



## Klasifikasi Tingkat Kemiskinan di Indonesia Menggunakan *Naive Bayes* dengan *RapidMiner*

Ahmad Rizki Sya'bani<sup>1\*</sup>, Wahyu Nur Hidayat<sup>2</sup>, Mikhael Valliano Benjamin<sup>3</sup>,  
Risca Lusiana Pratiwi<sup>4</sup>, Euis Widanegsih<sup>5</sup>

<sup>1,2,3,5</sup> Teknologi Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

<sup>4</sup> Sistem informasi, Fakultas Teknologi informasi, Universitas Nusa mandiri, Indonesia

\*Penulis Korespondensi: [ahmadrizkisyabani85@gmail.com](mailto:ahmadrizkisyabani85@gmail.com)

**Abstract.** Poverty is a multidimensional issue that has a significant impact on social and economic development in Indonesia. Accurate analysis of poverty levels is essential to support government policies in distributing aid and planning targeted development programs. This study aims to classify poverty levels in Indonesia using the Naive Bayes algorithm based on machine learning, assisted by the RapidMiner Studio software. The dataset consists of 155 entries with 12 key attributes reflecting social and economic indicators, such as household expenditure, education level, unemployment rate, and the Human Development Index (HDI). The research follows the CRISP-DM methodology, which includes the stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The modeling results show that the Naive Bayes algorithm achieves an accuracy of 94.19%, with high precision and recall values, indicating consistent model performance in classifying poor and non-poor categories. These findings suggest that the Naive Bayes-based machine learning approach can serve as an effective analytical tool to understand poverty patterns in Indonesia. The implementation of this model is expected to assist the government in making data-driven decisions to improve the effectiveness of poverty alleviation programs.

**Keywords:** Data Classification; Machine Learning; Naive Bayes; Poverty; RapidMiner.

**Abstrak.** Kemiskinan merupakan permasalahan multidimensi yang berdampak signifikan terhadap pembangunan sosial dan ekonomi di Indonesia. Analisis tingkat kemiskinan secara akurat diperlukan untuk mendukung kebijakan pemerintah dalam penyaluran bantuan dan perencanaan pembangunan yang tepat sasaran. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kemiskinan di Indonesia menggunakan algoritma Naive Bayes berbasis machine learning dengan bantuan perangkat lunak RapidMiner Studio. Dataset yang digunakan terdiri atas 155 entri dengan 12 atribut utama yang mencerminkan indikator sosial dan ekonomi, seperti pengeluaran rumah tangga, tingkat pendidikan, pengangguran, dan indeks pembangunan manusia (IPM). Penelitian dilakukan melalui tahapan CRISP-DM yang meliputi business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Hasil pemodelan menunjukkan bahwa algoritma Naive Bayes mampu menghasilkan akurasi sebesar 94,19% dengan nilai precision dan recall yang tinggi, menandakan performa model yang konsisten dalam mengklasifikasikan kategori miskin dan tidak miskin. Temuan ini menunjukkan bahwa pendekatan machine learning berbasis Naive Bayes dapat menjadi alat bantu analitis yang efektif untuk memahami pola kemiskinan di Indonesia. Implementasi model ini diharapkan dapat membantu pemerintah dalam pengambilan keputusan berbasis data guna meningkatkan efektivitas program penanggulangan kemiskinan.

**Kata kunci:** Klasifikasi Data; Klasifikasi Kemiskinan; Machine Learning; Naive Bayes; RapidMiner.

### 1. LATAR BELAKANG

Kemiskinan merupakan permasalahan sosial-ekonomi yang bersifat multidimensi dan masih menjadi tantangan utama dalam pembangunan nasional di Indonesia. Menurut Badan Pusat Statistik (BPS), kemiskinan adalah kondisi ketidakmampuan individu atau rumah tangga dalam memenuhi kebutuhan dasar seperti pangan, pendidikan, kesehatan, dan perumahan layak. Fenomena ini tidak hanya berkaitan dengan rendahnya pendapatan, tetapi juga dengan ketimpangan sosial, rendahnya kualitas sumber daya manusia, serta terbatasnya kesempatan kerja dan akses terhadap layanan publik (Putri et al., 2024). Oleh karena itu, pemahaman yang

komprehensif terhadap penyebab kemiskinan sangat dibutuhkan untuk mendukung kebijakan yang tepat sasaran.

Berbagai penelitian di Indonesia menunjukkan bahwa kemiskinan tidak dapat diatasi hanya dengan pendekatan ekonomi semata. Kurniawan (2020) menegaskan bahwa kemiskinan di Indonesia berkaitan erat dengan dimensi sosial dan budaya yang turut memperkuat ketimpangan ekonomi antarwilayah. Sementara itu, Kshirsagar & Wells, (2017) menyebutkan bahwa tingkat pendidikan, pertumbuhan ekonomi, dan pengangguran merupakan faktor yang paling memengaruhi peningkatan jumlah penduduk miskin. Penelitian Wulansari *et al.* (2023) juga menemukan bahwa ketimpangan distribusi sumber daya dan lemahnya produktivitas menjadi penyebab utama kemiskinan struktural di berbagai daerah di Indonesia.

