

## KLASIFIKASI KEMISKINAN DI INDONESIA DENGAN *DECISION TREE* MENGGUNAKAN *RAPIDMINER*

Oleh:

**Andi Setyawan<sup>1</sup>**

**An-nisa Fitriani<sup>2</sup>**

**Elkin Rilvani<sup>3</sup>**

Universitas Pelita Bangsa

Alamat: Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Sel., Kabupaten Bekasi, Jawa Barat (17530).

Korespondensi Penulis: [andisetyawan194@gmail.com](mailto:andisetyawan194@gmail.com), [annisafitri112501@gmail.com](mailto:annisafitri112501@gmail.com),  
[elkin.rilvani@pelitabangsa.ac.id](mailto:elkin.rilvani@pelitabangsa.ac.id).

***Abstract.** Poverty remains one of the most pressing issues in the context of development, particularly in developing countries such as Indonesia, where disparities in socio-economic conditions are still prevalent. Accurate and data-driven policy-making is essential to effectively address this issue, which necessitates the development of robust classification methods to identify and map poverty status among the population. This study aims to analyze the classification of poverty status using the Decision Tree method implemented through the RapidMiner application. The research utilizes a synthetic dataset comprising 150 data entries that simulate the socio-economic characteristics of individuals in Indonesia. Key variables considered in the dataset include age, level of education, employment status, monthly income, family size, and type of residential location (urban or rural). The classification process follows a supervised learning approach, producing an interpretable decision tree model that can be used to understand the decision-making logic. The results of the analysis demonstrate that the Decision Tree algorithm is capable of classifying poverty status with a high accuracy rate of 93%. Among the features, monthly income and employment status are identified as the most influential factors. The findings of this study are expected to support the formulation of more targeted and effective social assistance programs and poverty alleviation policies.*

---

Received July 03, 2025; Revised July 16, 2025; July 28, 2025

\*Corresponding author: [andisetyawan194@gmail.com](mailto:andisetyawan194@gmail.com)

# KLASIFIKASI KEMISKINAN DI INDONESIA DENGAN *DECISION TREE* MENGGUNAKAN *RAPIDMINER*

**Keywords:** *Decision Tree, Poverty, Classification, Data Mining, RapidMiner.*

**Abstrak.** Kemiskinan merupakan salah satu permasalahan utama yang masih dihadapi dalam proses pembangunan, khususnya di negara-negara berkembang seperti Indonesia. Ketimpangan sosial dan ekonomi yang masih tinggi menuntut adanya kebijakan yang tepat sasaran dan berbasis pada data yang akurat. Oleh karena itu, dibutuhkan metode klasifikasi yang mampu memetakan status kemiskinan masyarakat secara efektif. Penelitian ini bertujuan untuk menganalisis klasifikasi status kemiskinan menggunakan metode *Decision Tree* yang diimplementasikan melalui aplikasi *RapidMiner*. Data yang digunakan merupakan data sintesis sebanyak 150 entri yang menggambarkan kondisi sosial ekonomi penduduk Indonesia. Variabel-variabel yang digunakan dalam analisis meliputi umur, tingkat pendidikan, status pekerjaan, pendapatan bulanan, jumlah anggota keluarga, serta tipe lokasi tempat tinggal (perkotaan atau pedesaan). Proses klasifikasi dilakukan menggunakan pendekatan pembelajaran terawasi (*supervised learning*), yang menghasilkan model pohon keputusan yang mudah dipahami dan diinterpretasikan. Hasil penelitian menunjukkan bahwa metode *Decision Tree* mampu mengklasifikasikan status kemiskinan dengan tingkat akurasi mencapai 93%. Dari hasil analisis, diketahui bahwa pendapatan bulanan dan status pekerjaan merupakan variabel yang paling berpengaruh dalam menentukan status kemiskinan. Temuan ini diharapkan dapat menjadi landasan dalam perumusan kebijakan intervensi sosial yang lebih tepat sasaran dan efektif.

**Kata Kunci:** *Decision Tree, Kemiskinan, Klasifikasi, Data Mining, RapidMiner.*

## LATAR BELAKANG

Kemiskinan adalah salah satu tantangan terbesar dalam pembangunan sosial dan ekonomi yang berkelanjutan. Di Indonesia, meskipun terjadi penurunan angka kemiskinan secara nasional, namun distribusi kemiskinan yang tidak merata masih menjadi persoalan serius terutama di daerah terpencil. Badan Pusat Statistik (BPS) mencatat bahwa pada Maret 2023, persentase penduduk miskin di Indonesia mencapai 9,36% atau sekitar 25,9 juta jiwa. Oleh karena itu, upaya identifikasi wilayah rawan kemiskinan menjadi sangat penting untuk memastikan distribusi program bantuan sosial yang lebih tepat sasaran.

