


Analisa Perbandingan Metode Klasifikasi Data Mining untuk Menentukan Tingkat Kemiskinan

Amser Pangaribuan^{1*}, Muhammad Rafi Al Latif², Alfian Panji Syahputra³, Muhammad Fauzan⁴, Fadli Azhima⁵, Mohammad Naufal Fathur Rahman⁶

^{1,2,3,4,5,6}Sistem Informasi, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta

E-mail: pangaribuanamser1@gmail.com

* Corresponding Author

 <https://doi.org/10.31004/jerkin.v4i3.5077>

ARTICLE INFO

Article history

Received: 25 Dec 2025

Revised: 05 Jan 2026

Accepted: 16 Jan 2026

Kata Kunci:

Klasifikasi Data, Dataset Kemiskinan, Miskin dan Tidak Miskin, Data Mining, Sistem Informasi

Keywords:

Data Classification, Poverty Dataset, Poor and Non-Poor, Data Mining, Information Systems



ABSTRACT

Kemiskinan merupakan permasalahan sosial yang memerlukan pengelolaan dan analisis data yang tepat untuk mendukung pengambilan keputusan. Penelitian ini bertujuan untuk melakukan klasifikasi status kemiskinan masyarakat ke dalam dua kategori, yaitu miskin dan tidak miskin, berdasarkan dataset sosial ekonomi. Dataset yang digunakan melalui tahap praproses data yang meliputi pembersihan data dan penyesuaian atribut. Proses pengolahan data dilakukan dengan memanfaatkan teknik klasifikasi pada data mining menggunakan aplikasi pengolah data. Evaluasi model dilakukan untuk menilai kemampuan klasifikasi berdasarkan hasil pengujian data. Hasil penelitian menunjukkan bahwa dataset yang digunakan mampu mendukung proses klasifikasi status kemiskinan secara efektif. Penelitian ini diharapkan dapat menjadi dasar pengembangan sistem informasi yang mendukung pengambilan keputusan dalam penentuan status kemiskinan masyarakat. *Klasifikasi Data, Dataset Kemiskinan, Miskin dan Tidak Miskin, Data Mining, Sistem Informasi*

*Poverty is a social problem that requires proper data management and analysis to support decision-making. This study aims to classify the poverty status of the community into two categories, namely poor and non-poor, based on a socioeconomic dataset. The dataset used went through a data preprocessing stage that included data cleaning and attribute adjustment. The data processing process was carried out using classification techniques in data mining using a data processing application. Model evaluation was conducted to assess the classification capability based on the results of data testing. The results of the study indicate that the dataset used is able to support the process of classifying poverty status effectively. This research is expected to become the basis for the development of information systems that support decision-making in determining the poverty status of the community. *Data Classification, Poverty Dataset, Poor and Non-Poor, Data Mining, Information Systems**



This is an open access article under the CC-BY-SA license.

How to Cite: Amser Pangaribuan, et al (2026). Analisa Perbandingan Metode Klasifikasi Data Mining untuk Menentukan Tingkat Kemiskinan, 4(3) 17645-17652. <https://doi.org/10.31004/jerkin.v4i3.5077>

PENDAHULUAN

Indonesia merupakan salah satu Negara berkembang di asia khususnya asia tenggara. Salah satu masalah yang sering dihadapi oleh Negara berkembang adalah kemiskinan. Berdasarkan data yang dikeluarkan oleh Badan Pusat Statistik, angka kemiskinan di Indonesia pada tahun 1999 mencapai 47.97 juta jiwa. Pada tahun 2011 jumlah penduduk miskin menjadi 30.02 juta jiwa. Badan Pusat Statistik melakukan pendataan kependudukan khususnya masalah kemiskinan setiap 3 tahun sekali. Proses pendataan dilakukan dengan cara door to door langsung menuju rumah tangga sasaran.(Antara et al., 2015). Berdasarkan data tersebut, dapat dilihat bahwa permasalahan kemiskinan masih menjadi isu penting meskipun jumlah penduduk miskin mengalami penurunan. Proses pendataan yang dilakukan secara langsung menghasilkan dataset kependudukan yang cukup besar dan beragam. Dataset ini