Meskipun pemerintah telah menjalankan berbagai program seperti Program Keluarga Harapan (PKH), Bantuan Langsung Tunai (BLT), dan Dana Desa, efektivitasnya belum optimal. Lembaga Kita (2024) dalam penelitiannya menyebutkan bahwa salah satu kendala utama penyaluran bantuan sosial adalah belum adanya sistem klasifikasi data penerima yang berbasis analisis ilmiah dan teknologi data. Oleh karena itu, dibutuhkan pendekatan baru berbasis analisis data dan kecerdasan buatan (*artificial intelligence*) untuk menghasilkan model klasifikasi kemiskinan yang lebih akurat dan efisien.

Kemajuan teknologi informasi memberikan peluang besar dalam penerapan *data mining* dan *machine learning* untuk analisis kemiskinan. Pendekatan ini memungkinkan peneliti menemukan pola tersembunyi dalam data sosial-ekonomi yang besar dan kompleks. Menurut Imanda & Nurdin (2024), penerapan algoritma *machine learning* seperti *Naive Bayes* terbukti efektif dalam memprediksi kategori sosial-ekonomi masyarakat karena kemampuannya mengolah data heterogen dengan asumsi sederhana. *Naive Bayes* bekerja berdasarkan Teorema Bayes dengan menghitung probabilitas suatu kelas berdasarkan atribut yang diamati. Meskipun sederhana, metode ini mampu menghasilkan tingkat akurasi yang tinggi pada berbagai penelitian sosial-ekonomi di Indonesia (Mansur, 2024).

Secara global, Kshirsagar & Wells (2017) membuktikan bahwa algoritma *Naive Bayes* dapat digunakan untuk mengklasifikasikan tingkat kemiskinan rumah tangga di negara berkembang dengan data yang terbatas. Pendekatan serupa juga digunakan oleh Olearo *et al.* (2023) untuk memprediksi kemiskinan multidimensi melalui integrasi variabel ekonomi, pendidikan, dan kesehatan, yang menghasilkan model dengan tingkat keandalan tinggi. Sementara di Indonesia, Maharani *et al.* (2024) menekankan bahwa kemiskinan berdampak langsung pada rendahnya kualitas pendidikan anak, sehingga pengentasan kemiskinan perlu didukung analisis berbasis data agar kebijakan yang dihasilkan dapat menyentuh akar permasalahan.

Dalam konteks tersebut, pemanfaatan perangkat lunak RapidMiner menjadi penting karena mampu mengintegrasikan berbagai tahapan *machine learning* secara visual melalui model CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Proses ini meliputi enam tahap utama, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Dengan bantuan RapidMiner, penelitian dapat dilakukan secara sistematis tanpa memerlukan kemampuan pemrograman tingkat lanjut.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk mengklasifikasikan tingkat kemiskinan di Indonesia menggunakan algoritma *Naive Bayes* yang diimplementasikan melalui RapidMiner Studio. Penelitian ini diharapkan dapat menghasilkan model klasifikasi yang akurat dan efisien sebagai dasar pengambilan keputusan dalam kebijakan sosial, khususnya untuk menyalurkan bantuan secara tepat sasaran. Selain itu, hasil penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan metode *machine learning* di bidang sosial-ekonomi dan mendukung penerapan kebijakan berbasis data (*data-driven policy*) di Indonesia.

## 2. TINJAUAN TEORITIS

### Konsep Kemiskinan

Kemiskinan secara umum dipahami sebagai kondisi ketika individu atau rumah tangga tidak mampu memenuhi kebutuhan dasar seperti pangan, pendidikan, kesehatan, dan tempat tinggal yang layak (BPS, 2024). Menurut teori ekonomi klasik, kemiskinan muncul akibat rendahnya produktivitas tenaga kerja dan keterbatasan akses terhadap sumber daya produksi. Sementara menurut teori struktural, kemiskinan bersumber dari ketimpangan distribusi kekuasaan dan kesempatan ekonomi antar kelompok masyarakat (Todaro & Smith, 2020).

Dalam konteks multidimensi, kemiskinan tidak hanya dilihat dari sisi ekonomi, tetapi juga mencakup aspek sosial, pendidikan, dan kesehatan sebagaimana dijelaskan oleh Amartya Sen (1999) melalui *Capability Approach*, yang menekankan pada kemampuan seseorang untuk berfungsi secara efektif dalam kehidupan sosial dan ekonomi. Oleh karena itu, indikator kemiskinan yang digunakan dalam penelitian ini mencakup berbagai dimensi sosial-ekonomi seperti pengeluaran rumah tangga, pendidikan, pengangguran, dan indeks pembangunan manusia (IPM).

## Konsep Klasifikasi Data

Klasifikasi merupakan salah satu teknik utama dalam *data mining* yang bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan atribut yang dimilikinya (Han et al., 2012). Dalam konteks penelitian ini, klasifikasi digunakan untuk menentukan apakah suatu wilayah atau rumah tangga termasuk dalam kategori miskin atau tidak miskin.

