BAB II
TINJAUAN PUSTAKA

2.1 Kemiskinan

Kemiskinan merupakan kondisi ketidakmampuan seseorang atau rumah tangga dalam memenuhi kebutuhan dasar yang meliputi kebutuhan pangan dan non pangan. Kebutuhan tersebut mencakup makanan, sandang, tempat tinggal, pendidikan, serta layanan kesehatan yang layak. Kemiskinan menjadi indikator penting dalam mengukur tingkat kesejahteraan masyarakat pada suatu wilayah. Semakin tinggi tingkat kemiskinan, maka semakin rendah tingkat kesejahteraan masyarakat di wilayah tersebut.

Menurut Badan Pusat Statistik, kemiskinan merupakan kondisi ketika seseorang memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan yang telah ditetapkan. Garis kemiskinan tersebut dihitung berdasarkan kebutuhan minimum makanan dan non makanan yang harus dipenuhi agar seseorang dapat hidup secara layak (Badan Pusat Statistik, 2024). Pendekatan ini menunjukkan bahwa faktor ekonomi seperti penghasilan dan pekerjaan memiliki pengaruh besar dalam menentukan tingkat kemiskinan.

Kemiskinan juga dapat dipengaruhi oleh beberapa faktor sosial ekonomi lainnya, seperti tingkat pendidikan, jenis pekerjaan, serta kondisi tempat tinggal. Tingkat pendidikan yang rendah dapat membatasi peluang seseorang dalam memperoleh pekerjaan yang layak sehingga berpengaruh terhadap tingkat pendapatan. Selain itu, pekerjaan dengan pendapatan tidak tetap juga berpotensi meningkatkan risiko kemiskinan. Kondisi tempat tinggal seperti kepemilikan

rumah dan jenis lantai juga sering digunakan sebagai indikator dalam menentukan tingkat kesejahteraan masyarakat (Sari & Nugroho, 2022).

Dalam penelitian ini, indikator kemiskinan yang digunakan meliputi penghasilan, pekerjaan, pendidikan kepala keluarga, kepemilikan rumah, dan jenis lantai. Variabel tersebut dipilih karena memiliki keterkaitan langsung dengan kondisi sosial ekonomi masyarakat. Penggunaan indikator tersebut diharapkan mampu menggambarkan tingkat kemiskinan secara lebih objektif dan dapat digunakan dalam proses klasifikasi menggunakan metode *data mining*.

Dengan memanfaatkan indikator-indikator tersebut, proses klasifikasi tingkat kemiskinan dapat dilakukan secara sistematis menggunakan algoritma *Decision Tree*. Metode ini mampu mengelompokkan data masyarakat berdasarkan karakteristik tertentu sehingga dapat diketahui kategori miskin dan tidak miskin secara lebih cepat dan akurat.

2.2 Konsep Prediksi Tingkat Kemiskinan

Prediksi tingkat kemiskinan merupakan suatu proses untuk memperkirakan kondisi ekonomi masyarakat berdasarkan data yang tersedia dengan tujuan mengelompokkan masyarakat ke dalam kategori tertentu, seperti miskin dan tidak miskin. Proses ini dilakukan dengan menganalisis berbagai faktor yang mempengaruhi tingkat kesejahteraan, seperti penghasilan, pekerjaan, pendidikan, dan kondisi tempat tinggal.

Dalam perkembangannya, prediksi tingkat kemiskinan tidak hanya dilakukan secara konvensional, tetapi telah memanfaatkan teknologi berbasis *data mining*. *Data mining* merupakan proses pengolahan data dalam jumlah besar untuk

menemukan pola atau informasi yang tersembunyi yang dapat digunakan dalam pengambilan keputusan (Sari & Nugroho, 2022).

Konsep prediksi tingkat kemiskinan termasuk dalam permasalahan klasifikasi, dimana data masyarakat dikelompokkan ke dalam kelas tertentu berdasarkan atribut yang dimiliki. Salah satu metode yang sering digunakan dalam klasifikasi adalah algoritma *Decision Tree*. Metode ini bekerja dengan membentuk struktur pohon keputusan yang terdiri dari *root*, *branch*, dan *leaf*, sehingga menghasilkan aturan keputusan (*decision rules*) yang mudah dipahami.

Menurut penelitian oleh Pratama dan Wibowo (2024), penggunaan algoritma *Decision Tree* dalam prediksi penerima bantuan sosial mampu menghasilkan tingkat akurasi yang tinggi serta memberikan hasil klasifikasi yang transparan. Hal ini menunjukkan bahwa metode berbasis *data mining* lebih objektif dibandingkan dengan metode manual yang cenderung subjektif.

Selain itu, keberhasilan prediksi tingkat kemiskinan sangat dipengaruhi oleh pemilihan indikator yang digunakan. Indikator tersebut umumnya mencakup aspek ekonomi dan sosial, seperti penghasilan, pekerjaan, dan tingkat pendidikan. Pemilihan variabel yang relevan akan meningkatkan kualitas model prediksi yang dihasilkan (Saputra & Wibowo, 2023).

Dalam penelitian ini, proses prediksi dilakukan menggunakan algoritma *Decision Tree* dengan memanfaatkan data simulasi masyarakat. Data tersebut kemudian diolah melalui beberapa tahapan, yaitu *preprocessing*, pembentukan model, serta evaluasi menggunakan metode *cross validation* dan *confusion matrix*.

Hasil dari proses ini berupa model klasifikasi yang mampu memprediksi tingkat kemiskinan secara sistematis.

Keunggulan penggunaan algoritma *Decision Tree* dalam prediksi tingkat kemiskinan adalah kemampuannya dalam menghasilkan model yang sederhana, mudah dipahami, serta dapat divisualisasikan. Selain itu, metode ini juga mampu menunjukkan atribut yang paling berpengaruh dalam proses klasifikasi, sehingga dapat membantu dalam analisis faktor penyebab kemiskinan.

Dengan adanya konsep prediksi tingkat kemiskinan berbasis *data mining*, diharapkan proses identifikasi masyarakat miskin dapat dilakukan secara lebih cepat, akurat, dan objektif. Hasil prediksi ini juga dapat digunakan sebagai dasar dalam pengambilan kebijakan, khususnya dalam penyaluran bantuan sosial agar lebih tepat sasaran.