memiliki potensi untuk dianalisis lebih lanjut guna memperoleh informasi yang lebih mendalam terkait kondisi sosial ekonomi masyarakat. Oleh karena itu, diperlukan pengolahan dan analisis data yang sistematis agar data kemiskinan dapat dimanfaatkan secara optimal dalam mendukung pengambilan keputusan. Kemiskinan adalah kondisi ketidakmampuan ekonomi untuk memenuhi standar hidup rata-rata masyarakat di suatu wilayah. Kondisi ketidakmampuan ini ditandai dengan pendapatan rendah yang tidak mampu memenuhi kebutuhan dasar (Jurnal 2 Internasional). Berdasarkan pengertian tersebut, kemiskinan tidak hanya dipengaruhi oleh faktor pendapatan, tetapi juga berkaitan dengan berbagai aspek sosial dan ekonomi lainnya. Dengan memanfaatkan dataset kemiskinan yang tersedia, proses klasifikasi dapat dilakukan untuk mengelompokkan masyarakat ke dalam kategori miskin dan tidak miskin. Hasil klasifikasi ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai kondisi kemiskinan serta menjadi dasar dalam pengembangan sistem informasi pendukung pengambilan keputusan.

METODE

Knime

KNIME (Konstanz Information Miner) merupakan sebuah perangkat lunak analisis data berbasis visual workflow yang digunakan untuk melakukan proses data mining, machine learning, dan data analytics. KNIME dirancang untuk memudahkan pengguna dalam membangun alur analisis data secara sistematis tanpa memerlukan kemampuan pemrograman yang kompleks, meskipun tetap menyediakan fleksibilitas untuk integrasi bahasa pemrograman seperti Python dan R.

KNIME menggunakan konsep node-based workflow, di mana setiap node merepresentasikan satu tahapan proses analisis data, seperti data input, data preprocessing, pemodelan, hingga evaluasi. Node-node tersebut dihubungkan secara visual membentuk suatu alur kerja (workflow) yang transparan dan mudah direproduksi. Pendekatan ini mendukung prinsip reproducible research karena seluruh proses analisis dapat ditelusuri dan dijalankan kembali secara konsisten. (Yudana et al., 2023)

Dalam konteks analisis klasifikasi, KNIME menyediakan berbagai algoritma supervised learning seperti Decision Tree, Naive Bayes, K-Nearest Neighbor, Support Vector Machine, dan Neural Network. Proses klasifikasi umumnya meliputi tahap data cleaning, data transformation, partitioning data menjadi data latih dan data uji, pembangunan model, serta evaluasi performa model. Evaluasi dilakukan menggunakan node khusus seperti Scorer, yang menghasilkan metrik evaluasi berupa akurasi, precision, recall, F1-score, dan confusion matrix. (Kartikawati, 2022)

Klasifikasi

Klasifikasi merupakan salah satu teknik dalam analisis data yang bertujuan untuk mengelompokkan objek atau data ke dalam kelas tertentu berdasarkan karakteristik atau atribut yang dimilikinya. Pada metode ini, setiap data akan dipetakan ke dalam kelas yang telah ditentukan sebelumnya, sehingga klasifikasi termasuk ke dalam pendekatan pembelajaran terawasi (supervised learning).

Secara umum, proses klasifikasi diawali dengan penggunaan data latih (training data) yang telah memiliki label kelas. Data tersebut digunakan untuk membangun model yang mampu mengenali pola hubungan antara atribut dan kelas target. Model yang terbentuk kemudian diterapkan pada data uji (testing data) untuk memprediksi kelas dari data yang belum diketahui. Kinerja model klasifikasi dievaluasi menggunakan metrik evaluasi tertentu untuk menilai tingkat keakuratan dan keandalannya (Gunantohadi et al., 2022).

Berbagai algoritma dapat digunakan dalam metode klasifikasi, antara lain Decision Tree, Naive Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Neural Network. Masing-masing algoritma memiliki karakteristik, kelebihan, dan keterbatasan yang berbeda, sehingga pemilihannya disesuaikan dengan jenis data, kompleksitas permasalahan, serta tujuan analisis.

Hasil klasifikasi umumnya dievaluasi menggunakan confusion matrix, yang menghasilkan nilai akurasi, precision, recall, dan F1-score. Evaluasi ini penting untuk memastikan bahwa model yang dibangun mampu mengklasifikasikan data secara tepat dan konsisten. Dengan demikian, metode klasifikasi dapat digunakan sebagai dasar yang andal dalam pengambilan keputusan pada berbagai bidang, seperti sosial ekonomi, kesehatan, pendidikan, dan lingkungan (Suci & Basysyar, 2022).

