

**IDENTIFIKASI INDIKATOR PENYEBAB KEMISKINAN  
DI INDONESIA DENGAN ALGORITMA  
*DECISION TREE C4.5***

**SKRIPSI**

**BELLA FRANSISKA REJEKI SIMAMORA  
191402036**



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2024**

**IDENTIFIKASI INDIKATOR PENYEBAB KEMISKINAN DI INDONESIA  
DENGAN ALGORITMA *DECISION TREE C4.5***

**Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah  
Sarjana Teknologi Informasi**

**SKRIPSI**

**BELLA FRANSISKA REJEKI SIMAMORA**

**191402036**



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2024**

**PERSETUJUAN**

Judul : Identifikasi Indikator Penyebab Kemiskinan di  
Indonesia dengan Algoritma *Decision Tree C4.5*  
Kategori : Skripsi  
Nama : Bella Fransiska Rejeki Simamora  
Nomor Induk Mahasiswa : 191402036  
Program : Sarjana (S1) Teknologi Informasi  
Fakultas : Ilmu Komputer dan Teknologi Informasi

Medan, 11 Juli 2024

Komisi Pembimbing:

Pembimbing 2,



Ivan Jaya S.Si., M.Kom.  
NIP. 198407072015041001

Pembimbing 1,



Dedy Arisandi S.T., M.Kom.  
NIP. 197908312009121002

Diketahui/disetujui oleh  
Program Studi Teknologi Informasi  
Ketua,



Dedy Arisandi S.T., M.Kom.  
NIP. 197908312009121002

**PERNYATAAN****IDENTIFIKASI INDIKATOR PENYEBAB KEMISKINAN DI INDONESIA  
DENGAN ALGORITMA *DECISION TREE C4.5*****SKRIPSI**

Saya menyatakan bahwasanya skripsi ini merupakan hasil karya saya sendiri,  
terkecuali beberapa ringkasan maupun kutipan yang telah disebutkan sumbernya.

Medan, 1 Mei 2024

Bella Fransiska Rejeki Simamora  
191402036

## UCAPAN TERIMA KASIH

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, atas berkat-Nya yang luar biasa dalam menuntun langkah penulis dalam menyelesaikan skripsi yang berjudul “Identifikasi Indikator Penyebab Kemiskinan di Indonesia dengan Algoritma *Decision Tree* C4.5”, yang dimana menjadi syarat dalam memperoleh gelar sarjana komputer di Program Studi Teknologi Informasi Universitas Sumatera Utara.

Dengan penuh syukur, Penulis menyampaikan rasa terima kasih sedalamnya kepada seluruh pihak yang memberikan dukungan, bantuan, dan motivasi dalam proses penyelesaian skripsi ini. Penulis mengucapkan terima kasih kepada:

1. Tuhan Yesus Kristus yang selalu memberikan berkat dan kemampuan kepada penulis sehingga mampu menyelesaikan tanggung jawab hingga tahap akhir penulisan.
2. Keluarga penulis, Bapak Beriman Simamora, Kakak tercinta Brenda Agnes Novianty Simamora, kedua Adik penulis Brendan Natanael Pangihutan Simamora dan Bertrand Mangantar Bungaran yang selalu memberikan doa dan dukungan dan semangat sehingga akhirnya penulis dapat menyelesaikan skripsi ini.
3. Ibu Dr. Maya Silvi Lydia, M.Sc selaku Dekan Fasilkom-TI USU.
4. Bapak Dedy Arisandy S.T., M.Kom. selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan Dosen Pembimbing I yang telah membimbing, memberikan saran dan masukan kepada penulis hingga akhir penulisan.
5. Bapak Ivan Jaya, S.Si., M.Kom. selaku Sekretaris Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan Dosen Pembimbing II yang juga telah membimbing, memberikan saran dan masukan kepada penulis hingga akhir penulisan .
6. Bapak Dr. Romi Fadillah Rahmat B.Comp.Sc., M.Sc. dan Bapak Dr. Sawaluddin, M.IT. selaku Dosen Pembimbing penulis yang telah memberikan banyak saran dan masukan kepada penulis hingga akhir penulisan.
7. Dosen Pengajar di Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.