2.3 *Data Mining*

Data mining merupakan proses penggalian informasi dari kumpulan data dalam jumlah besar untuk menemukan pola tersembunyi yang dapat digunakan sebagai dasar pengambilan keputusan. Teknik ini memanfaatkan metode statistik, *machine learning*, dan basis data untuk mengidentifikasi hubungan antar variabel yang sebelumnya tidak terlihat secara langsung. Dalam perkembangan teknologi informasi saat ini, *data mining* banyak digunakan dalam berbagai bidang seperti bisnis, kesehatan, pendidikan, dan sosial ekonomi masyarakat.

Menurut Han, Kamber, dan Pei (2022), *data mining* adalah proses mengekstraksi pengetahuan yang berguna dari data berukuran besar dengan menggunakan teknik analisis tertentu sehingga menghasilkan informasi yang dapat

dimanfaatkan dalam proses pengambilan keputusan. Proses ini biasanya dilakukan melalui beberapa tahapan yang dikenal sebagai proses *Knowledge Discovery in Database* (*KDD*), yaitu seleksi data, *preprocessing*, transformasi data, *data mining*, dan evaluasi hasil.

Tahapan dalam *data mining* secara umum dapat dijelaskan sebagai berikut:

1. *Data Selection*

   Tahap ini merupakan proses pemilihan data yang relevan dengan tujuan penelitian. Data yang digunakan harus sesuai dengan variabel yang akan dianalisis sehingga hasil yang diperoleh lebih akurat.

2. *Preprocessing*

   Tahap ini dilakukan untuk membersihkan data dari kesalahan, duplikasi, serta nilai yang tidak konsisten. Proses ini bertujuan agar data yang digunakan siap untuk dianalisis.

3. *Transformation*

   Pada tahap ini data diubah ke dalam format yang sesuai agar dapat diproses oleh algoritma *data mining*. Transformasi dapat berupa normalisasi atau pengelompokan data.

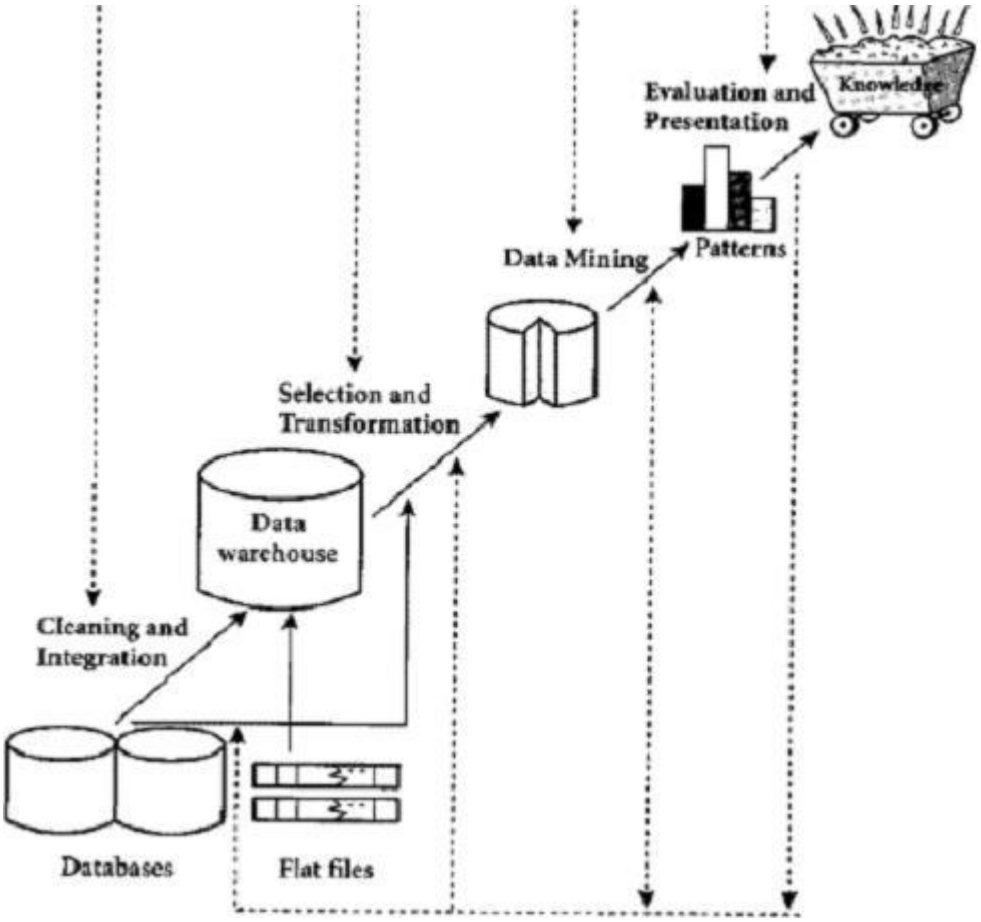
4. *Data Mining*

   Tahap ini merupakan proses utama untuk menemukan pola menggunakan algoritma tertentu. Dalam penelitian ini, algoritma yang digunakan adalah *Decision Tree* untuk melakukan klasifikasi tingkat kemiskinan.

5. *Evaluation*

   Tahap terakhir adalah mengevaluasi hasil model yang dihasilkan untuk

mengetahui tingkat akurasi dan performa klasifikasi. Evaluasi dilakukan menggunakan *confusion matrix* dan metode *cross validation.*



Gambar 2.1 Tahapan Data Mining

Sumber: BINUS University (2016)

Dalam penelitian ini, teknik *data mining* digunakan untuk mengklasifikasikan tingkat kemiskinan masyarakat berdasarkan variabel sosial ekonomi. Data yang digunakan terlebih dahulu melalui tahap *preprocessing* sebelum diproses menggunakan algoritma *Decision Tree*. Hasil dari proses ini berupa model klasifikasi yang dapat digunakan untuk memprediksi tingkat kemiskinan secara lebih cepat dan sistematis.

Dengan menggunakan pendekatan *data mining*, proses analisis data menjadi lebih efisien serta mampu menghasilkan informasi yang bermanfaat dalam mendukung pengambilan keputusan. Oleh karena itu, metode ini sangat sesuai digunakan dalam penelitian klasifikasi tingkat kemiskinan masyarakat.

2.4 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (*KDD*) merupakan proses keseluruhan dalam menemukan pengetahuan atau informasi yang bermanfaat dari kumpulan data dalam jumlah besar. *KDD* tidak hanya mencakup proses *data mining*, tetapi juga melibatkan tahapan lain seperti pemilihan data, pembersihan data, transformasi data, hingga evaluasi hasil.