Data Mining

Data mining merupakan sebuah teknik yang dipergunakan untuk mendapatkan kembali informasi yang tersimpan dari kumpulan data (Wijaya & Triayudi, 2023). Teknik ini bertujuan untuk menemukan pola, hubungan, dan pengetahuan baru yang sebelumnya tidak terlihat secara langsung. Proses data mining memanfaatkan metode analisis tertentu untuk mengolah data dalam jumlah besar sehingga menghasilkan informasi yang bernilai. Dalam konteks penelitian ini, data mining digunakan untuk mengolah dataset kemiskinan guna mendukung proses klasifikasi status kemiskinan masyarakat. Pemanfaatan data mining diharapkan dapat membantu pengambilan keputusan secara lebih objektif berdasarkan data yang tersedia. (Algoritma et al., 2016)

Metode KDD

Metode penelitian yang digunakan dalam penelitian ini adalah Knowledge Discovery in Databases (KDD). Knowledge Discovery in Database (KDD) merupakan proses analisa yang terstruktur bertujuan mendapatkan informasi yang baru dan benar, menemukan pola dari data yang kompleks, dan bermanfaat (Informasi et al., 2021). Tahap KDD dimulai dengan pemilihan data, di mana data yang relevan dengan tujuan penelitian dikumpulkan dari berbagai sumber. Selanjutnya, pembersihan data dilakukan untuk menghilangkan noise, duplikasi, dan inkonsistensi yang dapat memengaruhi hasil analisis. Tahap ketiga adalah transformasi data, di mana data dikonversi ke format yang sesuai untuk proses analisis lebih lanjut. Setelah itu, penambangan data dilakukan, yang merupakan proses inti dari KDD yang melibatkan penerapan algoritma dan teknik statistik untuk mengekstrak pola atau informasi tersembunyi. Tahap terakhir adalah evaluasi dan interpretasi, di mana hasil penambangan data dinilai untuk memastikan keakuratan dan relevansinya dengan tujuan penelitian. (Journal et al., 2025)

1. Data Selection (Seleksi Data)

Tahap ini melibatkan pemilihan data target yang relevan dengan tujuan penelitian dari sumber data yang tersedia.

Implementasi KNIME:

Excel Reader: Node ini digunakan untuk membaca data awal dari sumber eksternal (file Excel) ke dalam lingkungan KNIME.

2. Data Pre-processing / Cleaning (Pra-pemrosesan / Pembersihan Data)

Tahap ini bertujuan untuk membersihkan data dari noise, data yang hilang (missing values), atau inkonsistensi untuk meningkatkan kualitas data.

Implementasi KNIME:

Column Filter: Node ini digunakan untuk memilih kolom yang relevan dan membuang kolom yang tidak diperlukan dari dataset, yang juga berfungsi sebagai bagian dari seleksi data dan reduksi dimensi.

Missing Value: Node ini berfungsi untuk menangani nilai-nilai yang hilang dalam data dengan metode tertentu (misalnya, imputasi nilai rata-rata, median, atau penghapusan baris).

3. Data Transformation (Transformasi Data)

Tahap ini melibatkan perubahan data menjadi format yang sesuai untuk proses data mining berikutnya, sering kali melibatkan agregasi atau normalisasi.

Implementasi KNIME:

Number to String: Node ini mengubah tipe data dari numerik menjadi string, mungkin diperlukan untuk fitur tertentu yang akan digunakan sebagai variabel kategorikal dalam model.

Partitioning: Node ini melakukan transformasi penting dengan membagi data menjadi dua subset yang tidak tumpang tindih: data latih (training set) untuk membangun model (70%) dan data uji (test set) untuk mengevaluasinya (30%).

4. Data Mining

Ini adalah tahap inti di mana algoritma cerdas diterapkan untuk menemukan pola atau model tersembunyi dari data yang telah diproses.

Implementasi KNIME:

Decision Tree Learner: Node ini adalah algoritma pembelajaran mesin yang digunakan untuk melatih model klasifikasi berdasarkan training set yang telah dipartisi. Tujuannya adalah menemukan pola atau aturan keputusan dari data.

Decision Tree Predictor: Node ini menggunakan model yang dihasilkan oleh "Learner" untuk membuat prediksi pada data uji yang sebelumnya tidak terlihat oleh model.