8. Staff dan pegawai di Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.
9. Teman-teman seperjuangan penulis dalam masa perkuliahan Ineztri Situmeang, Siti Fadia Al Maswin, Annisa Putri Daulay, Huzaifah Lais dan Arsyia Fikri.
10. Sahabat penulis yang selalu memberikan dukungan dan semangat dalam penyelesaian tugas akhir ini, Rio Fernando Alexander, Frans Rivaldo, Paulina Novia, Rahel Permata Indah, Finer Mayland, Yuni Angelina Sitohang, Tarisa Ivanka dan Gusri Ramit Yora.
11. Teman-teman angkatan 2019 Teknologi Informasi Universitas Sumatera Utara yang memberikan dukungan dan motivasi kepada penulis.
12. Seluruh pihak yang tidak dapat penulis sebut yang telah memberikan bantuan, perhatian, semangat dan doa kepada penulis sehingga mampu menyelesaikan skripsi ini. Penulis mengucapkan terima kasih.

Medan, 1 Mei 2024

Bella Fransiska Rejeki Simamora

## IDENTIFIKASI INDIKATOR PENYEBAB KEMISKINAN DI INDONESIA DENGAN ALGORITMA *DECISION TREE C4.5*

### ABSTRAK

Saat ini kemiskinan masih menjadi masalah utama bagi seluruh negara di dunia. Hingga sampai saat ini, hampir keseluruhan negara di dunia masih dihadapkan dengan kemiskinan, begitupun dengan Indonesia. Dalam menanggapi masalah kemiskinan ini, menjadi kepentingan sebuah negara dalam mengembangkan sistem yang mampu untuk mencari indikator penyebab kemiskinan, agar pengentasan kemiskinan berdasarkan sumber daya yang dimiliki bisa dilakukan. Sehingga, penelitian ini menghasilkan sistem yang dapat memprediksi status kemiskinan dari setiap Kabupaten/Kota di Indonesia berdasarkan indikator-indikator yang diprediksi menjadi penyebab miskin tidaknya daerah tersebut. Status Kemiskinan dikelompokkan menjadi 2 kelas yaitu Miskin dan Tidak Miskin. Adapun data yang digunakan berjumlah 1028 data yang kemudian dibagi menjadi 514 *data training* dan 514 *data testing*. Data melewati tahap *preprocessing* yang terdiri dari *data splitting*, *handling missing value*, *data transformation*, *data binning* dan *encoding categorical variables* guna memaksimalkan data agar siap diolah oleh Model. Algoritma yang digunakan adalah *Decision Tree C4.5*. Berdasarkan hasil data latih, dapat diketahui bahwa nilai akurasi dari model *Decision Tree C4.5* mencapai 92.02%. Berdasarkan Evaluasi *Permutation Matrix* guna menentukan indikator penyebab kemiskinan, didapat bahwa PDRB per Kapita, Pengeluaran per Kapita, Jumlah Desa yang Memiliki Lembaga Perkreditan, Tingkat Pengangguran Terbuka (TPT) dan Rata-rata Lama Sekolah menjadi variabel paling berkontribusi dalam pemodelan ini.

**Kata Kunci :** Kemiskinan, PDRB, *Decision Tree*, Klasifikasi, *Permutation Matrix*.

**IDENTIFICATION OF POVERTY-CAUSING INDICATORS IN INDONESIA  
USING THE DECISION TREE C4.5 ALGORITHM**

**ABSTRACT**

*Poverty remains a significant issue for countries worldwide. To this day, nearly every nation, including Indonesia, grapples with this persistent challenge. In addressing poverty, it is essential for a country to develop a system capable of identifying the indicators that cause poverty, thereby enabling targeted poverty alleviation based on available resources. Consequently, this research aims to develop a system that can predict the poverty status of each Regency/City in Indonesia based on indicators hypothesized to influence poverty. Poverty status is classified into two categories: Poor and Not Poor. The dataset comprises 1,028 records, which are split into 514 training data and 514 testing data. The data undergoes preprocessing stages, including data splitting, handling missing values, data transformation, data binning, and encoding categorical variables, to ensure optimal preparation for model processing. The algorithm used is Decision Tree C4.5. The training data results indicate that the Decision Tree C4.5 model achieves an accuracy of 92.02%. Based on the Permutation Matrix evaluation to identify the indicators contributing to poverty, it is found that GDP per Capita, Expenditure per Capita, the Number of Villages with Credit Institutions, Open Unemployment Rate (TPT), and Average Years of Schooling are the most significant variables in this model.*

