepresentasikan kondisi sosial dan ekonomi di tingkat provinsi serta kabupaten/kota. Variabel-variabel tersebut meliputi informasi wilayah, persentase penduduk miskin, rata-rata lama sekolah, pengeluaran per kapita, Indeks Pembangunan Manusia, umur harapan hidup, akses sanitasi layak, akses air minum layak, tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, serta Produk Domestik Regional Bruto (PDRB). Selain itu, dataset juga memuat variabel Klasifikasi Kemiskinan yang digunakan sebagai label kelas dalam proses klasifikasi.

```
features = [
    'Persentase Penduduk Miskin (P0) Menurut Kabupaten/Kota (Persen)',
    'Rata-rata Lama Sekolah Penduduk 15+ (Tahun)',
    'Pengeluaran per Kapita Disesuaikan (Ribu Rupiah/Orang/Tahun)',
    'Indeks Pembangunan Manusia',
    'Umur Harapan Hidup (Tahun)',
    'Persentase rumah tangga yang memiliki akses terhadap sanitasi layak',
    'Persentase rumah tangga yang memiliki akses terhadap air minum layak',
    'Tingkat Pengangguran Terbuka',
    'Tingkat Partisipasi Angkatan Kerja',
    'PDRB atas Dasar Harga Konstan menurut Pengeluaran (Rupiah)'
]

X = df[features]
y = df['Klasifikasi Kemiskinan']
```

Pada tahap ini dilakukan penentuan variabel yang digunakan dalam proses pemodelan. Sejumlah indikator sosial ekonomi dipilih sebagai variabel independen (X), yang meliputi persentase penduduk miskin, rata-rata lama sekolah, pengeluaran per kapita, Indeks Pembangunan Manusia, umur harapan hidup, akses sanitasi layak, akses air minum layak, tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, serta Produk Domestik Regional Bruto (PDRB). Selanjutnya, variabel Klasifikasi Kemiskinan ditetapkan sebagai variabel dependen (y) yang berperan sebagai label kelas dalam proses klasifikasi menggunakan metode *Logistic Regression*.

```
[24]
✓ 0d
X = X.select_dtypes(include=['int64', 'float64'])
X = X.copy()

for col in X.columns:
    X[col].fillna(X[col].median(), inplace=True)

... /tmp/ipython-input-1010454139.py:5: FutureWarning: A value is tryin
The behavior will change in pandas 3.0. This inplace method will no

for example, when doing 'df[col].method(value, inplace=True)', try

  X[col].fillna(X[col].median(), inplace=True)
```

Pada tahap ini dilakukan proses pra-pemrosesan data dengan menyeleksi variabel numerik yang akan digunakan dalam pemodelan. Data yang dipilih hanya mencakup variabel dengan tipe numerik, yaitu bilangan bulat dan bilangan desimal, untuk memastikan kesesuaian dengan metode *Logistic Regression*.

Selanjutnya, dilakukan penanganan nilai kosong pada setiap variabel numerik dengan menggantinya menggunakan nilai median masing-masing variabel. Langkah ini bertujuan untuk menjaga kelengkapan data serta menghindari gangguan pada proses pelatihan dan pengujian model akibat adanya nilai yang hilang.

Pada titik ini, fungsi *train_test_split* dari pustaka *scikit-learn* digunakan untuk membagi dataset menjadi data uji dan data latih. Data dibagi menjadi 20% sebagai data uji dan 80% sebagai data latih. Data uji digunakan untuk mengevaluasi kinerja model, sedangkan data latih digunakan untuk membangunnya. Untuk memastikan bahwa hasil pembagian data konsisten, pengaturan random state digunakan.

```
model = LogisticRegression(max_iter=1000)
model.fit(X_train, y_train)
```

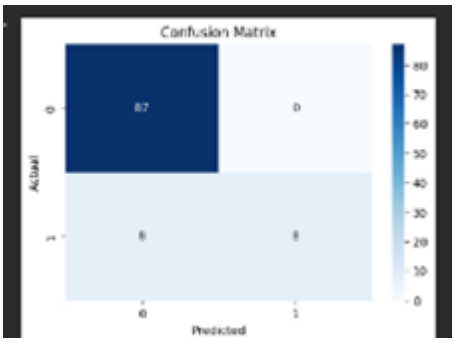
```
LogisticRegression
LogisticRegression(max_iter=1000)
```

Pada tahap ini dilakukan pembuatan dan pelatihan model menggunakan algoritma *Logistic Regression* dari pustaka *scikit-learn*. Model diinisialisasi dengan parameter *max_iter* sebesar 1000 untuk memastikan proses optimasi dapat mencapai kondisi konvergen. Selanjutnya, model dilatih menggunakan data latih (X_train) dan label target (y_train) sehingga model siap digunakan pada tahap prediksi dan evaluasi kinerja

```
Akurasi: 0.9223300970873787

Classification Report:
              precision    recall  f1-score   support

           0       0.92      1.00      0.96        87
           1       1.00      0.50      0.67        16

    accuracy                           0.92       103
   macro avg       0.96      0.75      0.81       103
weighted avg       0.93      0.92      0.91       103
```