Proses klasifikasi dilakukan melalui pendekatan *supervised learning*, di mana sistem belajar dari data latih (training data) yang sudah diketahui labelnya. Model kemudian diuji menggunakan data uji (testing data) untuk mengukur tingkat akurasi dan generalisasi (Tan et al., 2016). Teknik ini sangat berguna untuk analisis sosial-ekonomi karena dapat mengungkap pola tersembunyi dalam data yang kompleks dan beragam.

## Algoritma Naïve Baiyes

Naive Bayes merupakan salah satu algoritma *supervised learning* berbasis probabilistik yang didasarkan pada Teorema Bayes. Algoritma ini menghitung peluang suatu data masuk ke dalam kelas tertentu berdasarkan nilai atribut yang diamati dengan asumsi independensi antar atribut (Han et al., 2012).

Rumus umum Teorema Bayes adalah:

$$P(C|X)=P(X|C)P(C)P(X)$$

Di mana:

$P(C|X)$  adalah probabilitas suatu kelas CC diberikan data XX,

$P(X|C)$  adalah probabilitas data XX muncul jika diketahui kelas CC,

$P(C)$  adalah probabilitas awal kelas CC,

$P(X)$  adalah probabilitas total dari data XX.

Meskipun sederhana, Naive Bayes memiliki performa yang tinggi dan efisien, terutama untuk dataset berukuran kecil hingga menengah. Penelitian terdahulu menunjukkan bahwa algoritma ini cocok untuk analisis kemiskinan karena kemampuannya mengolah variabel sosial-ekonomi dengan akurasi di atas 90% (Imanda & Nurdin, 2024; Chairani et al., 2021).

## Model CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan kerangka kerja standar dalam pelaksanaan penelitian berbasis data mining. Model ini mencakup enam tahapan utama, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment (Chapman et al., 2000).

Model CRISP-DM membantu peneliti dalam memastikan setiap tahap dilakukan secara sistematis, mulai dari identifikasi masalah hingga penerapan model. Dalam konteks penelitian ini, RapidMiner digunakan untuk menerapkan seluruh tahapan CRISP-DM secara visual dan terintegrasi.

### 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk mengklasifikasikan tingkat kemiskinan di Indonesia berdasarkan data sosial dan ekonomi. Model penelitian disusun berdasarkan kerangka CRISP-DM (Cross Industry Standard Process for Data Mining) yang terdiri atas enam tahap utama, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment (Ahmad et al., 2023). Tahapan tersebut digambarkan secara visual pada Diagram 1, yang menunjukkan alur penelitian dari proses identifikasi masalah hingga penerapan model klasifikasi kemiskinan.



**Gambar 1.** Tahapan Penelitian Menggunakan Model *CRISP-DM*.

Dataset yang digunakan dalam penelitian ini terdiri dari 155 entri data dengan 12 atribut yang menggambarkan kondisi sosial dan ekonomi wilayah di Indonesia, seperti rata-rata pengeluaran rumah tangga, tingkat pendidikan, tingkat pengangguran, dan indeks pembangunan manusia (IPM). Sebelum dilakukan pemodelan, data melalui tahap pra-pemrosesan (*preprocessing*) untuk memastikan kualitas dan konsistensi, meliputi penghapusan nilai kosong, normalisasi data numerik, serta pemilihan atribut yang relevan terhadap variabel target (Mulyani & Pajri, 2024). Dataset diimpor ke RapidMiner menggunakan operator Read CSV seperti terlihat pada Gambar 2.