Menurut Kusrini dan Luthfi (2021), *KDD* adalah proses sistematis yang digunakan untuk mengekstraksi informasi penting dari data yang sebelumnya tidak diketahui, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan. Proses ini sangat penting dalam penelitian berbasis *data mining* karena membantu memastikan bahwa data yang digunakan telah melalui tahapan yang benar sebelum dilakukan analisis.

Secara umum, tahapan dalam proses *KDD* terdiri dari beberapa langkah sebagai berikut:

1. **Data Selection**

   Tahap ini merupakan proses pemilihan data yang relevan dengan tujuan penelitian. Data yang digunakan harus sesuai dengan variabel yang akan dianalisis agar hasil yang diperoleh lebih akurat.

2. **Data Cleaning (Preprocessing)**

   Pada tahap ini dilakukan pembersihan data dari kesalahan, duplikasi, serta nilai yang tidak konsisten. Proses ini bertujuan untuk meningkatkan kualitas data sebelum dilakukan analisis.

3. **Data Transformation**

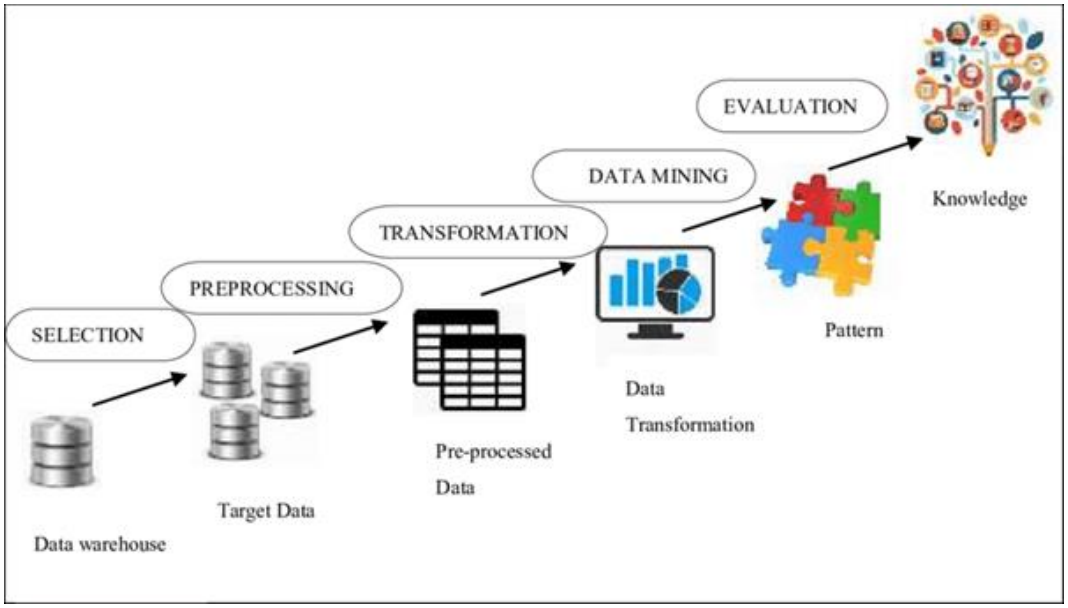
   Data yang telah dibersihkan kemudian diubah ke dalam format yang sesuai agar dapat diproses oleh algoritma. Transformasi ini dapat berupa pengelompokan data atau normalisasi.

4. **Data Mining**

   Tahap ini merupakan inti dari proses *KDD*, dimana dilakukan pencarian pola menggunakan algoritma tertentu. Dalam penelitian ini, algoritma yang digunakan adalah *Decision Tree* untuk melakukan klasifikasi tingkat kemiskinan.

5. **Evaluation (Interpretation)**

   Tahap terakhir adalah mengevaluasi hasil yang diperoleh dari proses *data mining*. Evaluasi dilakukan untuk mengetahui apakah pola yang ditemukan sudah sesuai dengan tujuan penelitian serta memiliki tingkat akurasi yang baik.



Gambar 2.2 Alur KDD

Sumber: BINUS (2021)

Dalam penelitian ini, proses *KDD* diterapkan secara sistematis mulai dari penyusunan dataset menggunakan *Microsoft Excel*, kemudian dilakukan tahap *preprocessing* untuk membersihkan data, dilanjutkan dengan proses klasifikasi menggunakan algoritma *Decision Tree* pada aplikasi *Orange Data Mining*, serta evaluasi hasil menggunakan *confusion matrix* dan *cross validation*.

Penerapan konsep *KDD* dalam penelitian ini bertujuan untuk memastikan bahwa proses pengolahan data dilakukan secara terstruktur sehingga menghasilkan model klasifikasi yang akurat dan dapat dipercaya. Selain itu, penggunaan tahapan *KDD* juga membantu dalam memahami alur proses analisis data secara menyeluruh.

Dengan demikian, *KDD* menjadi kerangka kerja yang penting dalam penelitian ini karena mengintegrasikan seluruh proses mulai dari pengolahan data hingga

menghasilkan informasi yang bermanfaat dalam memprediksi tingkat kemiskinan masyarakat.

2.5 Algoritma *Decision Tree*

Algoritma *Decision Tree* merupakan salah satu metode klasifikasi dalam teknik *data mining* yang digunakan untuk menghasilkan model prediksi dalam bentuk struktur pohon. Struktur tersebut terdiri dari akar (*root*), cabang (*branch*), dan daun (*leaf*) yang menggambarkan aturan keputusan berdasarkan atribut tertentu. Metode ini banyak digunakan karena mudah dipahami serta mampu menjelaskan proses pengambilan keputusan secara transparan.

Menurut Han, Kamber, dan Pei (2022), *Decision Tree* adalah metode klasifikasi yang membangun model dalam bentuk pohon dengan membagi data ke dalam beberapa kelompok berdasarkan atribut yang paling berpengaruh terhadap variabel target. Proses pembentukan pohon dilakukan secara rekursif hingga setiap cabang menghasilkan kelas yang homogen. Metode ini sangat efektif digunakan untuk data kategorikal maupun numerik.

Algoritma *Decision Tree* bekerja dengan memilih atribut terbaik sebagai akar (*root node*) berdasarkan nilai perhitungan tertentu seperti *entropy*, *information gain*, atau *gain ratio*. Atribut dengan nilai tertinggi akan dipilih sebagai pemisah utama. Proses ini kemudian dilanjutkan pada setiap cabang hingga seluruh data dapat diklasifikasikan dengan baik (Suyanto, 2021).