Dalam beberapa tahun terakhir, pendekatan berbasis teknologi informasi seperti data mining semakin banyak digunakan dalam upaya pemetaan kemiskinan. Data mining merupakan proses mengekstraksi pola atau informasi tersembunyi dari kumpulan data besar untuk mendukung proses pengambilan keputusan. Salah satu teknik data mining yang sering digunakan dalam klasifikasi adalah algoritma *Decision Tree*. Algoritma ini populer karena hasilnya dapat divisualisasikan dalam bentuk pohon keputusan yang mudah dipahami, bahkan oleh pengambil kebijakan non-teknis. (Jollyta, N. Irawan, dan R. Dewi, 2023)

Berdasarkan hal tersebut, penelitian ini dilakukan untuk menerapkan metode *Decision Tree* dalam mengklasifikasikan status kemiskinan masyarakat Indonesia dengan menggunakan dataset sintetis. Penelitian ini bertujuan untuk mengetahui variabel mana yang paling memengaruhi status kemiskinan serta mengevaluasi performa algoritma *Decision Tree* dalam proses klasifikasi.

## **KAJIAN TEORITIS**

### **Konsep Kemiskinan dan Pendekatan Data Mining**

Kemiskinan adalah situasi di mana seseorang atau sekelompok orang tidak mampu memenuhi kebutuhan dasar dengan layak. Di Indonesia, masalah kemiskinan menjadi fokus utama pembangunan yang memerlukan intervensi yang tepat dan berbasis data. Salah satu metode modern untuk memetakan dan mengklasifikasikan kemiskinan adalah dengan menggunakan data mining, yang merupakan proses analisis data dalam jumlah besar untuk mengidentifikasi pola-pola tersembunyi yang bermanfaat dalam pengambilan keputusan (Witten et al. , 2016).

Klasifikasi adalah salah satu teknik utama dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan beberapa atribut input. Dalam konteks penelitian ini, klasifikasi digunakan untuk menentukan status kemiskinan individu atau keluarga berdasarkan atribut sosial ekonomi seperti usia, pendidikan, pekerjaan, pendapatan, dan lain-lain.

### **Algoritma *Decision Tree***

Algoritma *decision tree* berfungsi dengan membagi data berdasarkan atribut yang memberikan informasi paling signifikan (information gain atau gain ratio). Proses ini

# KLASIFIKASI KEMISKINAN DI INDONESIA DENGAN *DECISION TREE* MENGGUNAKAN *RAPIDMINER*

menghasilkan model pohon keputusan di mana setiap simpul mewakili atribut dan setiap daun menggambarkan prediksi kelas akhir. Kelebihan dari algoritma ini adalah kemudahan dalam interpretasi dan visualisasi yang sangat berguna untuk membantu pengambilan keputusan oleh pihak-pihak yang tidak memiliki latar belakang teknis.

Menurut Quinlan (1993), algoritma C4. 5 (pengembangan dari ID3) yang memanfaatkan gain ratio sangat efisien dalam membangun model klasifikasi dari data yang memiliki atribut campuran dan mampu menangani nilai yang hilang. Model ini sangat tepat diterapkan pada permasalahan sosial seperti klasifikasi kemiskinan.

## Review Penelitian Sebelumnya

Di bawah ini adalah beberapa penelitian sebelumnya yang relevan dan menjadi sumber referensi penting dalam mengembangkan sistem klasifikasi status kemiskinan menggunakan *decision tree* dan alat bantu seperti *RapidMiner*:

1. Jollyta, Irawan, dan Dewi (2023)

Penelitian ini berjudul "Penerapan *Decision Tree* dalam Pengklasifikasian Status Kemiskinan Berdasarkan Data Sosial Ekonomi". Penelitian ini memanfaatkan algoritma *decision tree* untuk mengelompokkan masyarakat ke dalam kategori miskin dan tidak miskin berdasarkan variabel seperti pengeluaran rumah tangga, pekerjaan, dan pendidikan. Hasil penelitian menunjukkan bahwa *decision tree* memiliki tingkat akurasi yang tinggi dan menghasilkan model klasifikasi yang mudah dimengerti oleh pembuat kebijakan. Penelitian ini menekankan pentingnya interpretabilitas model dalam konteks kebijakan sosial.