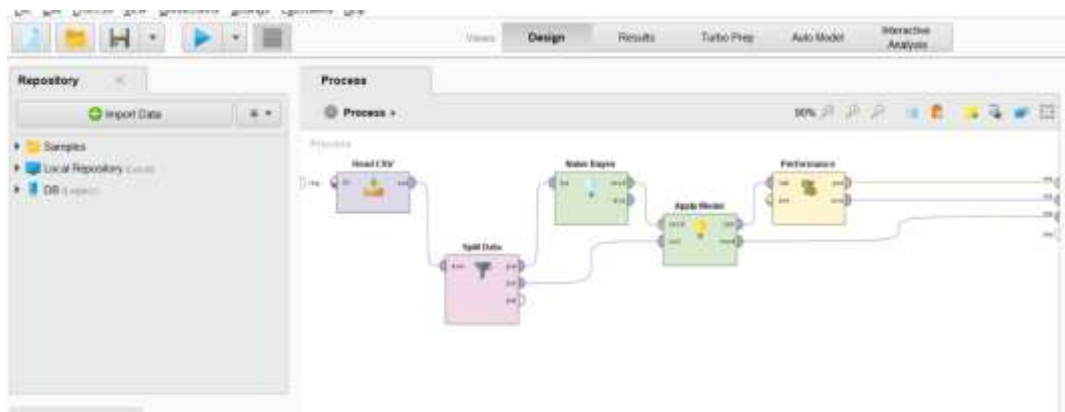
| Row No. | Provinsi | Kab/Kota        | Persentase ... | Rata-rata L... | Pengeluaran... | Indeks Pem... | Uraian Harap... | Persentase ... | Persentase ... | Tingkat Pem... | Tng  |
|---------|----------|-----------------|----------------|----------------|----------------|---------------|-----------------|----------------|----------------|----------------|------|
| 1       | ACEH     | Bireuh          | 18.98          | 9.48           | 7149           | 66.410        | 65.180          | 71.560         | 87.480         | 8.719          | 71.4 |
| 2       | ACEH     | Aceh Singkil    | 20.98          | 8.68           | 8776           | 66.220        | 67.430          | 69.560         | 79.660         | 8.369          | 62.8 |
| 3       | ACEH     | Aceh Selatan    | 13.18          | 8.68           | 8180           | 67.440        | 64.400          | 62.550         | 79.660         | 9.489          | 60.6 |
| 4       | ACEH     | Aceh Tenggara   | 13.41          | 9.67           | 8030           | 66.440        | 68.220          | 62.710         | 88.710         | 8.439          | 68.6 |
| 5       | ACEH     | Aceh Timur      | 14.45          | 8.21           | 8877           | 67.830        | 68.740          | 66.790         | 83.160         | 7.139          | 65.4 |
| 6       | ACEH     | Aceh Tengah     | 15.26          | 9.68           | 10780          | 73.370        | 69.860          | 90.590         | 90.100         | 2.619          | 76.1 |
| 7       | ACEH     | Aceh Barat      | 16.81          | 9.68           | 9993           | 71.670        | 67.990          | 69.630         | 94.220         | 7.089          | 60.0 |
| 8       | ACEH     | Aceh Besar      | 14.05          | 10.33          | 9044           | 73.560        | 69.790          | 87.400         | 82.360         | 7.709          | 61.6 |
| 9       | ACEH     | Pada            | 19.59          | 9              | 9860           | 70.700        | 66.560          | 94.100         | 89.240         | 7.289          | 60.2 |
| 10      | ACEH     | Bireuh          | 13.25          | 9.29           | 8867           | 72.330        | 71.280          | 61.890         | 93.630         | 4.329          | 65.5 |
| 11      | ACEH     | Aceh Utara      | 17.42          | 8.64           | 9201           | 69.460        | 68.810          | 79.970         | 91.080         | 8.319          | 68.4 |
| 12      | ACEH     | Aceh Barat D... | 16.34          | 8.67           | 8428           | 66.390        | 63.660          | 65.710         | 95.340         | 4.049          | 57.5 |
| 13      | ACEH     | Sumatra         | 15.53          | 8.6            | 8845           | 67.660        | 66.430          | 67.670         | 83.630         | 8.449          | 79.0 |

**Gambar 2.** Proses *Read CSV* pada RapidMiner.

Proses pembelajaran model dilakukan dengan membagi dataset menjadi dua bagian, yaitu 70% sebagai data latih dan 30% sebagai data uji menggunakan operator *Split Data*. Algoritma yang digunakan adalah Naive Bayes, salah satu metode *supervised learning* yang berfungsi mengklasifikasikan data berdasarkan prinsip Teorema Bayes dengan asumsi independensi antar atribut. Algoritma Naive Bayes merupakan metode klasifikasi probabilistik berdasarkan Teorema Bayes, dengan asumsi bahwa antar variabel bersifat independen (Mulyani & Pajri, 2024). Rumus umum Naive Bayes adalah sebagai berikut:

$$P(C|X) = (P(X|C) * P(C)) / P(X)$$

Pemodelan dilakukan dengan urutan operator *Read CSV* → *Split Data* → *Naive Bayes* → *Apply Model* → *Performance*, sebagaimana diterapkan pada RapidMiner Studio. Tahapan pemodelan ditunjukkan pada Gambar 3 berikut. Model kemudian dievaluasi dengan menggunakan *confusion matrix* untuk membandingkan hasil prediksi terhadap data aktual.



**Gambar 3.** Proses Klasifikasi di RapidMiner (*Read CSV* → *Split Data* → *Naive Bayes* → *Apply Model* → *Performance*).

Evaluasi dilakukan menggunakan confusion matrix seperti pada tabel berikut:

**Tabel 1.** Evaluasi menggunakan confusion matrix.

|        | True 0 | True 1 |
|--------|--------|--------|
| Pred 0 | 128    | 1      |
| Pred 1 | 8      | 18     |

Rumus akurasi:

$$\text{Akurasi} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \times 100\%$$

$$\text{Akurasi} = (128 + 18) / (128 + 18 + 1 + 8) \times 100\% = 94.19\%$$

Selain menggunakan confusion matrix, hasil prediksi yang diperoleh dari model Naive Bayes disimpan dalam bentuk file hasil.csv, yang kemudian divisualisasikan pada Gambar 4.