Struktur *Decision Tree* secara umum terdiri dari tiga bagian utama, yaitu:

1. *Root Node*

   Merupakan node paling atas dalam struktur pohon yang digunakan sebagai

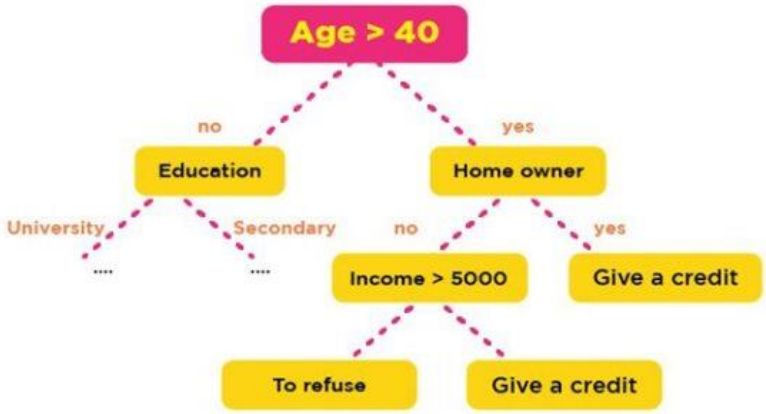
dasar pengambilan keputusan pertama. Atribut pada bagian ini memiliki pengaruh paling besar terhadap klasifikasi.

2. *Branch*

   Merupakan cabang yang menghubungkan antara node satu dengan node lainnya. Cabang menunjukkan hasil dari pengujian suatu atribut.

3. *Leaf Node*

   Merupakan node akhir yang berisi hasil klasifikasi. Pada bagian ini data telah masuk ke dalam kategori tertentu seperti miskin atau tidak miskin.



Gambar 2.3 Struktur Decision Tree

Sumber: BINUS University (2020)

Keunggulan algoritma *Decision Tree* antara lain sebagai berikut:

1. Mudah dipahami dan diinterpretasikan
2. Tidak memerlukan perhitungan yang kompleks
3. Dapat menangani data numerik dan kategorikal
4. Dapat divisualisasikan dalam bentuk pohon keputusan

5. Proses klasifikasi berjalan cepat

Namun demikian, algoritma *Decision Tree* juga memiliki beberapa kelemahan, di antaranya:

1. Rentan terhadap *overfitting* apabila data terlalu spesifik
2. Sensitif terhadap perubahan data
3. Struktur pohon dapat menjadi kompleks jika jumlah atribut banyak

Dalam penelitian ini, algoritma *Decision Tree* digunakan untuk mengklasifikasikan tingkat kemiskinan masyarakat berdasarkan variabel penghasilan, pekerjaan, pendidikan kepala keluarga, kepemilikan rumah, dan jenis lantai. Proses pembentukan pohon keputusan dilakukan menggunakan aplikasi *Orange Data Mining* sehingga menghasilkan model klasifikasi yang mudah dipahami. Model tersebut kemudian digunakan untuk memprediksi tingkat kemiskinan masyarakat ke dalam kategori miskin dan tidak miskin.

2.6 *Entropy*

Entropy merupakan ukuran yang digunakan untuk mengetahui tingkat ketidakpastian atau tingkat keacakan suatu data. Dalam algoritma *Decision Tree*, nilai *entropy* digunakan untuk menentukan seberapa baik suatu atribut dalam memisahkan data ke dalam kelas tertentu. Semakin kecil nilai *entropy*, maka data semakin homogen, sedangkan semakin besar nilai *entropy*, maka data semakin heterogen.

Menurut Han, Kamber, dan Pei (2022), *entropy* digunakan untuk mengukur tingkat impurity pada suatu kumpulan data. Jika semua data berada pada satu kelas, maka nilai *entropy* bernilai nol. Sebaliknya, jika data terbagi secara merata pada

beberapa kelas, maka nilai *entropy* akan semakin besar. Oleh karena itu, *entropy* menjadi dasar dalam menentukan atribut terbaik dalam pembentukan *Decision Tree*.

Rumus *entropy* dapat dituliskan sebagai berikut:

$$Entropy(S) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Keterangan:

- *S* = himpunan data
- *pi* = proporsi data pada kelas ke-i
- *n* = jumlah kelas

Nilai *entropy* berada pada rentang 0 sampai 1. Jika nilai *entropy* mendekati 0 maka data semakin homogen, sedangkan jika nilai *entropy* mendekati 1 maka data semakin tidak homogen. Dalam proses pembentukan *Decision Tree*, atribut yang menghasilkan pembagian data dengan nilai *entropy* paling kecil akan dipilih sebagai node terbaik.

Pada penelitian ini, perhitungan *entropy* digunakan untuk menentukan atribut yang paling berpengaruh dalam klasifikasi tingkat kemiskinan. Dataset yang digunakan terdiri dari dua kelas yaitu miskin dan tidak miskin. Nilai *entropy* awal dihitung berdasarkan distribusi jumlah data pada masing-masing kelas, kemudian digunakan dalam perhitungan *information gain* untuk menentukan akar pohon keputusan.

Dengan menggunakan perhitungan *entropy*, proses pemilihan atribut dalam algoritma *Decision Tree* menjadi lebih sistematis sehingga model klasifikasi yang dihasilkan lebih optimal dan mudah dipahami.

2.7 *Information Gain*

Information Gain merupakan ukuran yang digunakan untuk menentukan atribut terbaik dalam pembentukan algoritma *Decision Tree*. Nilai *Information Gain* diperoleh dari selisih antara nilai *entropy* sebelum dilakukan pemisahan data dengan nilai *entropy* setelah data dibagi berdasarkan suatu atribut. Atribut yang memiliki nilai *Information Gain* tertinggi akan dipilih sebagai node terbaik karena mampu memberikan informasi paling besar dalam proses klasifikasi.

Menurut Han, Kamber, dan Pei (2022), *Information Gain* digunakan untuk mengukur seberapa efektif suatu atribut dalam mengurangi ketidakpastian (*entropy*) pada dataset. Semakin besar nilai *Information Gain*, maka semakin baik atribut tersebut dalam memisahkan data ke dalam kelas yang berbeda. Oleh karena itu, *Information Gain* menjadi salah satu metode utama dalam pemilihan atribut pada algoritma *Decision Tree*.