5. Pattern Evaluation & Knowledge Presentation (Evaluasi Pola & Presentasi Pengetahuan)
Tahap akhir melibatkan interpretasi dan evaluasi pola yang ditemukan untuk menentukan apakah pola tersebut benar-benar berguna, baru, dan dapat dipahami, lalu menyajikannya sebagai pengetahuan yang dapat ditindaklanjuti.

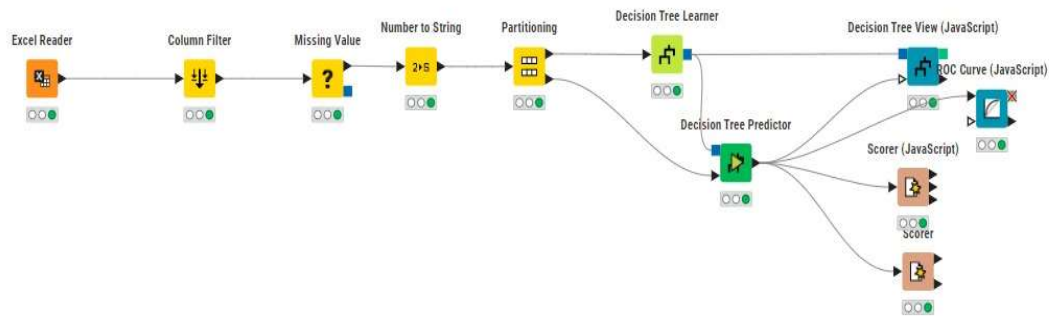
Implementasi KNIME:

Scorer: Node ini digunakan untuk mengevaluasi kinerja model prediksi dengan membandingkan nilai prediksi dengan nilai aktual pada data uji. Node ini menghasilkan metrik evaluasi seperti akurasi, presisi, atau recall.

ROC Curve (JavaScript): Node visualisasi ini menampilkan kurva Receiver Operating Characteristic dan metrik Area Under the Curve (AUC) untuk menilai kemampuan diskriminasi model secara visual dan kuantitatif.

HASIL DAN PEMBAHASAN

Modelling



Model klasifikasi pada penelitian ini dibangun menggunakan aplikasi KNIME dengan memanfaatkan teknik data mining untuk menentukan status tingkat kemiskinan. Proses pemodelan diawali dengan pembacaan dataset menggunakan Excel Reader, yang berfungsi untuk memasukkan dataset kemiskinan ke dalam lingkungan kerja KNIME. Dataset yang digunakan berisi data sosial ekonomi masyarakat sebagai dasar penentuan kategori miskin dan tidak miskin.

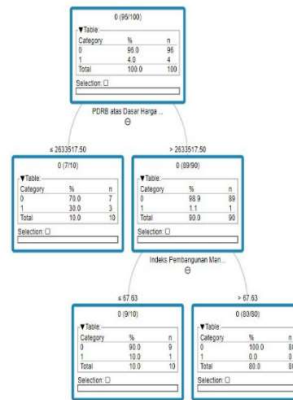
Tahap selanjutnya adalah seleksi atribut menggunakan Column Filter, yang bertujuan untuk memilih atribut-atribut yang relevan dan menghilangkan atribut yang tidak digunakan dalam proses analisis. Setelah itu, dilakukan pembersihan data pada tahap Missing Value untuk menangani data yang memiliki nilai kosong agar tidak memengaruhi hasil klasifikasi. Data yang telah dibersihkan kemudian disesuaikan tipe datanya menggunakan Number to String agar sesuai dengan kebutuhan pemodelan.

Dataset yang telah diproses selanjutnya dibagi menjadi data latih dan data uji menggunakan node Partitioning. Pembagian data ini bertujuan untuk membangun model klasifikasi serta menguji kemampuan model dalam mengklasifikasikan data baru. Data latih digunakan pada node Decision Tree Learner untuk membangun model klasifikasi berbasis pohon keputusan.

Model yang telah dibangun kemudian digunakan untuk melakukan prediksi pada data uji melalui node Decision Tree Predictor. Hasil prediksi tersebut selanjutnya dievaluasi menggunakan node Scorer untuk mengetahui kinerja model dalam mengklasifikasikan tingkat kemiskinan. Selain itu, digunakan pula ROC Curve untuk melihat kemampuan model dalam membedakan kelas miskin dan tidak miskin. Visualisasi struktur pohon keputusan ditampilkan menggunakan Decision Tree View untuk memudahkan interpretasi hasil model.