| Row No. | Provinsi | Kab/Kota      | Persentase | Rate-rata L. | Pengukuran | Indeks Pem. | Umar Harap. | Persentase | Persentase | Tingkat Pen... |
|---------|----------|---------------|------------|--------------|------------|-------------|-------------|------------|------------|----------------|
| 1       | ACEH     | Pada          | 10.530     | 0            | 9893       | 70.730      | 66.990      | 54.100     | 88.240     | 7.280          |
| 2       | ACEH     | Aceh Barat D. | 16.340     | 8.670        | 8428       | 88.990      | 66.990      | 65.710     | 99.540     | 4.040          |
| 3       | ACEH     | Kota Banda A. | 7.810      | 12.630       | 18891      | 88.710      | 71.820      | 99.880     | 88.370     | 8.940          |
| 4       | ACEH     | Kota Langsa   | 10.960     | 11.120       | 12067      | 77.440      | 69.430      | 91.810     | 97.140     | 7.210          |
| 5       | ACEH     | Kota Subul... | 17.650     | 8.030        | 7385       | 65.270      | 64.070      | 66.710     | 93.870     | 6.260          |
| 6       | SUMATERA | Mendaling N.  | 5.490      | 8.630        | 9771       | 87.190      | 62.690      | 35.730     | 73.780     | 6.120          |
| 7       | SUMATERA | Del Serang    | 4.910      | 10.100       | 12291      | 75.830      | 71.770      | 96.370     | 98.180     | 9.130          |
| 8       | SUMATERA | Humbang H.    | 9.880      | 9.710        | 8016       | 88.410      | 88.910      | 91.880     | 91.880     | 1.940          |
| 9       | SUMATERA | Samar         | 12.980     | 9.660        | 8594       | 70.830      | 71.410      | 91.090     | 85.640     | 3.790          |
| 10      | SUMATERA | Padang Lar.   | 9.920      | 9.380        | 10055      | 70.110      | 67.220      | 67.170     | 77.580     | 3.180          |
| 11      | SUMATERA | Labuhan Bel.  | 8.530      | 8.900        | 11902      | 71.890      | 68.810      | 84.850     | 84.860     | 4.710          |
| 12      | SUMATERA | Nias Utara    | 26.860     | 8.770        | 8195       | 62.820      | 69.860      | 48.090     | 88.170     | 3              |
| 13      | DIKATA   | Kota Bont...  | 78.410     | 8.690        | 8413       | 81.880      | 88.880      | 98.000     | 91.830     | 8.940          |

**Gambar 4.** Tampilan Output hasil.csv dari RapidMiner

Tahap akhir penelitian, yaitu *deployment*, dilakukan dengan menyimpan hasil klasifikasi dalam file *hasil.csv* dan menampilkan visualisasi performa model (Jefri & Fatah, 2025).

#### 4. HASIL DAN PEMBAHASAN

Berdasarkan hasil pengujian menggunakan operator *Performance (Classification)* pada RapidMiner Studio, diperoleh nilai akurasi sebesar 94.19%. Tampilan nilai akurasi serta metrik evaluasi lainnya ditunjukkan pada Gambar 5.

|              | true 0 | true 1 |                         |
|--------------|--------|--------|-------------------------|
| pred 0       | 128    | 1      | class precision: 99.22% |
| pred 1       | 8      | 18     | class precision: 69.23% |
| class recall | 94.12% | 94.74% |                         |

**Gambar 5.** Tampilan Akurasi Model Naive Bayes di RapidMiner.

Penelitian ini menghasilkan model klasifikasi tingkat kemiskinan di Indonesia menggunakan algoritma *Naive Bayes* dengan bantuan perangkat lunak RapidMiner Studio. Hasil utama menunjukkan bahwa model yang dibangun memiliki performa yang sangat baik dengan akurasi sebesar 94,19 %. Nilai ini diperoleh setelah dilakukan serangkaian tahapan pra-pemrosesan, pembagian data, pemodelan, serta evaluasi menggunakan *confusion matrix* dan metrik performa lainnya. Angka akurasi ini menunjukkan bahwa model mampu mengenali pola sosial-ekonomi masyarakat Indonesia dengan tingkat ketepatan yang tinggi, sehingga dapat digunakan sebagai alat bantu dalam pengambilan keputusan berbasis data (*data-driven decision making*) di bidang kebijakan public (Novianti, 2019).

Proses klasifikasi dilakukan dengan menggunakan 155 data yang mewakili kondisi sosial dan ekonomi di beberapa wilayah Indonesia, yang memiliki 12 atribut utama seperti tingkat pengeluaran rumah tangga, pendidikan kepala keluarga, status pekerjaan, tingkat pengangguran, serta indeks pembangunan manusia (IPM). Data dibagi menjadi dua bagian dengan proporsi 70% sebagai data latih dan 30% sebagai data uji menggunakan operator *Split Data* pada RapidMiner. Pembagian ini dilakukan untuk menjaga validitas model dan menghindari *overfitting*, yaitu kondisi di mana model terlalu menyesuaikan dengan data pelatihan sehingga gagal menggeneralisasi pada data baru.

Hasil evaluasi diperoleh melalui *confusion matrix* yang menunjukkan distribusi hasil klasifikasi antara nilai aktual dan prediksi. Berdasarkan tabel *confusion matrix*, diperoleh 128 data benar diklasifikasikan sebagai “tidak miskin” (*true negative*), 18 data benar diklasifikasikan sebagai “miskin” (*true positive*), 8 data salah diklasifikasikan (*false positive*), dan 1 data salah tidak terklasifikasi (*false negative*). Dengan demikian, akurasi model dihitung menggunakan rumus:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{18 + 128}{18 + 128 + 8 + 1} \times 100\% = 94,19\%.$$

Nilai ini membuktikan bahwa sebagian besar data berhasil diprediksi dengan benar oleh algoritma *Naive Bayes*. Selain akurasi, nilai *precision*, *recall*, dan *F-measure* juga diperoleh di atas 0,9, yang menunjukkan keseimbangan antara kemampuan model mengenali data miskin dan tidak miskin. Visualisasi hasil evaluasi model ditunjukkan dalam Gambar 5, yang memperlihatkan tampilan performa klasifikasi pada RapidMiner Studio.