2. Qisthiano, Hidayat, dan Saputro (2023)

Dalam studi yang berjudul "Analisis Klasifikasi Status Sosial Ekonomi Menggunakan Metode *Decision Tree C4. 5*", peneliti membandingkan kinerja algoritma *decision tree C4. 5* dengan algoritma lainnya dalam klasifikasi status sosial ekonomi masyarakat di Jawa Tengah. Dataset yang digunakan mencakup atribut seperti jenis pekerjaan, pengeluaran bulanan, dan kepemilikan aset. Hasil analisis menunjukkan bahwa C4. 5 dapat memberikan akurasi lebih dari 90% dan menghasilkan

pohon keputusan yang logis, sehingga cocok digunakan pada data pemerintah daerah.

3. Ardilla, Kurniawan, dan Latifah (2021)

Penelitian ini berjudul "Model Klasifikasi Kemiskinan Berbasis Data Mining untuk Menunjang Kebijakan Pemerintah". Penelitian ini mengembangkan model klasifikasi kemiskinan menggunakan *decision tree* dengan data dari Dinas Sosial. Peneliti menggunakan atribut seperti kepemilikan rumah, jumlah tanggungan dalam keluarga, dan jenis pekerjaan kepala keluarga. Hasilnya menunjukkan bahwa atribut pekerjaan dan pendapatan bulanan merupakan faktor yang paling berpengaruh dalam model klasifikasi. Penelitian ini juga menunjukkan bahwa klasifikasi otomatis dapat membantu penentuan penerima bantuan sosial secara lebih objektif dan akurat.

4. Sari dan Maulidina (2022)

Dalam studi yang berjudul "Penerapan Data Mining untuk Prediksi Kemiskinan Menggunakan Algoritma *Decision Tree* dan *Naïve Bayes*", Sari serta Maulidina melakukan perbandingan antara kedua algoritma tersebut dalam mengklasifikasikan kondisi ekonomi di Provinsi Jawa Timur. Hasil penelitian menunjukkan bahwa meskipun *Naïve Bayes* lebih cepat, algoritma *decision tree* berhasil memberikan akurasi dan presisi yang lebih tinggi dalam mengidentifikasi kategori kemiskinan. Penelitian ini juga menunjukkan bahwa *decision tree* memiliki keunggulan dalam menghasilkan model yang bisa dijelaskan (*explainable*) kepada para pembuat kebijakan.

5. Fitriana dan Permana (2023)

Tujuan dari penelitian ini adalah untuk mengklasifikasikan status ekonomi dari keluarga penerima manfaat di Kabupaten Sleman dengan menggunakan *decision tree* yang dioperasikan melalui *RapidMiner*. Penelitian ini melibatkan 200 data rumah tangga serta atribut-atribut seperti pendapatan, kepemilikan kendaraan, pendidikan kepala keluarga, dan status tempat tinggal, dan berhasil menghasilkan model dengan akurasi mencapai 91,7%. Penelitian ini menekankan bahwa *RapidMiner*

# **KLASIFIKASI KEMISKINAN DI INDONESIA DENGAN *DECISION TREE* MENGGUNAKAN *RAPIDMINER***

sangat berperan dalam visualisasi serta proses pengolahan data yang terstruktur tanpa membutuhkan kode yang rumit.

## ***RapidMiner* sebagai Alat Analisis**

*RapidMiner* adalah platform analitik yang berbasis antarmuka grafis pengguna (GUI), yang memungkinkan pemakai untuk membangun proses analisis data secara visual. Perangkat lunak ini mendukung berbagai teknik pembelajaran mesin, termasuk klasifikasi, regresi, dan pengelompokan. Dalam penelitian ini, *RapidMiner* dimanfaatkan untuk menerapkan algoritma *decision tree* pada data kemiskinan yang dihasilkan secara sintesis. Pengguna dapat mengatur parameter seperti metode pemilihan atribut (rasio keuntungan, informasi yang diperoleh), kedalaman maksimal pohon, dan ukuran minimum cacah agar bisa mengatur kompleksitas dari model yang dibuat.