Pada tahap evaluasi model, dilakukan prediksi terhadap data uji menggunakan model *Logistic Regression* yang telah dilatih. Hasil tersebut kemudian dibandingkan dengan label aktual untuk menilai kinerja model. Evaluasi dilakukan menggunakan beberapa metrik, yaitu akurasi serta *classification report* yang mencakup nilai *precision*, *recall*, dan *f1-score* pada masing-masing kelas.

Metrik-metrik tersebut digunakan untuk memberikan gambaran mengenai tingkat ketepatan model dalam melakukan klasifikasi, serta kemampuan model untuk membedakan tingkat kemiskinan secara keseluruhan untuk setiap kelas.



Confusion matrix divisualisasikan dalam bentuk *heatmap* untuk mengevaluasi kinerja model klasifikasi. Visualisasi ini menampilkan perbandingan antara label aktual dan hasil prediksi model pada setiap kelas. Setiap sel pada matriks menunjukkan jumlah data yang diklasifikasikan dengan benar maupun yang mengalami kesalahan klasifikasi.

Penggunaan *heatmap* pada *confusion matrix* bertujuan untuk mempermudah interpretasi hasil evaluasi, sehingga pola kesalahan dan tingkat ketepatan model dalam membedakan masing-masing kelas tingkat kemiskinan dapat diamati secara lebih jelas.

Evaluasi Kinerja Model Menggunakan *Confusion Matrix* dan *Classification*

Evaluasi kinerja model dilakukan untuk mengetahui kemampuan metode *Logistic Regression* dalam mengklasifikasikan tingkat kemiskinan

berdasarkan indikator sosial ekonomi yang digunakan. Pada penelitian ini, proses evaluasi tidak hanya mengacu pada nilai akurasi, tetapi juga didukung oleh penggunaan *confusion matrix* dan *classification report* guna memperoleh gambaran performa model secara lebih rinci dan menyeluruh.

Variabel target Klasifikasi Kemiskinan direpresentasikan dalam bentuk numerik melalui proses *label encoding*, di mana kelas 0 menunjukkan kategori tingkat kemiskinan yang lebih rendah, sedangkan kelas 1 menunjukkan kategori tingkat kemiskinan yang lebih tinggi. Pengkodean ini bertujuan untuk memudahkan proses pemodelan dan evaluasi menggunakan metode *Logistic Regression*.

Berdasarkan hasil pengujian menggunakan data uji, model menghasilkan nilai akurasi sebesar sekitar 92,23%. Nilai tersebut menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar oleh model. Hasil evaluasi menggunakan *confusion matrix* memperlihatkan bahwa jumlah prediksi yang tepat lebih dominan dibandingkan dengan kesalahan klasifikasi pada masing-masing kelas tingkat kemiskinan. Hal ini mengindikasikan bahwa model mampu membedakan antara kelas dengan tingkat kemiskinan yang berbeda secara cukup baik, meskipun masih terdapat sejumlah kecil kesalahan prediksi.

Selain itu, hasil *classification report* menunjukkan bahwa model memiliki performa yang sangat baik pada kelas 0, yang ditunjukkan oleh nilai *recall* yang tinggi, sehingga seluruh data pada kelas tersebut dapat dikenali dengan baik oleh model. Sementara itu, pada kelas 1, nilai *precision* menunjukkan tingkat ketepatan prediksi yang baik, meskipun nilai *recall* masih relatif lebih rendah. Kondisi ini mengindikasikan bahwa model cenderung lebih berhati-hati dalam memprediksi kelas tingkat kemiskinan yang lebih tinggi, yang kemungkinan dipengaruhi oleh distribusi data yang tidak seimbang antar kelas.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *Logistic Regression* memiliki kinerja yang cukup stabil dan konsisten dalam mengklasifikasikan tingkat kemiskinan. Dengan demikian, metode ini dapat dijadikan sebagai salah satu pendekatan yang efektif dan dapat diandalkan dalam analisis serta pengelompokan tingkat kemiskinan di Indonesia.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode *Logistic Regression* mampu digunakan untuk mengklasifikasikan tingkat kemiskinan di Indonesia dengan memanfaatkan berbagai indikator sosial ekonomi yang tersedia. Proses pemodelan dan evaluasi yang dilakukan menggunakan platform Google Colab menunjukkan bahwa model dapat menghasilkan klasifikasi dengan tingkat ketepatan yang cukup baik. Hasil evaluasi menggunakan *classification report* menunjukkan adanya perbedaan kinerja model pada masing-masing kelas. Pada kelas 0, model menghasilkan nilai *precision* sebesar 0,92, *recall* sebesar 1,00, dan *f1-score* sebesar 0,96, yang mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam mengenali data pada kelas tersebut. Sementara itu, pada kelas 1 diperoleh nilai *precision* sebesar 1,00, *recall* sebesar 0,50, dan *f1-score* sebesar 0,67. Hasil ini menunjukkan bahwa meskipun prediksi model pada kelas 1 tergolong

tepat, kemampuan model dalam mengenali seluruh data pada kelas tersebut masih terbatas.