Nilai akurasi sebesar 94,19% tergolong tinggi untuk analisis data sosial, di mana sering ditemukan variabel yang heterogen dan sulit diprediksi secara pasti. Hasil ini memperkuat efektivitas *Naive Bayes* dalam mengolah data dengan karakteristik kompleks namun berstruktur jelas. Algoritma ini menghitung probabilitas posterior untuk setiap kelas berdasarkan nilai atribut yang diamati, dan meskipun menggunakan asumsi independensi antaratribut, performanya tetap optimal untuk dataset berskala menengah seperti yang digunakan dalam penelitian ini. Sejalan dengan pendapat Chairani *et al.* (2021), *Naive Bayes* termasuk algoritma dengan efisiensi tinggi dan akurasi stabil, terutama ketika diterapkan pada dataset sosial yang memiliki variabilitas tinggi.

Hasil penelitian ini juga sejalan dengan temuan Putri *et al.* (2024) yang menerapkan algoritma *Naive Bayes* untuk klasifikasi rumah tangga miskin di Kota Padang dan memperoleh akurasi 93,5%. Demikian pula, Imanda & Nurdin (2024) membuktikan bahwa model serupa mampu menghasilkan akurasi di atas 90% dalam pemetaan keluarga miskin di Kabupaten Majalengka. Hasil-hasil tersebut menunjukkan konsistensi efektivitas *Naive Bayes* untuk data sosial-ekonomi Indonesia, di mana pola kemiskinan sering kali dapat diidentifikasi melalui indikator pendidikan, pekerjaan, dan pengeluaran rumah tangga.

File *hasil.csv* yang dihasilkan oleh RapidMiner memuat hasil prediksi untuk setiap entri data dan menjadi bukti keluaran model. Tampilan file ini disajikan pada Gambar 4, yang memperlihatkan kolom label aktual dan hasil prediksi untuk setiap sampel. Sebagian besar prediksi menunjukkan kecocokan antara label aktual dan label hasil klasifikasi, sehingga memperkuat nilai akurasi yang diperoleh. Hasil ini menegaskan kemampuan algoritma dalam mengenali karakteristik kemiskinan secara probabilistik berdasarkan kombinasi atribut sosial-ekonomi.

Dalam konteks *machine learning*, keberhasilan model ini juga dipengaruhi oleh tahap pra-pemrosesan data (*data preprocessing*) yang tepat. Normalisasi data dan penghapusan nilai kosong berperan penting dalam menjaga konsistensi hasil. Penelitian sebelumnya oleh Aji *et al.* (2023) menunjukkan bahwa kualitas data memiliki pengaruh signifikan terhadap stabilitas model klasifikasi berbasis *Naive Bayes*. Oleh karena itu, perhatian terhadap integritas data menjadi langkah penting dalam penerapan algoritma pembelajaran mesin untuk analisis sosial.

Jika dibandingkan dengan algoritma lain seperti *Decision Tree* atau *K-Nearest Neighbor (KNN)*, *Naive Bayes* memiliki keunggulan pada efisiensi komputasi dan kestabilan hasil. Penelitian Ahmad et al. (2023) menemukan bahwa *Naive Bayes* lebih cepat dalam memproses dataset kemiskinan dibandingkan *Decision Tree*, dengan selisih waktu komputasi mencapai 35% lebih efisien. Hal ini menjadikan *Naive Bayes* pilihan ideal untuk implementasi dalam konteks kebijakan publik, di mana data sering diperbarui secara berkala dan model perlu dijalankan secara efisien.

Dari hasil analisis probabilitas kondisi di RapidMiner, diketahui bahwa atribut yang paling berpengaruh terhadap klasifikasi adalah rata-rata pengeluaran rumah tangga, tingkat pendidikan, dan indeks pembangunan manusia (IPM). Ketiga variabel ini menunjukkan korelasi yang kuat dengan status kemiskinan. Misalnya, daerah dengan rata-rata pengeluaran rumah tangga di bawah batas garis kemiskinan BPS cenderung diklasifikasikan sebagai “miskin”, sedangkan wilayah dengan IPM tinggi umumnya masuk kategori “tidak miskin”. Hasil ini mendukung temuan Lembaga Kita (2024) yang menegaskan bahwa kemiskinan di Indonesia sangat dipengaruhi oleh akses terhadap pendidikan dan kemampuan ekonomi keluarga.