**Keywords:** *Poverty, GDP, Decision Tree, Classification, Permutation Matrix.*



## DAFTAR ISI

PERSETUJUAN	ii
PERNYATAAN	iii
UCAPAN TERIMA KASIH	iv
ABSTRAK	vi
<i>ABSTRACT</i>	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah	3
1.5. Manfaat Penelitian	4
1.6. Metodologi Penelitian	4
1.7. Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	7
2.1. <i>Data Mining</i>	7
2.2. <i>Algoritma Decision Tree C4.5</i>	8
2.3. Kemiskinan	10
2.4. Faktor Penyebab Kemiskinan	11
2.4.1. Faktor Pekerjaan	11
2.4.2. Faktor Pendidikan	11
2.4.3. Faktor Gender	12
2.4.4. Faktor Akses terhadap Pelayanan dan Infrastruktur Dasar	12
2.4.5. Faktor Lokasi Geografis	13
2.4.6. Faktor Kesehatan	14
2.4.7. Faktor Pendapatan Daerah	15
2.5. Penelitian Terdahulu	15
2.6. Perbedaan Penelitian	19
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	20
3.1. <i>Dataset</i>	20

3.2.	Arsitektur Umum	23
3.2.1.	Input Data	24
3.2.2.	<i>Pre-processing</i>	27
3.2.3.	<i>Training Model Building</i>	31
3.2.4.	<i>Testing Model Building</i>	37
3.2.5.	<i>Output</i>	37
3.3.	Perancangan Aplikasi Sistem	37
3.3.1.	Gambaran tampilan <i>homepage</i>	38
3.3.2.	Gambaran tampilan halaman <i>training</i>	39
3.3.3.	Gambaran tampilan halaman <i>testing</i>	39
3.4.	Metode Evaluasi	40
3.4.1.	<i>K-Fold Cross Validation</i>	40
3.4.2.	<i>Confusion Matrix</i>	41
3.4.3.	<i>Permutation Importance</i>	42
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM		43
4.1.	Implementasi Sistem	43
4.1.1.	Spesifikasi Perangkat Keras dan Perangkat Lunak	43
4.1.2.	Penerapan Perancangan Tampilan Antarmuka	43
4.2.	Evaluasi <i>Training Model</i>	47
4.3.	Evaluasi <i>Testing Model</i>	49
BAB 5 KESIMPULAN DAN SARAN		52
5.1.	Kesimpulan	52
5.2.	Saran	52
DAFTAR PUSTAKA		53
LAMPIRAN		55

## DAFTAR GAMBAR

Gambar 3. 1 Arsitektur Umum	24
Gambar 3. 2 <i>Flowchart</i> tahapan <i>Pre-processing</i>	27
Gambar 3. 3 Gambaran <i>Homepage</i>	38
Gambar 3. 4 Gambaran tampilan <i>training data</i>	39
Gambar 3. 5 Gambaran tampilan <i>testing data</i>	40
Gambar 4. 1 Tampilan <i>Homepage</i>	44
Gambar 4. 2 Tampilan <i>Halaman Training</i>	45
Gambar 4. 3 Tampilan Halaman Hasil <i>Training</i>	45
Gambar 4. 4 Tampilan Halaman <i>Testing</i>	46
Gambar 4. 5 Tampilan Halaman <i>Hasil Testing</i>	47
Gambar 4. 6 <i>Permutation Importance Training Model</i>	49
Gambar 4. 7 <i>Permutation Importance Testing Model</i>	51