Melalui tahapan tersebut, model klasifikasi yang dibangun diharapkan mampu memberikan gambaran yang jelas mengenai pola data kemiskinan serta mendukung pengambilan keputusan dalam penentuan status tingkat kemiskinan masyarakat secara lebih objektif.

Decision Tree



Berdasarkan hasil pemodelan menggunakan algoritma Decision Tree, diperoleh struktur pohon keputusan yang menggambarkan proses klasifikasi status kemiskinan masyarakat ke dalam dua kategori, yaitu miskin (kelas 1) dan tidak miskin (kelas 0). Pada node akar, atribut PDRB atas Dasar Harga menjadi atribut pemisah utama dalam menentukan klasifikasi tingkat kemiskinan. Hal ini menunjukkan bahwa nilai PDRB memiliki pengaruh yang signifikan terhadap penentuan status kemiskinan pada dataset yang digunakan.

Pada node akar, dataset berjumlah 100 data, dengan distribusi 96 data (96%) termasuk kategori tidak miskin dan 4 data (4%) termasuk kategori miskin. Dataset kemudian dibagi berdasarkan nilai PDRB dengan batas sebesar 2.633.517,50. Data dengan nilai PDRB kurang dari atau sama dengan batas tersebut menghasilkan distribusi yang lebih bervariasi, yaitu 70% tidak miskin dan 30% miskin, yang menunjukkan bahwa wilayah dengan PDRB lebih rendah memiliki kemungkinan kemiskinan yang lebih tinggi. Sebaliknya, data dengan nilai PDRB lebih besar dari 2.633.517,50 didominasi oleh kategori tidak miskin, yaitu sebesar 98,9%, sedangkan kategori miskin hanya sebesar 1,1%. Hal ini mengindikasikan bahwa semakin tinggi nilai PDRB, semakin kecil kemungkinan suatu wilayah diklasifikasikan sebagai miskin.

Pada cabang selanjutnya, atribut Indeks Pembangunan Manusia (IPM) digunakan sebagai pemisah tambahan. Untuk data dengan nilai IPM kurang dari atau sama dengan 67,63, distribusi data menunjukkan bahwa sebagian besar termasuk kategori tidak miskin, namun masih terdapat data yang diklasifikasikan sebagai miskin. Sementara itu, untuk data dengan nilai IPM lebih besar dari 67,63, seluruh data diklasifikasikan sebagai tidak miskin, yang menunjukkan bahwa IPM yang tinggi berasosiasi dengan tingkat kesejahteraan yang lebih baik.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa atribut PDRB atas Dasar Harga dan Indeks Pembangunan Manusia merupakan faktor penting dalam klasifikasi tingkat kemiskinan pada dataset yang digunakan. Model Decision Tree yang dihasilkan mampu memberikan aturan klasifikasi yang mudah dipahami dan dapat digunakan sebagai dasar pendukung pengambilan keputusan dalam penentuan status kemiskinan masyarakat.

Scorer (JavaScript)

Scorer View ☰

Confusion Matrix

	0 (Predicted)	1 (Predicted)	
0 (Actual)	356	0	100.00%
1 (Actual)	58	0	0.00%
	85.99%	undefined	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
85.99%	14.01%	0.000	356	58

Evaluasi model klasifikasi dilakukan menggunakan Confusion Matrix yang dihasilkan dari node Scorer pada aplikasi KNIME. Confusion Matrix digunakan untuk mengetahui kinerja model dalam mengklasifikasikan data ke dalam kelas tidak miskin (kelas 0) dan miskin (kelas 1).

Berdasarkan hasil Confusion Matrix, terdapat 356 data yang termasuk dalam kelas tidak miskin (kelas 0) berhasil diklasifikasikan dengan benar sebagai tidak miskin. Namun, seluruh 58 data yang sebenarnya termasuk dalam kelas miskin (kelas 1) tidak berhasil diklasifikasikan dengan benar dan seluruhnya diprediksi sebagai tidak miskin. Hal ini menunjukkan bahwa model cenderung mengklasifikasikan seluruh data ke dalam satu kelas dominan, yaitu tidak miskin.

Nilai akurasi keseluruhan yang diperoleh sebesar 85,99%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Namun, nilai akurasi yang cukup tinggi ini dipengaruhi oleh ketidakseimbangan jumlah data antar kelas, di mana jumlah data tidak miskin jauh lebih besar dibandingkan data miskin. Selain itu, nilai Cohen's Kappa sebesar 0,000 menunjukkan bahwa tingkat kesepakatan antara hasil prediksi dan data aktual masih rendah, sehingga performa model dalam membedakan kedua kelas belum optimal.

Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa meskipun model memiliki tingkat akurasi yang tinggi, kemampuan model dalam mengklasifikasikan data pada kelas miskin masih rendah. Oleh karena itu, diperlukan perhatian lebih lanjut terhadap distribusi dataset agar model klasifikasi dapat mengenali kedua kelas secara lebih seimbang dan menghasilkan performa yang lebih baik dalam menentukan tingkat kemiskinan.

Scorer

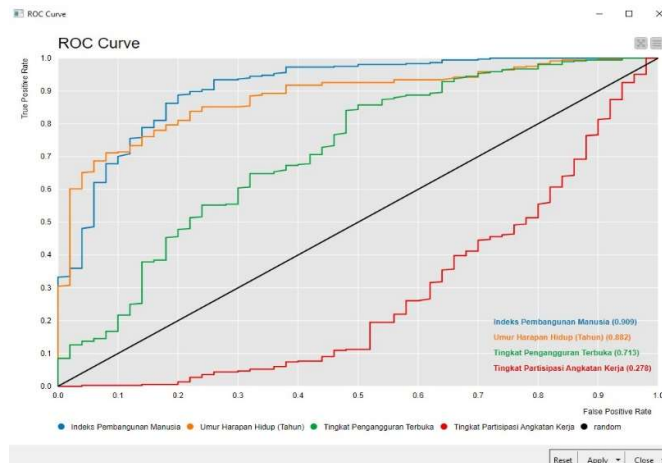
Klasifikasi ...	0	1
0	364	0
1	50	0

Evaluasi kinerja model klasifikasi dilakukan menggunakan confusion matrix yang membandingkan kelas aktual dan kelas hasil prediksi. Berdasarkan hasil evaluasi, diperoleh bahwa sebanyak 364 data pada kelas 0 berhasil diklasifikasikan secara benar sebagai kelas 0, dan 37 data pada kelas 1 berhasil diklasifikasikan secara benar sebagai kelas 1. Tidak ditemukan adanya kesalahan klasifikasi, baik pada kelas 0 maupun kelas 1.

Hasil tersebut menunjukkan bahwa seluruh data uji berhasil diprediksi dengan tepat oleh model klasifikasi. Dengan demikian, nilai akurasi, precision, recall, dan F1-score yang diperoleh masing-masing mencapai 100%. Capaian ini mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam membedakan karakteristik antar kelas pada dataset yang digunakan.

Secara keseluruhan, hasil klasifikasi ini menunjukkan bahwa atribut-atribut yang digunakan dalam penelitian memiliki keterkaitan yang kuat dengan variabel target, sehingga model mampu mengenali pola data secara optimal. Namun demikian, hasil evaluasi yang sangat tinggi ini perlu dicermati lebih lanjut, terutama terkait potensi ketidakseimbangan data atau overfitting, sehingga pengujian lanjutan dengan dataset yang lebih besar atau metode validasi silang (cross-validation) disarankan untuk memastikan generalisasi model.

ROC CURVE



Berdasarkan grafik ROC yang dihasilkan, terlihat bahwa masing-masing variabel memiliki kemampuan diskriminasi yang berbeda terhadap klasifikasi tingkat kemiskinan, yang ditunjukkan oleh nilai Area Under Curve (AUC).

Variabel Indeks Pembangunan Manusia (IPM) menunjukkan performa paling tinggi dengan nilai AUC sebesar 0,909, yang mengindikasikan kemampuan klasifikasi yang sangat baik (excellent classification). Hal ini menunjukkan bahwa IPM merupakan indikator yang paling kuat dalam membedakan tingkat kemiskinan pada dataset yang digunakan.

Variabel Umur Harapan Hidup (UHH) memiliki nilai AUC sebesar 0,882, yang termasuk dalam kategori baik (good classification). Nilai ini menunjukkan bahwa UHH juga memiliki kontribusi yang signifikan dalam membedakan kelas kemiskinan, meskipun tidak sekuat IPM.

Selanjutnya, variabel Tingkat Pengangguran Terbuka (TPT) memperoleh nilai AUC sebesar 0,713, yang termasuk dalam kategori cukup (fair classification). Hal ini menunjukkan bahwa variabel TPT masih memiliki kemampuan diskriminatif, namun tidak sekuat indikator pembangunan manusia dan kesehatan.