Faktor pendidikan juga menjadi aspek kunci dalam mengurangi kemiskinan karena berpengaruh langsung terhadap peluang kerja dan produktivitas masyarakat. Menurut Aji et al. (2023), peningkatan kualitas pendidikan mampu menurunkan tingkat kemiskinan melalui perbaikan kompetensi tenaga kerja dan peningkatan pendapatan rumah tangga. Dalam konteks ini, model klasifikasi berbasis *Naive Bayes* tidak hanya bermanfaat untuk identifikasi status kemiskinan, tetapi juga dapat digunakan untuk memetakan wilayah yang membutuhkan intervensi kebijakan pendidikan dan ekonomi.

Keunggulan lain dari penelitian ini terletak pada penggunaan RapidMiner yang menyediakan lingkungan analisis visual tanpa memerlukan kode pemrograman. Hal ini mempercepat proses eksperimen dan memungkinkan analisis data sosial dilakukan dengan lebih mudah oleh peneliti dari berbagai bidang. Fitur *Apply Model* dan *Performance Evaluation* yang dimiliki RapidMiner membantu menampilkan hasil analisis secara langsung, baik dalam bentuk tabel maupun grafik performa model. Dengan demikian, RapidMiner berperan penting dalam memastikan bahwa seluruh tahapan CRISP-DM dari *data understanding* hingga *evaluation* berjalan secara sistematis.

Selain untuk keperluan akademik, hasil penelitian ini memiliki potensi besar untuk diaplikasikan dalam kebijakan publik (Martinez & Cooray, 2025). Model klasifikasi ini dapat digunakan oleh instansi pemerintah seperti BPS atau Kementerian Sosial untuk memperbaiki

akurasi pendataan masyarakat miskin. Dengan model *machine learning* yang terlatih, distribusi bantuan sosial dapat diarahkan lebih tepat sasaran, sehingga mengurangi risiko ketidaktepatan penerima manfaat. Hal ini sejalan dengan rekomendasi Hidayat & Henryka (2024) yang menekankan pentingnya pemanfaatan analisis data dalam kebijakan sosial agar lebih efisien dan transparan.

Namun, perlu dicatat bahwa hasil model ini masih bergantung pada kualitas dan jumlah data. Jika dataset diperluas, misalnya mencakup data antarprovinsi atau lintas waktu (*time series*), akurasi model bisa berubah. Oleh karena itu, pengembangan penelitian selanjutnya dapat diarahkan pada perbandingan algoritma lain seperti *Random Forest*, *Support Vector Machine (SVM)*, atau *Neural Network* untuk menilai kestabilan model. Selain itu, metode *feature selection* juga dapat digunakan untuk mengidentifikasi atribut yang paling relevan terhadap klasifikasi kemiskinan secara lebih mendalam (Khalik & Arifin, 2023).

Dengan demikian, hasil penelitian ini menegaskan bahwa penggunaan algoritma *Naive Bayes* mampu memberikan hasil klasifikasi kemiskinan yang akurat, efisien, dan mudah diimplementasikan. Nilai akurasi 94,19% menunjukkan tingkat ketepatan tinggi dalam memprediksi status ekonomi rumah tangga berdasarkan data sosial yang sederhana. Pendekatan ini dapat dijadikan dasar pengembangan sistem informasi kemiskinan berbasis kecerdasan buatan yang lebih komprehensif. Secara praktis, hasil penelitian ini memberikan kontribusi terhadap upaya digitalisasi kebijakan sosial dan pembangunan berkelanjutan di Indonesia, sejalan dengan visi *Indonesia Emas 2045* yang menekankan pentingnya transformasi digital dalam tata kelola pemerintahan dan pelayanan publik.

## 5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa penerapan algoritma *Naive Bayes* dalam proses klasifikasi tingkat kemiskinan di Indonesia menggunakan RapidMiner Studio terbukti efektif dan akurat. Model yang dibangun mampu mengklasifikasikan status kemiskinan berdasarkan 12 atribut sosial dan ekonomi dengan tingkat akurasi sebesar 94,19%, yang menunjukkan bahwa metode ini memiliki kemampuan prediktif tinggi serta stabil dalam mengolah data heterogen. Atribut yang paling berpengaruh dalam penentuan kategori kemiskinan adalah pengeluaran rumah tangga, tingkat pendidikan, dan indeks pembangunan manusia (IPM), yang seluruhnya berkaitan erat dengan dimensi kesejahteraan masyarakat. Penggunaan perangkat lunak RapidMiner memberikan kemudahan bagi peneliti untuk melakukan pra-pemrosesan, pelatihan model, serta evaluasi hasil secara sistematis melalui pendekatan CRISP-DM. Implementasi model klasifikasi berbasis kecerdasan buatan seperti ini dapat menjadi solusi

strategis dalam meningkatkan ketepatan penyaluran program bantuan sosial dan evaluasi kesejahteraan masyarakat secara data-driven. Untuk penelitian selanjutnya, disarankan melakukan perbandingan antar-algoritma seperti Decision Tree, Random Forest, atau Neural Network dengan dataset yang lebih besar agar diperoleh model prediksi kemiskinan yang lebih komprehensif dan adaptif terhadap variasi data wilayah di Indonesia.