Rumus *Information Gain* dituliskan sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|\ S_i\ |}{|\ S\ |} Entropy(S_i)$$

Keterangan:

- *Gain(S,A)* = nilai *Information Gain* atribut A
- *Entropy(S)* = nilai *entropy* seluruh data
- *Si* = subset data berdasarkan atribut A
- |*Si*| = jumlah data pada subset ke-i
- |*S*| = jumlah seluruh data
- *Entropy(Si)* = nilai *entropy* pada subset ke-i

Dalam algoritma *Decision Tree*, setiap atribut akan dihitung nilai *Information Gain*-nya. Atribut dengan nilai tertinggi akan dipilih sebagai akar (*root node*) karena dianggap paling mampu membagi data secara optimal. Proses ini kemudian dilakukan secara berulang pada setiap cabang hingga terbentuk struktur pohon keputusan yang lengkap.

Pada penelitian ini, perhitungan *Information Gain* digunakan untuk menentukan atribut yang paling berpengaruh dalam klasifikasi tingkat kemiskinan. Atribut seperti penghasilan, pekerjaan, dan pendidikan dihitung nilai *Information Gain*-nya, kemudian atribut dengan nilai tertinggi digunakan sebagai akar dalam pembentukan model *Decision Tree*.

Dengan menggunakan *Information Gain*, proses pemilihan atribut menjadi lebih objektif karena didasarkan pada perhitungan matematis. Hal ini membantu menghasilkan model klasifikasi yang lebih akurat dalam memprediksi tingkat kemiskinan masyarakat.

2.8 *Gain Ratio*

Gain Ratio merupakan metode yang digunakan untuk memperbaiki kelemahan dari *Information Gain* dalam pemilihan atribut pada algoritma *Decision Tree*. *Information Gain* cenderung memilih atribut yang memiliki jumlah kategori lebih banyak, sehingga dapat menyebabkan bias dalam pembentukan pohon keputusan. Oleh karena itu, digunakan *Gain Ratio* untuk menormalkan nilai *Information Gain* dengan membaginya terhadap nilai *Split Information*.

Menurut penelitian oleh *Sari dan Nugraha* (2022), *Gain Ratio* digunakan untuk mengurangi kecenderungan pemilihan atribut yang memiliki banyak nilai unik

dengan cara membandingkan nilai *Information Gain* terhadap distribusi pembagian data. Dengan demikian, atribut yang dipilih sebagai node pada *Decision Tree* merupakan atribut yang benar-benar memiliki kemampuan terbaik dalam memisahkan data.

Rumus *Gain Ratio* dapat dituliskan sebagai berikut:

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)}$$

Dimana nilai *Split Information* dihitung menggunakan rumus berikut:

$$SplitInfo(S,A) = -\sum_{i=1}^{n} \frac{|\,S_i\,|}{|\,S\,|} \log_2 \frac{|\,S_i\,|}{|\,S\,|}$$

Keterangan:

1. *GainRatio(S,A)* = nilai *Gain Ratio* atribut A
2. *Gain(S,A)* = nilai *Information Gain* atribut A
3. *SplitInfo(S,A)* = nilai pembagi atribut A
4. *Si* = subset data ke-i
5. |*Si*| = jumlah data subset ke-i
6. |*S*| = jumlah seluruh data

Dalam algoritma *Decision Tree*, setiap atribut dihitung nilai *Gain Ratio*-nya, kemudian atribut dengan nilai terbesar dipilih sebagai node terbaik. Penggunaan *Gain Ratio* menghasilkan pohon keputusan yang lebih stabil dan tidak bias terhadap atribut dengan jumlah kategori yang banyak.

Pada penelitian ini, perhitungan *Gain Ratio* digunakan untuk menentukan atribut terbaik dalam klasifikasi tingkat kemiskinan. Atribut seperti penghasilan, pekerjaan, dan pendidikan dibandingkan nilai *Gain Ratio*-nya untuk menentukan

akar (*root node*) pada model *Decision Tree*. Atribut dengan nilai *Gain Ratio* tertinggi akan menjadi faktor utama dalam pembentukan pohon keputusan.

Dengan menggunakan *Gain Ratio*, model klasifikasi yang dihasilkan menjadi lebih optimal karena pemilihan atribut dilakukan secara lebih objektif dan seimbang. Hal ini membantu meningkatkan akurasi model dalam memprediksi tingkat kemiskinan masyarakat.

2.9 *Confusion Matrix*

Confusion Matrix merupakan metode evaluasi yang digunakan untuk mengukur performa model klasifikasi dengan membandingkan hasil prediksi dengan kondisi sebenarnya. Metode ini menampilkan jumlah data yang diklasifikasikan dengan benar maupun yang salah dalam bentuk matriks. Dengan menggunakan *confusion matrix*, dapat diketahui tingkat akurasi serta kesalahan yang terjadi pada model klasifikasi.

Menurut *Gorunescu* yang dikutip dalam penelitian oleh *Saputra dan Wibowo* (2023), *confusion matrix* digunakan untuk mengevaluasi kinerja algoritma klasifikasi dengan menampilkan jumlah prediksi benar dan salah berdasarkan kategori kelas. Matriks ini sangat penting karena tidak hanya menunjukkan tingkat akurasi, tetapi juga memberikan informasi detail mengenai kesalahan klasifikasi yang terjadi.

Secara umum, *confusion matrix* terdiri dari empat komponen utama, yaitu:

1. *True Positive (TP)*

   Data yang sebenarnya termasuk kelas positif dan diprediksi sebagai kelas positif.

2. *True Negative (TN)*

   Data yang sebenarnya termasuk kelas negatif dan diprediksi sebagai kelas negatif.

3. *False Positive (FP)*

   Data yang sebenarnya termasuk kelas negatif tetapi diprediksi sebagai kelas positif.

4. *False Negative (FN)*

   Data yang sebenarnya termasuk kelas positif tetapi diprediksi sebagai kelas negatif.

Berdasarkan nilai tersebut, beberapa ukuran evaluasi dapat dihitung untuk mengetahui performa model klasifikasi. Ukuran yang umum digunakan antara lain *accuracy*, *precision*, *recall*, dan *F1-score*.