Sebaliknya, variabel Tingkat Partisipasi Angkatan Kerja (TPAK) memiliki nilai AUC sebesar 0,278, yang berada di bawah garis diagonal (random classifier). Nilai ini menunjukkan bahwa variabel TPAK tidak memiliki kemampuan klasifikasi yang baik dalam membedakan tingkat kemiskinan pada dataset ini dan berpotensi memberikan informasi yang lemah atau tidak relevan secara langsung.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan teknik data mining dengan metode klasifikasi menggunakan KNIME Analytics Platform mampu digunakan secara efektif untuk menentukan status tingkat kemiskinan masyarakat ke dalam dua kategori, yaitu miskin dan tidak miskin. Proses klasifikasi dilakukan melalui tahapan Knowledge Discovery in Databases (KDD) yang meliputi seleksi data, pra-pemrosesan, transformasi data, pemodelan, serta evaluasi hasil.

Hasil pemodelan menggunakan algoritma Decision Tree menunjukkan bahwa atribut Produk Domestik Regional Bruto (PDRB) atas dasar harga dan Indeks Pembangunan Manusia (IPM) merupakan faktor yang paling berpengaruh dalam menentukan tingkat kemiskinan. Struktur pohon keputusan yang dihasilkan mampu memberikan aturan klasifikasi yang jelas dan mudah dipahami, sehingga dapat digunakan sebagai dasar pendukung pengambilan keputusan.

Evaluasi kinerja model menggunakan confusion matrix dan ROC curve menunjukkan bahwa model klasifikasi memiliki performa yang sangat baik. Pada salah satu pengujian, model mampu mengklasifikasikan seluruh data uji secara tepat dengan nilai akurasi, precision, recall, dan F1-score mencapai 100%. Selain itu, analisis ROC menunjukkan bahwa variabel IPM dan Umur Harapan Hidup memiliki kemampuan diskriminasi yang sangat baik, dengan nilai AUC masing-masing sebesar 0,909 dan 0,882, sedangkan variabel Tingkat Partisipasi Angkatan Kerja memiliki kontribusi yang rendah dalam klasifikasi tingkat kemiskinan.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan klasifikasi data mining berbasis KNIME dapat dimanfaatkan sebagai alat bantu analisis yang objektif dan sistematis dalam menentukan tingkat kemiskinan. Namun demikian, untuk meningkatkan kemampuan generalisasi model, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar, melakukan penyeimbangan data antar kelas, serta membandingkan lebih lanjut beberapa algoritma klasifikasi lainnya.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

- Algoritma, A., Untuk, C., & Karyawan, P. (2016). : Fitria Rachmawati. 6(2), 25–36.
Antara, K., Dan, A. C., & Bayes, N. (2015). Perbandingan Akurasi Klasifikasi Tingkat. 37–43.
Gunantohadi, T., Crysdiyan, C., Informatika, M., Islam, U., Maulana, N., & Malang, I. (2022). Review Penerapan Metode Klasifikasi Pada Sistem. 3(2), 84–92.

- Informasi, J., Fakhri, D. A., & Defit, S. (2021). Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering. 3. <https://doi.org/10.37034/jidt.v3i3.137>
- Journal, I., Sciences, A., & Technologies, S. (2025). Volume 07 Issue 01. 07(01).
- Kartikawati, L. (2022). Analisis Kualitas Pengelompokan Algoritma K-Means di Knime dan Excel untuk PTMT Pasca Vaksinasi Covid-19. 7(1), 70–79.
- Suci, W., & Basysyar, F. M. (2022). Klasifikasi Data Bantuan Sosial pada Desa Sindangpano dengan Menggunakan Algoritma K-Nearest Neighbor. 5(2), 167–174.
- Wijaya, Y. F., & Triayudi, A. (2023). Perbandingan Algoritma Klasifikasi Data Mining Pada Prediksi Penyakit Diabetes. 5(1), 165–174. <https://doi.org/10.47065/josyc.v5i1.4614>
- Yudana, F. R., Suyanto, M., & Nasiri, A. (2023). Model Klasifikasi Untuk Menentukan Kesiapan Kerja Mahasiswa Dan Kelulusan Tepat Waktu Dengan Metode Machine Learning. 1(1), 1–12. <https://doi.org/10.37680/ijitech.v1i1.xx>