## DAFTAR REFERENSI

- Ahmad, C. F., Suarna, N., & Dwilestari, G. (2023). Klasifikasi data kemiskinan menggunakan metode Naïve Bayes untuk mengetahui tingkat kemiskinan (Studi kasus: Desa Karangasem). *Jurnal Informatika dan Teknologi Informasi*, 2(2), 203–208. <https://doi.org/10.56854/jt.v2i2.190>
- Aji, P., Rahman, D., Hernawati, E., & Yualinda, S. (2023). Poverty level prediction based on e-commerce data using Naïve Bayes algorithm and similarity-based feature selection. *IJAIT*, 7(2), 114–126. <https://doi.org/10.25124/ijait.v7i02.5374>
- Amartya, S. (1999). *Development as freedom*. Oxford University Press.
- Badan Pusat Statistik. (2024). *Profil kemiskinan di Indonesia Maret 2024*. BPS. <https://www.bps.go.id>
- Chairani, A., Hariyanto, F. E., Luh, N., Andini, E., & S, Z. C. (2021). Klasifikasi rumah tangga miskin menggunakan metode Naive Bayes (Studi kasus: Provinsi Papua tahun 2017). *Jurnal Sains Matematika dan Statistika*, 7(1), 89–95. <https://doi.org/10.24014/jsms.v7i1.11924>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Hidayat, M. F., & Henryka, D. N. (2024). Forecasting poverty ratios in Indonesia: A time series modeling approach. *Jurnal EMACS*, 6(3), 219–227. <https://doi.org/10.21512/emacsjournal.v6i3.11968>
- Imanda, N., & Nurdin. (2024). Penerapan algoritma Naive Bayes pada klasifikasi penduduk kurang mampu dan mampu di Tanoh Anou Idi Rayeuk. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3), 4674–4681. <https://doi.org/10.23960/jitet.v12i3S1.5463>
- Jefri, & Fatah, Z. (2025). Klasifikasi data mining untuk memprediksi kelulusan mahasiswa. *Jurnal Ilmiah Multidisiplin Ilmu*, 2(1), 29–37. <https://doi.org/10.69714/mhjq1v85>
- Khalik, M. F. M., & Arifin, F. (2023). Klasifikasi indeks kedalaman kemiskinan Provinsi Sulawesi Selatan berbasis Decision Tree, K-Nearest Neighbor, Naive Bayes, Neural Network, dan Random Forest. *Jurnal Edukasi dan Penelitian Informatika*, 9(2), 282–288.

- Kshirsagar, V., & Wells, R. (2017). Household poverty classification in data-scarce environments: A machine learning approach. 31st Conference on Neural Information Processing Systems, 1–17.
- Kurniawan, D. (2020). Kemiskinan di Indonesia dan solusinya. *Jurnal Indonesia*, 1(1), 1–18.
- Maharani, C., Ningrum, D. A., Fatmawati, A. E., & Fadilla, A. (2024). Dampak kemiskinan terhadap kualitas pendidikan anak di Indonesia: Rekomendasi kebijakan yang efektif. *Journal of Macroeconomics and Social Development*, 1(3), 1–10. <https://economics.pubmedia.id/index.php/jmsd>
- Mansur, Y. (2024). Analisis perkembangan penduduk miskin, karakteristik kemiskinan dan kedalaman kemiskinan di Indonesia. *Jurnal Ekonomi dan Manajemen Teknologi (EMT)*, 8(1), 18–31. <https://doi.org/10.35870/emt.v8i1.1930>
- Martinez, R. G., & Cooray, M. (2025). Enhancing poverty targeting with spatial machine learning: An application to Indonesia. *Jurnal Econ*, 1(1), 1–47.
- Mulyani, S., & Pajri, A. E. (2024). Klasifikasi tingkat kemiskinan di Indonesia menggunakan algoritma Naïve Bayes. *Journal of Computers Sciences and Informatics*, 1(2), 53–57. <https://doi.org/10.34304/scientific.v1i2.333>
- Novianti, D. (2019). Implementasi algoritma Naïve Bayes pada dataset hepatitis menggunakan RapidMiner. *Paradigma – Jurnal Komputer dan Informatika*, 21(1), 49–54. <https://doi.org/10.31294/p.v20i2>
- Olearo, L., Adda, F. D., Messina, V., Cremaschi, M., Bandini, S., & Gasparini, F. (2023). An artificial intelligence approach to predict multidimensional poverty of older people from unlabelled data. *CEUR Workshop Proceedings*, 1–15.
- Putri, A. K., Fitria, D., & Mukhti, T. O. (2024). Classification of poor households in Padang City using the Naïve Bayes algorithm with Synthetic Minority Oversampling Technique. *UNP Journal of Statistics and Data Science*, 2(4), 446–452. <https://doi.org/10.24036/ujsds/vol2-iss4/241>
- Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining* (2nd ed.). Pearson Education.
- Todaro, M. P., & Smith, S. C. (2020). *Economic development* (13th ed.). Pearson Education.
- Wulansari, R. Y., Fadhilah, N., & Huda, M. (2023). Faktor yang mempengaruhi kemiskinan di Indonesia. *Journal of Economic, Management, Accounting and Technology*, 6(1), 82–95. <https://doi.org/10.32500/jematech.v6i1.3928>