Rumus *accuracy* dapat dituliskan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Rumus *precision* adalah sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

Rumus *recall* adalah sebagai berikut:

$$Recall = \frac{TP}{TP + FN}$$

Rumus *F1-Score* adalah sebagai berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Dalam penelitian ini, *confusion matrix* digunakan untuk mengevaluasi performa algoritma *Decision Tree* dalam mengklasifikasikan tingkat kemiskinan masyarakat. Hasil evaluasi menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar serta jumlah kesalahan yang terjadi. Nilai tersebut kemudian digunakan untuk menghitung *accuracy*, *precision*, *recall*, dan *F1-score* guna mengetahui tingkat performa model.

Dengan menggunakan *confusion matrix*, performa model dapat dianalisis secara lebih rinci sehingga dapat diketahui apakah model yang dibangun sudah optimal dalam memprediksi tingkat kemiskinan.

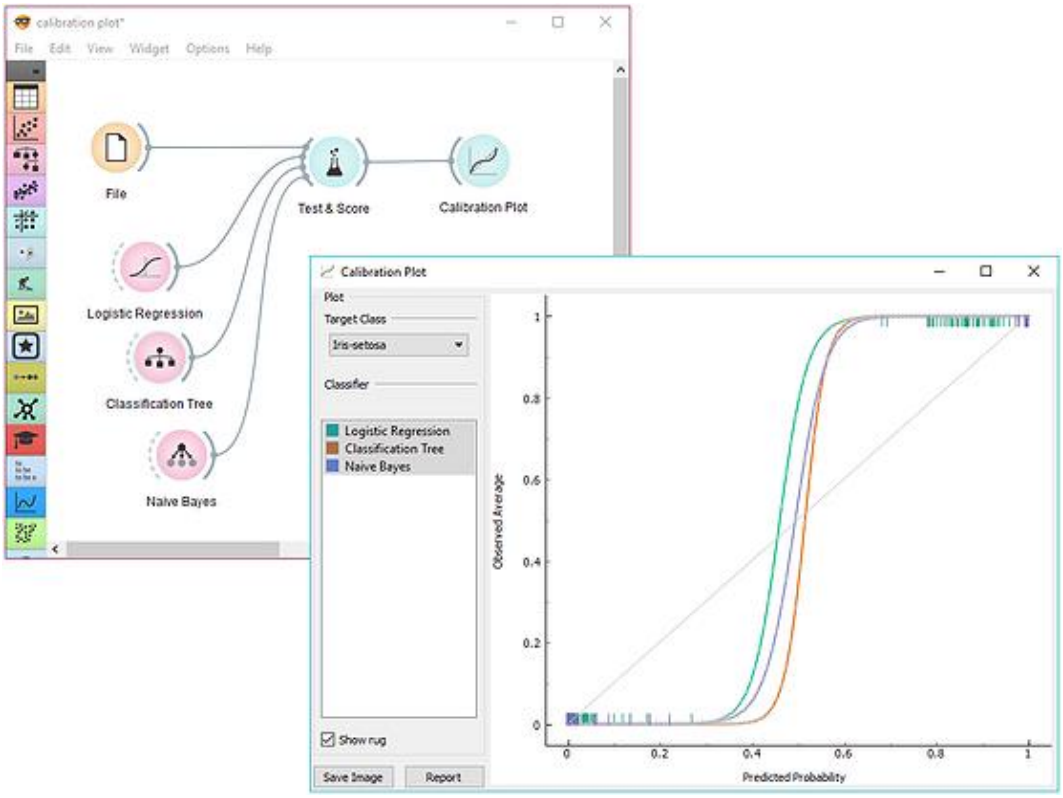
2.10 Orange Data Mining

Orange Data Mining merupakan salah satu aplikasi berbasis *open source* yang digunakan untuk melakukan analisis data berbasis *data mining* dan *machine learning*. Aplikasi ini memiliki tampilan visual berbasis *drag and drop* sehingga memudahkan pengguna dalam membangun model tanpa harus melakukan pemrograman secara langsung.

Menurut Saputra dan Hidayat (2021), *Orange Data Mining* merupakan tools yang efektif dalam proses analisis data karena menyediakan berbagai fitur untuk klasifikasi, visualisasi, serta evaluasi model secara interaktif. Hal ini menjadikan aplikasi ini cocok digunakan dalam penelitian, khususnya bagi pengguna yang ingin memahami proses *data mining* secara visual.

Dalam penelitian ini, *Orange Data Mining* digunakan sebagai alat utama dalam proses pengolahan data dan pembangunan model klasifikasi menggunakan

algoritma *Decision Tree*. Proses analisis dilakukan melalui beberapa tahapan yang disusun dalam bentuk *workflow*.



Gambar 2.4 Tampilan Orange Data Mining

Sumber: Orange Data Mining (2025)

Adapun tahapan penggunaan *Orange Data Mining* dalam penelitian ini adalah sebagai berikut:

1. **File**

   Digunakan untuk memasukkan dataset yang telah disusun sebelumnya menggunakan *Microsoft Excel* ke dalam sistem.

2. **Select Columns**

   Digunakan untuk menentukan atribut yang berperan sebagai *feature*, *target*,

dan *meta*. Pada tahap ini dilakukan pemisahan antara variabel input dan variabel output.

3. **Decision Tree**

   Digunakan untuk membangun model klasifikasi berdasarkan algoritma *Decision Tree*. Pada tahap ini sistem akan memproses data dan menghasilkan struktur pohon keputusan.

4. **Test and Score**

   Digunakan untuk melakukan pengujian model menggunakan metode *cross validation*. Hasil dari tahap ini berupa nilai evaluasi seperti akurasi, *precision, recall,* dan *F1-score*.

5. **Confusion Matrix**

   Digunakan untuk mengevaluasi performa model secara lebih rinci dengan melihat jumlah prediksi yang benar dan salah.

6. **Tree Viewer**

   Digunakan untuk menampilkan visualisasi struktur pohon keputusan yang dihasilkan, sehingga memudahkan dalam memahami aturan klasifikasi.

Keunggulan dari *Orange Data Mining* adalah kemampuannya dalam menyajikan proses analisis secara visual sehingga memudahkan pengguna dalam memahami alur *data mining*. Selain itu, aplikasi ini juga mendukung berbagai algoritma klasifikasi yang dapat digunakan sesuai dengan kebutuhan penelitian.

Dalam konteks penelitian ini, penggunaan *Orange Data Mining* sangat membantu dalam mempercepat proses analisis serta menghasilkan model klasifikasi yang mudah dipahami dan divisualisasikan. Dengan demikian, aplikasi

ini berperan penting dalam mendukung keberhasilan penelitian dalam memprediksi tingkat kemiskinan masyarakat.