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	16
Tabel 3. 1 Data hasil <i>scrapping</i> pada indikator penyebab kemiskinan	20
Tabel 3. 2 Deskripsi Indikator Penyebab Kemiskinan	21
Tabel 3. 3 Variabel Rata-rata Lama Sekolah	25
Tabel 3. 4 Variabel Tingkat Pengangguran Terbuka	25
Tabel 3. 5 Variabel Pengeluaran per Kapita	25
Tabel 3. 6 Variabel PDRB per Kapita	25
Tabel 3. 7 Variabel Angka Harapan Hidup	25
Tabel 3. 8 Variabel Rumah Tangga dengan Air Minum Layak	26
Tabel 3. 9 Variabel Rumah Tangga dengan Penggunaan Listrik PLN	26
Tabel 3. 10 Variabel Rumah Tangga dengan Sanitasi layak	26
Tabel 3. 11 Variabel Jumlah Desa yang Memiliki Lembaga Keterampilan	26
Tabel 3. 12 Variabel Jumlah Desa yang Memiliki Lembaga Perkreditan	26
Tabel 3. 13 Variabel Jumlah Desa yang Memiliki Sekolah Dasar	26
Tabel 3. 14 Variabel Pekerja Tani Sektor Informal	27
Tabel 3. 15 <i>Data Splitting</i>	28
Tabel 3. 16 Tahap <i>Handling Missing Values</i>	28
Tabel 3. 17 Tahap <i>Data Transformation</i>	29
Tabel 3. 18 Tahap <i>Data Binning</i>	30
Tabel 3. 19 <i>Tahap Encoding Categorical Variables</i>	31
Tabel 3. 20 Agresi Dataset Penyebab Kemiskinan	31
Tabel 3. 21 Sampel Perhitungan <i>Entropy</i> Rata-rata Lama Sekolah	34
Tabel 3. 22 Perhitungan <i>Entropy</i> dan <i>Gain</i> untuk Node Akar	34
Tabel 3. 23 Implementasi Evaluasi <i>Confusion Matrix</i>	41
Tabel 4. 1 Hasil <i>Confusion Matrix</i>	48
Tabel 4. 2 Contoh Hasil <i>Testing Model</i>	49

## **BAB 1**

### **PENDAHULUAN**

#### **1.1. Latar Belakang**

Saat ini kemiskinan menjadi permasalahan utama keseluruhan negara di dunia, dimana seluruh keseluruhan negara di dunia masih dihadapkan dengan kemiskinan, begitupun dengan Indonesia. Dalam menanggapi masalah kemiskinan ini, banyak negara berkomitmen untuk mengurangi kemiskinan dengan menjadikannya sebagai salah satu tujuan dalam Sustainable Development Goals (SDGs). Pada tahun 2015, tercatat lebih dari 736 juta jiwa yang dimana setara dengan sepuluh persen penduduk memiliki kebutuhan biaya hidup kurang dari \$1.90 per hari, dimana banyak diantaranya yang memiliki masalah kurangnya makanan, air bersih dan juga sanitasi. Namun proses pengentasan kemiskinan dalam skala internasional yang tidak merata di setiap negara menyebabkan penyelesaian kemiskinan ini sangat sulit untuk diselesaikan.

Menurut *World Bank* (2000), kemiskinan merupakan suatu ketidakcukupan kesejahteraan. Adapun standar yang ditentukan oleh *World Bank* dalam mengukur kemiskinan adalah pendapatan yang kurang dari \$2 per hari. Saat ini, nilai yang digunakan oleh *World Bank* dan BPS dalam menentukan kemiskinan memiliki perbedaan cara hitung. Badan Pusat Statistik (BPS) menggunakan konsep kemampuan dalam mencukupi kebutuhan dasar (*basic needs method*) dalam menghitung kemiskinan di Indonesia. Berdasarkan konsep ini, kemiskinan didefinisikan sebagai ketidakmampuan dalam mencukupi kebutuhan dasar makanan dan non-makanan dari segi ekonomi. Dalam hal ini, adapun bahan yang digunakan BPS adalah data dari Survei Sosial Ekonomi Nasional (SUSENAS) Konsumsi dan Pengeluaran, SUSENAS KOR, serta Survei Paket Komoditi Kebutuhan Dasar.

Saat ini Kemiskinan di Indonesia menunjukkan tren yang menurun dan melambat, sesuai dengan yang telah didefinisikan oleh *World Bank* (2018). Tercatat bahwasanya pemerintah belum mampu untuk memenuhi Target Kerja Pemerintah setiap tahunnya. Terdata pada Maret 2020 persentase penduduk turun meski hanya menjadi 9,78 persen.

Berdasarkan Rencana Kerja Pemerintah Tahun 2022, saat ini pemerintah menargetkan kemiskinan mencapai angka 8,5-9,0 persen.

Masalah selanjutnya yakni solusi dalam pengentasan kemiskinan yang dilakukan di Indonesia cenderung tidak merata. Salah satu masalah utama yang saat ini ada di Indonesia adalah rendahnya tingkat akurasi atau ketepatan sasaran (tingginya *inclusion* dan *exclusion error*) penerima program perlindungan sosial. Sehingga, perlunya ditetapkan standar kemiskinan yang tepat dalam mengambil keputusan.