Penggunaan *RapidMiner* terbukti sangat efisien dalam penelitian baik di ranah akademis maupun profesional karena dapat menghasilkan model dalam waktu singkat, serta dilengkapi dengan fitur evaluasi yang komprehensif seperti *confusion matrix*, akurasi, presisi, dan *recall*. Kapasitasnya untuk melakukan pra-pemrosesan data dan pemodelan dalam satu alur kerja terintegrasi menjadikan alat ini sangat sesuai untuk penelitian tentang klasifikasi kemiskinan.

## **METODE PENELITIAN**

Metodologi penelitian ini dirancang untuk memberikan penjelasan menyeluruh dan sistematis mengenai langkah-langkah dalam penerapan algoritma *Decision Tree* untuk klasifikasi status kemiskinan menggunakan aplikasi *RapidMiner*. Penelitian dilakukan secara kuantitatif dengan pendekatan eksperimen melalui tahapan-tahapan yang dapat direplikasi, mulai dari persiapan data hingga evaluasi performa model klasifikasi.

### **Menyiapkan Dataset**

Langkah pertama yang dilakukan dalam penelitian ini adalah menyiapkan dataset yang akan digunakan sebagai sumber data pelatihan dan pengujian. Dataset yang digunakan merupakan data sintesis yang disusun oleh peneliti dan mencerminkan atribut

sosial ekonomi masyarakat Indonesia, dengan total 150 entri data. Setiap entri memiliki atribut sebagai berikut:

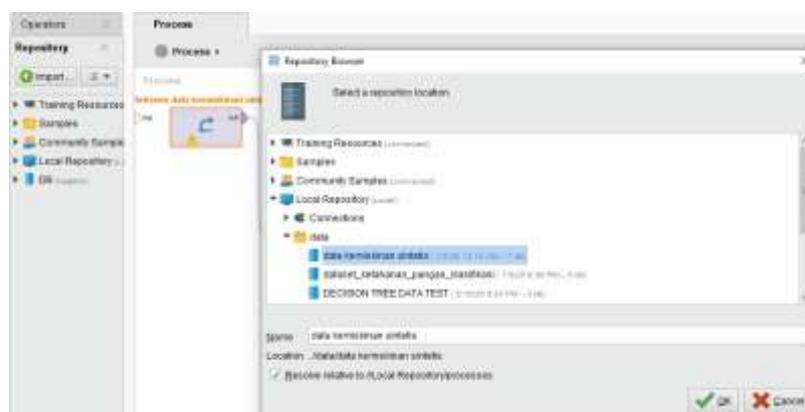
- a. Umur (*Age*)
- b. Tingkat Pendidikan (*Education Level*)
- c. Status Pekerjaan (*Employment Status*)
- d. Pendapatan Bulanan (*Monthly Income*)
- e. Jumlah Anggota Keluarga (*Family Size*)
- f. Jenis Lokasi (*Location Type*)
- g. Label: Status Kemiskinan (*Poverty Status*)

Selanjutnya, data ini dimasukkan ke dalam *repository* pada aplikasi *RapidMiner* untuk nantinya dapat dilakukan pengolahan data dan analisa menggunakan algoritma *Decision tree*.

### Mengimpor Dataset

Dataset yang telah disiapkan terlebih dahulu disimpan dalam repository *RapidMiner*. Untuk memanggil dataset tersebut ke dalam proses kerja (*process panel*), digunakan operator *Retrieve*. Operator ini memungkinkan pengguna mengambil dataset yang telah disimpan sebelumnya dalam format internal *RapidMiner*, sehingga lebih efisien dibanding impor eksternal. Dataset yang di-*retrieve* ini kemudian digunakan sebagai input utama untuk seluruh proses klasifikasi selanjutnya, mulai dari pengaturan role, pembagian data, hingga pembuatan model klasifikasi.

**Gambar 1. Retrieve data dari repository**



Sumber : Data olahan peneliti menggunakan *RapidMiner*.

# KLASIFIKASI KEMISKINAN DI INDONESIA DENGAN *DECISION TREE* MENGGUNAKAN *RAPIDMINER*

## Menentukan Label (*Set Role*)

Setelah data berhasil dimasukkan ke *RapidMiner*, langkah berikutnya adalah mengatur atribut target yang akan diprediksi, yaitu "*Poverty Status*". Ini dilakukan menggunakan operator *Set Role*, dengan menetapkan atribut tersebut sebagai label, sementara atribut lainnya tetap sebagai attribute default.