2.11 Microsoft Word

Microsoft Word merupakan salah satu aplikasi pengolah kata (*word processing*) yang banyak digunakan dalam penyusunan dokumen, termasuk laporan penelitian dan skripsi. Aplikasi ini menyediakan berbagai fitur yang memudahkan pengguna dalam menulis, mengedit, serta memformat dokumen secara sistematis dan terstruktur.

Dalam penelitian ini, *Microsoft Word* digunakan sebagai alat utama dalam penyusunan laporan skripsi mulai dari Bab I hingga Bab V. Fitur-fitur seperti pengaturan paragraf, pembuatan daftar isi otomatis, penomoran halaman, serta pengelolaan sitasi sangat membantu dalam menghasilkan dokumen yang rapi dan sesuai dengan standar penulisan ilmiah.

Selain itu, *Microsoft Word* juga mendukung penggunaan *styles* yang memudahkan dalam pengaturan format judul, subjudul, serta isi teks. Dengan penggunaan fitur ini, struktur dokumen menjadi lebih konsisten dan profesional. Menurut Rahman (2021), penggunaan aplikasi pengolah kata seperti *Microsoft Word* sangat penting dalam penyusunan karya ilmiah karena mampu meningkatkan efisiensi dan kerapian dokumen.

Dengan demikian, *Microsoft Word* berperan sebagai alat pendukung dalam dokumentasi hasil penelitian sehingga laporan yang dihasilkan dapat disusun secara sistematis, mudah dibaca, dan sesuai dengan kaidah penulisan ilmiah.

2.12 Microsoft Excel

Microsoft Excel merupakan aplikasi pengolah data berbasis spreadsheet yang digunakan untuk melakukan perhitungan, pengolahan data, serta penyajian data dalam bentuk tabel dan grafik. Aplikasi ini banyak digunakan dalam berbagai bidang, termasuk penelitian, karena kemampuannya dalam mengelola data secara cepat dan akurat.

Dalam penelitian ini, *Microsoft Excel* digunakan untuk menyusun dataset yang berisi data simulasi kondisi sosial ekonomi masyarakat. Dataset tersebut terdiri dari 100 data kepala keluarga dengan atribut seperti penghasilan, pekerjaan, pendidikan, kepemilikan rumah, dan jenis lantai.

Selain itu, *Microsoft Excel* juga digunakan dalam tahap awal *preprocessing*, seperti pengecekan data, penghapusan data yang tidak relevan, serta penyeragaman format data. Penggunaan fungsi-fungsi dasar dalam *Excel* membantu dalam memastikan bahwa data yang digunakan sudah bersih dan siap untuk diproses lebih lanjut.

Menurut Putra dan Santoso (2022), *Microsoft Excel* memiliki keunggulan dalam pengolahan data karena menyediakan berbagai fungsi perhitungan serta kemampuan dalam mengelola data dalam jumlah besar secara efisien. Hal ini menjadikan *Excel* sebagai salah satu tools yang penting dalam tahap awal penelitian berbasis *data mining*.

Setelah dataset selesai disusun di *Microsoft Excel*, data tersebut kemudian diimpor ke dalam aplikasi *Orange Data Mining* untuk dilakukan proses klasifikasi menggunakan algoritma *Decision Tree*. Dengan demikian, *Microsoft Excel*

berperan sebagai alat bantu dalam pengolahan data awal sebelum dilakukan proses analisis lebih lanjut menggunakan metode *data mining*.

2.13 Profil Kelurahan Bakaran Batu

Kelurahan Bakaran Batu merupakan salah satu wilayah administratif yang berada di Kecamatan Rantau Selatan, Kabupaten Labuhanbatu, Provinsi Sumatera Utara. Wilayah ini menjadi salah satu pusat aktivitas masyarakat dengan karakteristik sosial ekonomi yang cukup beragam, sehingga relevan dijadikan sebagai objek penelitian dalam analisis tingkat kemiskinan berbasis *data mining*. Secara geografis, Kelurahan Bakaran Batu terletak di wilayah Kecamatan Rantau Selatan yang merupakan bagian dari Kabupaten Labuhanbatu dengan posisi yang cukup strategis karena berada di kawasan perkotaan yang menjadi pusat kegiatan ekonomi dan pemerintahan. Secara umum, wilayah ini berbatasan dengan kelurahan lain di Kecamatan Rantau Selatan di sebelah utara, Kecamatan Rantau Utara di sebelah selatan, kawasan pemukiman dan lahan masyarakat di sebelah timur, serta pusat aktivitas ekonomi dan perdagangan di sebelah barat. Letak yang strategis ini menjadikan Kelurahan Bakaran Batu memiliki akses yang cukup baik terhadap berbagai fasilitas umum seperti pendidikan, kesehatan, dan perdagangan.
Kondisi demografis masyarakat di Kelurahan Bakaran Batu terdiri dari berbagai kelompok usia, tingkat pendidikan, serta latar belakang pekerjaan. Masyarakat di wilayah ini didominasi oleh usia produktif yang sebagian besar bekerja di sektor informal maupun formal. Tingkat pendidikan masyarakat bervariasi mulai dari sekolah dasar hingga perguruan tinggi, dengan mata pencaharian yang beragam seperti buruh, petani, wiraswasta, dan pegawai negeri. Selain itu, komposisi

keluarga juga berbeda-beda berdasarkan jumlah tanggungan, yang secara tidak langsung mempengaruhi kondisi ekonomi rumah tangga. Keberagaman kondisi demografis ini menjadi faktor penting dalam analisis tingkat kemiskinan karena setiap variabel memiliki kontribusi terhadap kesejahteraan masyarakat.

Kondisi sosial ekonomi masyarakat di Kelurahan Bakaran Batu menunjukkan adanya perbedaan tingkat kesejahteraan antar rumah tangga. Sebagian masyarakat memiliki penghasilan tetap dengan kondisi ekonomi yang stabil, sementara sebagian lainnya masih berada pada kategori ekonomi menengah ke bawah. Faktor-faktor yang mempengaruhi kondisi tersebut antara lain tingkat penghasilan kepala keluarga, jenis pekerjaan, tingkat pendidikan, kepemilikan aset seperti rumah, serta kondisi tempat tinggal. Indikator-indikator ini sejalan dengan konsep yang digunakan oleh Badan Pusat Statistik (2024) dalam menentukan tingkat kemiskinan suatu wilayah, sehingga variabel-variabel tersebut digunakan dalam penelitian ini sebagai dasar dalam proses klasifikasi menggunakan metode *data mining*.