Masalah kemiskinan ini memanglah sebuah masalah mendasar yang kompleks dan bersifat multidimensional (Badan Pusat Statistik, 2020). Sifat multidimensional ini menyebabkan kemiskinan bisa dilihat dari berbagai faktor, sehingga indikator yang bisa digunakan dalam pengukurannya sangatlah banyak. Dari kumpulan indikator yang banyak tersebut, berpotensi memiliki indikator paling berpengaruh yang mampu menjelaskan kemiskinan. Salah satu cara untuk menemukan indikator tersebut adalah dengan melakukan klasifikasi. Pada penelitian ini diharapkan mampu untuk menemukan indikator-indikator penyebab kemiskinan yang paling berpengaruh.

Klasifikasi adalah proses analisa sebuah data dalam menentukan model yang dapat menguraikan atau mengelompokkan data-data berdasarkan kelas yang digunakan memprediksi kelas berdasarkan objek lain yang tidak diketahui kelasnya. Berbagai metode klasifikasi telah digunakan oleh peneliti sebelumnya, salah satunya merupakan metode *Decision Tree*. *Decision tree* merupakan salah satu algoritma *machine learning* dengan metode klasifikasi yang bersifat *supervised learning* menggunakan pembentukan dari pohon keputusan yang diproses data yang ada. Terdapat berbagai jenis algoritma yang menggunakan klasifikasi *Decision Tree*, antara lain *Random Forest*, *Boosted Trees*, *ID3*, *Rotation Forest*, *C4.5*, *QUEST*, *Chi-Square Automatic Interaction Detector (CHAID)*, *CRUISE*, dan *MARS* (Kokakoc dan Keser, 2019). Salah satu algoritma yang paling dikenal adalah *C4.5*.

Penelitian sebelumnya menunjukkan efektivitas algoritma *C4.5* dalam berbagai konteks. Penelitian oleh Ramadhani et al. (2022) menggunakan Algoritma *C4.5* untuk menentukan kelayakan penerima bantuan sosial di Kantor Kelurahan Martoba, dengan hasil yang efektif dalam menyelesaikan masalah penentuan kelayakan penerima bantuan sosial. Sementara itu, penelitian oleh Abu et al. (2020) melakukan klasifikasi

kemiskinan multidimensi di Malaysia dengan Random Forest dan Decision Tree C4.5, menemukan bahwa C4.5 memiliki interpretasi hasil yang cukup mudah, penghilangan fitur yang kurang penting dan juga efisiensi yang lebih kompleks dibandingkan model lainnya. Kelebihan utamanya meliputi kemampuannya menangani atribut kontinu dan data hilang, serta melakukan pemangkasan pohon untuk mengurangi *overfitting*. Algoritma C4.5 juga mendukung klasifikasi multi-kelas dan menghasilkan model yang mudah diinterpretasikan, menjadikannya pilihan populer dalam berbagai aplikasi klasifikasi.

Dengan didasari oleh latar belakang tersebut, penulis melakukan sebuah penelitian pengklasifikasian seluruh kabupaten di Indonesia berdasarkan status kemiskinannya dan juga untuk menentukan indikator-indikator yang menjadi penyebab utama di Indonesia. Model klasifikasi ini juga digunakan untuk memprediksi kemiskinan di suatu daerah berdasarkan data pemerintah dan juga guna mengevaluasi ketepatan sasaran program pemerintah dalam pengentasan kemiskinan di tahun bersangkutan. Penelitian ini diberi judul **“Identifikasi Indikator Penyebab Kemiskinan di Indonesia dengan Algoritma Decision Tree C4.5.”**

## **1.2. Rumusan Masalah**

Dalam menciptakan pelaksanaan strategi dan kebijakan yang baik dan peningkatan kualitas aparat pemerintah, dirasa perlu untuk menetapkan perhitungan dalam menggambarkan kemiskinan yang ada di Indonesia secara terbaik. Maka dari itu, rumusan masalah pada penelitian ini yaitu pengidentifikasian indikator-indikator yang mampu menjelaskan kemiskinan secara terbaik dengan menggunakan algoritma boosted C4.5.

## **1.3. Tujuan Penelitian**

Tujuan penelitian ini yakni untuk mengidentifikasi indikator yang paling memberikan pengaruh dalam menggambarkan kemiskinan di Indonesia dengan menggunakan algoritma *Decision Tree C4.5*.