**Gambar 2. Set Role Atribut Data**



Sumber : Data olahan peneliti menggunakan *RapidMiner*

## Membagi Data (*Split Data*)

Proses dilanjutkan dengan membagi dataset menjadi dua bagian, yaitu data pelatihan (*training*) dan data pengujian (*testing*). Operator *Split Data* digunakan dengan rasio pembagian 80% untuk pelatihan dan 20% untuk pengujian. Hal ini penting untuk menghindari *overfitting* dan untuk menguji kemampuan generalisasi model.

**Gambar 3. Konfigurasi parameter partisi dataset**



Sumber : Data olahan peneliti menggunakan *RapidMiner*

## Membangun Model *Decision Tree*

Pembangunan model klasifikasi dilakukan menggunakan operator *Decision Tree* dalam aplikasi *RapidMiner*. Operator ini menyediakan sejumlah parameter pengaturan yang dapat disesuaikan, dan dalam penelitian ini digunakan pengaturan default dari *RapidMiner*. Meskipun *RapidMiner* tidak menyebutkan secara eksplisit nama algoritma yang digunakan (seperti ID3, C4.5, atau CART), namun konfigurasi *default*-nya menggunakan kriteria pemilihan atribut berupa gain ratio, yang umum digunakan dalam algoritma C4.5.

**Gambar 4. Split dataset dan data test**

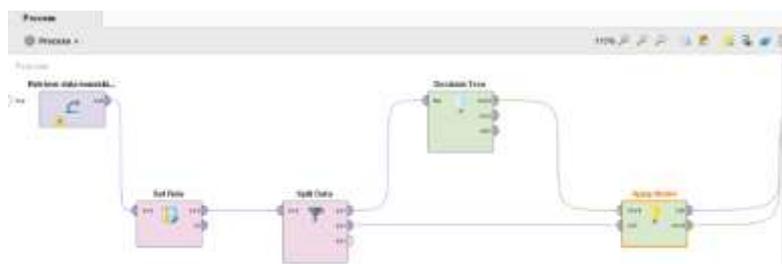


Sumber : Data olahan peneliti menggunakan *RapidMiner*

## Menerapkan Model (*Apply Model*)

Model yang telah dibentuk kemudian digunakan untuk memprediksi status kemiskinan pada data pengujian. Operator *Apply Model* menggabungkan model hasil pelatihan dengan dataset pengujian. Output-nya adalah data prediksi terhadap label target.

**Gambar 5. Penerapan *Apply Model***



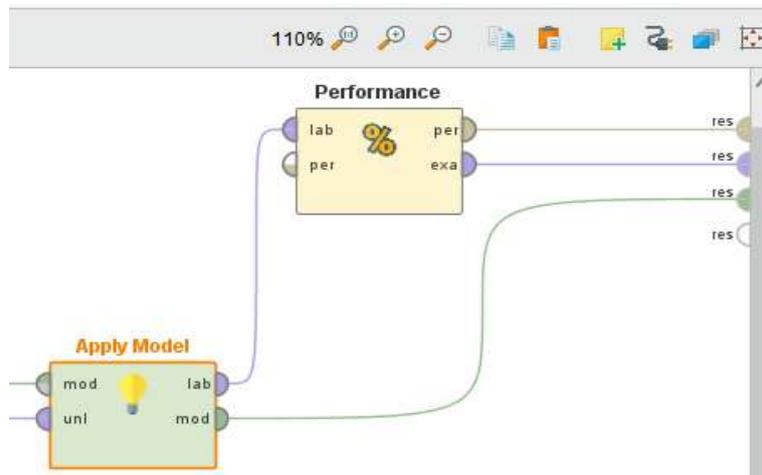
Sumber : Data olahan peneliti menggunakan *RapidMiner*

# KLASIFIKASI KEMISKINAN DI INDONESIA DENGAN *DECISION TREE* MENGGUNAKAN *RAPIDMINER*

## Evaluasi Performa Model

Tahap akhir adalah mengevaluasi akurasi hasil klasifikasi menggunakan operator Performance (*Classification*). Evaluasi dilakukan terhadap hasil prediksi untuk mengukur akurasi, *precision*, *recall*, dan menghasilkan *confusion matrix*. Hasil evaluasi ini digunakan sebagai dasar analisis performa dari algoritma *Decision Tree* dalam klasifikasi status kemiskinan.