Secara keseluruhan, model *Logistic Regression* menghasilkan nilai akurasi sebesar 92,23%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Indikator sosial ekonomi seperti persentase penduduk miskin, tingkat pendidikan, pengeluaran per kapita, dan Indeks Pembangunan Manusia memiliki kontribusi yang signifikan dalam proses klasifikasi tingkat kemiskinan. Oleh karena itu, metode regresi logistik dapat digunakan sebagai salah satu metode alternatif untuk menganalisis dan mengorganisasikan tingkat kemiskinan di Indonesia. Untuk penelitian selanjutnya, disarankan dilakukan penanganan ketidakseimbangan data antar kelas atau perbandingan dengan metode klasifikasi lain guna meningkatkan kinerja model.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada semua orang yang telah membantu, membantu, dan berkontribusi pada pelaksanaan penelitian ini. Khususnya, mereka berterima kasih kepada dosen pembimbing yang telah dengan sabar memberikan bimbingan, arahan, dan saran yang bermanfaat selama seluruh proses perencanaan, pelaksanaan, dan penyusunan laporan penelitian. Selain itu, penulis mengucapkan terima kasih kepada semua organisasi dan lembaga terkait yang telah memberikan data dan bantuan yang diperlukan untuk penelitian ini. Selain itu, penghargaan juga diberikan kepada semua pihak yang telah memberikan bantuan moral, saran, dan bantuan secara langsung maupun tidak langsung selama proses penelitian. Semoga penelitian ini bermanfaat dan membantu pengembangan ilmu pengetahuan. Selain itu, semoga menjadi referensi untuk penelitian yang akan datang.

REFERENCES

[1] Ermila. (n.d.). *Klasifikasi Tingkat Kemiskinan di Indonesia*. Retrieved from kaggle: https://www.kaggle.com/datasets/ermila/klasifikasi-tingkat-kemiskinan-di-indonesia

[2] Hidayat, M. R. (2021). “Evaluasi Kinerja Model Klasifikasi Menggunakan Confusion Matrix dan Classification Report". *Jurnal Informatika dan Komputasi, vol. 9, no. 1*.

[3] Immanudin, N. d. (2019). “Penerapan Metode Klasifikasi Data dalam Analisis Sosial Ekonomi”. *Komputika: Jurnal Sistem Komputer, vol. 8, no. 2*.

[4] Indonesia, K. K. (2023). *“Statistik Kemiskinan dan Kesejahteraan"*. Retrieved from https://www.kemenkeu.go.id

[5] R. Kustijono, T. S. (2019). “Pemanfaatan Analisis Data dalam Pengambilan Keputusan Berbasis Sosial Ekonomi”. *Jurnal Ilmu Komputer dan Informatika, vol. 8, no. 1*.

[6] Sapto Hadi Riono, R. D. (2023). Myedusolve personality clustering analysis using k-means machine learningalgorithm. *AMCA JOURNAL OF SCIENCE AND TECHNOLOGY*, 40-42.

[7] Sapto Hadi Riono, R. H. (2025). Clustering Analysis of Rice Productivity of Regions in Three Major Provinces of Java Island Using Python-based K-Means Algorithm. *SMARTICS Journal*, 1-8.

[8] Saputra, A. W. (2022). "Penerapan Machine Learning untuk Analisis Tingkat Kesejahteraan”. *Jurnal Teknologi Informasi, vol. 15, no. 2*.

[9] Setyawan, H. P. (2020). “Analisis Faktor-Faktor Kemiskinan Menggunakan Model Regresi Logistik". *Jurnal Ekonomi & Statistik, vol. 12, no. 1*.

[10] Statistik, B. P. (2023). *“Kemiskinan dan Ketimpangan Penduduk Indonesia"*. Retrieved from https://www.bps.go.id

[11] Wahono, A. S. (2015). “Analisis dan Penerapan Metode Klasifikasi Data untuk Permasalahan Sosial”. *Journal of Intelligent System, vol. 1, no. 2*.

[12] F. A. Fitrony, L. D. Supraba, T. Rantung, I. M. A. Agastya, and K. Kusrini, “Analysis to Predict the Number of New Students At UNU Pasuruan using Arima Method”, SISFOKOM, vol. 14, no. 1, pp. 123–128, Jan. 2025.