## **1.4. Batasan Masalah**

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan data yang terdiri atas data 514 Kabupaten / Kota di Indonesia pada tahun 2021 dan tahun 2022.
2. Penelitian ini memfokuskan pada 12 Indikator, yaitu Rata-rata Lama Sekolah, Tingkat Pengangguran Terbuka, Pengeluaran per Kapita, PDRB per Kapita, Angka Harapan Hidup, Rumah Tangga dengan Air Minum Layak, Rumah Tangga dengan Penggunaan Listrik PLN, Rumah Tangga dengan Anitasi layak, Desa dengan Lembaga Keterampilan, Desa dengan Lembaga Perkreditan, Desa dengan Sekolah Dasar dan Pekerja Tani Sektor Informal.
3. Penelitian ini tidak mempertimbangkan dampak dan faktor eksternal lainnya (seperti bencana alam atau perubahan harga komoditas global) yang mungkin mempengaruhi tingkat kemiskinan pada periode data. Fokus analisis hanya pada nilai indikator-indikator yang bersumber dari BPS, Survei Sosial Ekonomi Nasional dan Potensi Desa.
4. *Output* dari penelitian ini adalah sistem berbasis web yang memberikan klasifikasi kemiskinan dan identifikasi 5 Indikator penyebab yang paling berpengaruh.

### **1.5. Manfaat Penelitian**

Manfaat yang didapat dari penelitian ini adalah sebagai berikut:

1. Mengidentifikasi dan memberikan pemahaman yang lebih baik mengenai indikator-indikator yang mempengaruhi dan menyebabkan kemiskinan di Indonesia.
2. Sebagai masukan bagi pemerintah dan lembaga terkait untuk pengambilan kebijakan yang lebih tepat sasaran dalam upaya pengentasan kemiskinan.
3. Mempelajari dan mengimplementasikan Algoritma *Decision Tree* C4.5 untuk mengklasifikasikan tingkat kemiskinan serta mengidentifikasi indikator-indikator yang paling berpengaruh.

### **1.6. Metodologi Penelitian**

Adapun tahapan yang akan dilakukan pada penelitian ini adalah :

#### **1. Studi Literatur**

Tahap ini merupakan proses pengumpulan data dan referensi yang dimana berkaitan dengan *data mining* dan algoritma *Decision Tree* C4.5, kemiskinan di



Indonesia serta pengumpulan data sekunder variabel-variabel penyebab kemiskinan di Indonesia dari jurnal, skripsi, buku, artikel, dan sumber lainnya.

## **2. Analisis Permasalahan dan Perancangan Sistem**

Akan dilakukan analisis berdasarkan informasi dan referensi yang sudah dikumpulkan guna mendapatkan pemahaman mengenai algoritma *Decision Tree C4.5*. Berdasarkan analisis yang telah dilakukan pada tahap sebelumnya, dilakukan perancangan sistem berupa perancangan arsitektur umum, pengolahan data dan perancangan antarmuka. Hasil pembelajaran literatur diimplementasikan dalam proses perancangan sistem.

## **3. Implementasi**

Pada tahap ini, perancangan sistem yang sudah dibuat pada tahap sebelumnya diimplementasikan sehingga menghasilkan sebuah sistem yang sesuai dengan tujuan penelitian.

## **4. Pengujian Sistem**

Pada tahap ini, akan dilakukan proses pengujian pada sistem yang telah dibuat serta akan memastikan sistem tersebut sudah dapat berjalan dan memiliki nilai akurasi sebagaimana yang diharapkan.

## **5. Penyusunan Laporan**

Pada tahap ini akan dilakukan penyusunan laporan berdasarkan keseluruhan penelitian yang menjelaskan analisis masalah, rancangan, hingga implementasi dan hasil penelitian.

### **1.7. Sistematika Penulisan**

Sistematika penulisan pada penelitian ini meliputi lima bagian yakni :

#### **Bab 1: Pendahuluan**

Pada bab satu akan menjelaskan perihal dasar dari penelitian ini yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

#### **Bab 2: Landasan Teori**

Bab dua berisikan ringkasan teori dasar yang penting guna memahami masalah dalam penelitian ini seperti Data Mining, Klasifikasi dan Algoritma *Decision Tree C4.5*,

Kemiskinan serta Indikator-indikator Penyebab Kemiskinan dan juga disertakan dengan beberapa penelitian terdahulu sebagai referensi pada penelitian ini.