**Gambar 6. Penambahan Operator Performace**



Sumber : Data olahan peneliti menggunakan RapidMiner

## HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan model klasifikasi berbasis algoritma *Decision Tree* yang dibangun dan diuji menggunakan aplikasi RapidMiner. Berdasarkan hasil evaluasi model dengan menggunakan operator Performance (*Classification*), diperoleh nilai akurasi sebesar 93.33%. Ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengklasifikasikan status kemiskinan berdasarkan atribut sosial ekonomi yang digunakan. Hasil evaluasi performa model dapat dilihat pada *confusion matrix* berikut:

**Gambar 7. Hasil test performance menggunakan *Decision Tree***

accuracy: 93.33%

	true Vulnerable	true Not Poor	true Poor	class precision
pred. Vulnerable	16	1	0	94.12%
pred. Not Poor	0	6	0	100.00%
pred. Poor	1	0	6	85.71%
class recall	94.12%	85.71%	100.00%	

Sumber : Data olahan peneliti menggunakan *RapidMiner*.

## KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* dapat digunakan secara efektif untuk mengklasifikasikan status kemiskinan masyarakat berdasarkan sejumlah atribut sosial ekonomi. Model klasifikasi yang dibangun menggunakan aplikasi *RapidMiner* mampu mencapai akurasi sebesar 93.33%, dengan tingkat presisi dan *recall* yang tinggi pada masing-masing kelas. Atribut seperti pendapatan bulanan, jumlah anggota keluarga, dan status pekerjaan menjadi faktor dominan dalam pengambilan keputusan. Struktur pohon keputusan yang dihasilkan bersifat eksplisit dan mudah dipahami, sehingga dapat dimanfaatkan dalam mendukung pengambilan kebijakan sosial yang berbasis data.

Meskipun penelitian ini menggunakan data sintetis, metode yang diterapkan terbukti mampu menghasilkan klasifikasi yang akurat dan logis. Oleh karena itu, pendekatan ini dapat dijadikan acuan awal untuk pengembangan sistem klasifikasi serupa dengan data nyata. Penelitian lanjutan disarankan untuk menguji performa algoritma *Decision Tree* terhadap data riil dari lembaga resmi, serta melakukan perbandingan dengan metode klasifikasi lainnya agar diperoleh model yang lebih optimal dan dapat diandalkan dalam pengambilan keputusan yang kompleks.

## DAFTAR REFERENSI

- Ardilla, R. F., Kurniawan, S., & Latifah, T. (2021). Model klasifikasi kemiskinan berbasis *data mining* untuk menunjang kebijakan pemerintah. In *Prosiding Seminar Nasional Teknologi Informasi* (pp. 123–128).
- Badan Pusat Statistik. (2023, Juli 17). *Persentase penduduk miskin Maret 2023 turun menjadi 9,36%*. <https://www.bps.go.id/pressrelease/2023/07/17/1882/persentase-penduduk-miskin-maret-2023.html>
- Fitriana, A., & Permana, F. (2023). Klasifikasi status ekonomi keluarga penerima manfaat menggunakan *decision tree* di *RapidMiner*. *Jurnal Sistem Informasi dan Komputerisasi*, 8(2), 112–120.
- Jollyta, I., Irawan, N., & Dewi, R. (2023). Penerapan *decision tree* dalam pengklasifikasian status kemiskinan berdasarkan data sosial ekonomi. *Jurnal Teknologi dan Sistem Komputer*, 11, 67–73.

## **KLASIFIKASI KEMISKINAN DI INDONESIA DENGAN *DECISION TREE* MENGGUNAKAN *RAPIDMINER***

- Qisthiano, Q., Hidayat, H., & Saputro, Y. (2023). Analisis klasifikasi status sosial ekonomi menggunakan metode *decision tree C4.5*. *Jurnal Penelitian Sistem Informasi (JPenSi)*, 6, 45–52.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
- RapidMiner Documentation. (2024). *Decision tree operator*. [https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/decision\\_tree.html](https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/decision_tree.html)
- RapidMiner.(2025). *RapidMiner Studio educational license*. <https://rapidminer.com/educational-program/>
- Sari, R. R., & Maulidina, N. (2022). Penerapan *data mining* untuk prediksi kemiskinan menggunakan algoritma *decision tree* dan *Naïve Bayes*. *Jurnal Ilmiah Teknologi dan Informasi*, 9(1), 34–40.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed., pp. 57–96). Elsevier.