Selain itu, Kelurahan Bakaran Batu juga memiliki berbagai sarana dan prasarana yang mendukung aktivitas masyarakat, seperti fasilitas pendidikan mulai dari sekolah dasar hingga menengah, fasilitas kesehatan seperti puskesmas dan klinik, sarana ibadah, pasar, serta akses jalan yang cukup memadai. Ketersediaan fasilitas ini berpengaruh terhadap tingkat kesejahteraan masyarakat, terutama dalam hal akses terhadap layanan dasar. Dari segi perumahan, kondisi rumah masyarakat cukup beragam, mulai dari rumah permanen hingga semi permanen, dengan status kepemilikan yang bervariasi seperti milik sendiri maupun kontrak. Kondisi fisik

rumah, termasuk jenis lantai seperti semen atau keramik, juga menjadi indikator penting dalam menentukan tingkat kesejahteraan masyarakat.

Kelurahan Bakaran Batu memiliki potensi berupa letak yang strategis sebagai pusat aktivitas ekonomi, ketersediaan tenaga kerja dari masyarakat usia produktif, serta akses terhadap fasilitas umum yang cukup baik. Namun demikian, wilayah ini juga menghadapi beberapa permasalahan seperti ketimpangan tingkat penghasilan masyarakat, masih adanya tingkat pendidikan yang rendah, serta ketidakstabilan pekerjaan pada sektor informal. Permasalahan-permasalahan tersebut menjadi faktor yang berkontribusi terhadap munculnya tingkat kemiskinan di wilayah ini.

Kelurahan Bakaran Batu dipilih sebagai objek penelitian karena memiliki karakteristik yang sesuai dengan tujuan penelitian, yaitu menganalisis dan memprediksi tingkat kemiskinan masyarakat. Variasi kondisi sosial ekonomi yang ada memungkinkan penerapan metode *data mining* untuk menemukan pola yang mempengaruhi tingkat kemiskinan. Dalam penelitian ini, data yang digunakan merupakan data simulasi yang merepresentasikan kondisi masyarakat di wilayah tersebut dengan variabel seperti penghasilan, pekerjaan, pendidikan kepala keluarga, kepemilikan rumah, dan jenis lantai. Dengan adanya profil wilayah ini, diharapkan pembaca dapat memahami konteks penelitian secara lebih jelas serta mengetahui alasan pemilihan lokasi penelitian dalam penerapan algoritma *Decision Tree*.

Gambar 2.5 Area Kelurahan Bakaran Batu

2.14 Penelitian Terdahulu

Penelitian terdahulu digunakan sebagai acuan untuk mengetahui perbandingan metode, variabel, serta hasil yang diperoleh dari penelitian sebelumnya. Selain itu, penelitian terdahulu juga membantu memperkuat alasan penggunaan algoritma *Decision Tree* dalam klasifikasi tingkat kemiskinan.

Beberapa penelitian terdahulu yang relevan dengan penelitian ini adalah sebagai berikut:

1. Penelitian oleh Rahman dan Putri (2023)

Penelitian ini membahas penerapan algoritma *Decision Tree* untuk klasifikasi tingkat kemiskinan menggunakan variabel penghasilan, pekerjaan, dan pendidikan. Metode yang digunakan adalah *data mining* dengan algoritma *Decision Tree* serta evaluasi menggunakan *confusion matrix*. Hasil penelitian menunjukkan bahwa

model yang dihasilkan memiliki tingkat akurasi sebesar 96% dalam mengklasifikasikan data kemiskinan. Penelitian ini menunjukkan bahwa algoritma *Decision Tree* efektif digunakan dalam klasifikasi tingkat kemiskinan berdasarkan indikator sosial ekonomi.

2. Penelitian oleh Sari dan Nugroho (2022)

Penelitian ini menerapkan algoritma *Decision Tree* untuk mengklasifikasikan tingkat kesejahteraan masyarakat. Variabel yang digunakan meliputi penghasilan, pendidikan, pekerjaan, dan kondisi tempat tinggal. Pengujian dilakukan menggunakan metode *cross validation*. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* mampu menghasilkan model klasifikasi yang mudah dipahami dengan tingkat akurasi sebesar 94%. Penelitian ini menyimpulkan bahwa metode *Decision Tree* dapat digunakan sebagai alat bantu dalam menentukan tingkat kemiskinan masyarakat.

3. Penelitian oleh Pratama dan Wibowo (2024)

Penelitian ini menggunakan algoritma *Decision Tree* untuk klasifikasi penerima bantuan sosial. Data yang digunakan berupa data masyarakat dengan atribut penghasilan, jumlah tanggungan, pekerjaan, dan pendidikan. Hasil evaluasi menggunakan *confusion matrix* menunjukkan nilai akurasi sebesar 97%. Penelitian ini menyatakan bahwa algoritma *Decision Tree* mampu membantu proses pengambilan keputusan dalam menentukan penerima bantuan sosial secara lebih objektif.

4. Penelitian oleh Saputra dan Hidayat (2021)

Penelitian ini membandingkan algoritma *Decision Tree* dengan metode klasifikasi lainnya dalam menentukan tingkat kemiskinan. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* memiliki keunggulan dalam hal interpretasi model dan kemudahan visualisasi dibandingkan metode lainnya. Nilai akurasi yang diperoleh mencapai 95% sehingga metode ini dinilai cukup efektif dalam klasifikasi data sosial ekonomi masyarakat.

Berdasarkan beberapa penelitian terdahulu tersebut, dapat disimpulkan bahwa algoritma *Decision Tree* memiliki performa yang baik dalam melakukan klasifikasi tingkat kemiskinan. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada penggunaan dataset simulasi masyarakat Kelurahan Bakaran Batu serta variabel yang digunakan yaitu penghasilan, pekerjaan, pendidikan kepala keluarga, kepemilikan rumah, dan jenis lantai. Selain itu, penelitian ini juga menggunakan aplikasi *Orange Data Mining* untuk membangun model klasifikasi dan melakukan evaluasi menggunakan *confusion matrix*.

Dengan demikian, penelitian ini diharapkan dapat melengkapi penelitian sebelumnya serta memberikan hasil klasifikasi tingkat kemiskinan yang lebih mudah dipahami melalui model pohon keputusan.