### **Bab 3: Analisis dan Perancangan Sistem**

Pada bab tiga akan membahas analisis terkait seperti arsitektur umum, pemodelan sistem, tahap preprocessing data, olah latih dan pengujian data, serta penerapan Algoritma *Decision Tree* C4.5.

### **Bab 4: Implementasi dan Pengujian**

Bab empat berisikan bagaimana implementasi berdasarkan perancangan sistem yang telah disusun. Lebih lanjut, akan dilakukan evaluasi terhadap sistem tersebut guna memantau kinerja model yang telah dibangun.

### **Bab 5: Kesimpulan dan Saran**

Bab terakhir berisi ringkasan dan kesimpulan berdasarkan penelitian yang telah dilakukan, yang juga berisikan saran dan masukan guna pengembangan pada penelitian selanjutnya.

## BAB 2

### LANDASAN TEORI

#### **2.1. Data Mining**

*Data mining* merupakan proses analisa berbagai data yang berbeda guna menghasilkan informasi yang baru baik berdasarkan pola, pengetahuan guna memperkecil biaya pengeluaran serta meningkatkan keuntungan pada proses Analisa data.

Menurut Tan (2017), penggunaan dan manfaat *data mining* dapat dilihat melalui dua sisi, yakni sisi komersial dan keilmuan. Menurut sisi komersial, *data mining* dipakai guna menangani konsekuensi dari ledakan volume data yang diproses menyesuaikan dengan bagaimana penyimpanan, ekstraksi dan pemanfaatan data tersebut.

Berdasarkan penelitian oleh Rerung (2018), *Data mining* dibedakan menjadi beberapa jenis menurut tugas dan fungsi yang dilakukan, yaitu :

##### 1. Deskripsi

Deskripsi merupakan analisis sederhana guna menemukan kemungkinan dalam menggambarkan suatu pola atau kecenderungan yang terdapat pada data.

##### 2. Estimasi

Estimasi dapat dikatakan mirip dengan klasifikasi, namun variabel data yang digunakan cenderung numerik. Pada fungsi ini, peninjauan berikutnya adalah untuk menentukan estimasi nilai berdasarkan variabel target yang didasarkan oleh nilai pada variabel prediksi.

##### 3. Klasifikasi

Klasifikasi merupakan pengelompokan data ke dalam beberapa target variabel kategori berdasarkan kelompok yang dipahami oleh metode yang digunakan. Semisal, pemisahan/klasifikasi pendapatan yang dibedakan dalam tiga jenis kategori yakni pendapatan rendah, sedang dan tinggi.

##### 4. Prediksi

Sama halnya dengan estimasi dan juga klasifikasi, penambahannya adalah terdapat nilai yang diprediksi pada masa mendatang. Adapun beberapa jenis metode dan

fungsi yang terdapat pada klasifikasi dan juga estimasi, dapat juga digunakan untuk fungsi prediksi.

## 5. Klasterisasi

Klaster merupakan kumpulan *record* pada data yang memiliki sifat mirip antar satu dan lainnya dan sekaligus memiliki perbedaan dengan kumpulan yang lain. Klasterisasi tidak menjalankan fungsi lainnya seperti klasifikasi, estimasi maupun prediksi, namun klasterisasi mencoba untuk mengelompokkan seluruh data yang memiliki kesamaan/kemiripan (homogen).

## 6. Asosiasi

Asosiasi adalah fungsi guna menentukan atribut atau pola yang ada dalam satu waktu.

## 2.2. Algoritma *Decision Tree* C4.5

*Decision tree* merupakan salah satu cara untuk melakukan klasifikasi dan prediksi untuk menghasilkan pendukung keputusan dengan menggunakan struktur pohon atau hierarki. *Decision tree* mengolah data menjadi aturan-aturan keputusan yang dalam penelitian ini dapat diimplementasikan pada sistem pendukung keputusan yang akan dibuat. Menurut Frank et al., (1998), algoritma ini bekerja dari atas ke bawah dengan memeriksa setiap level atribut dan membaginya menjadi bagian terbaik dari kelas dan menangani sub-masalah yang dihasilkan dari pemisahan ini secara rekursif.

Algoritma C4.5 adalah sebuah algoritma guna menghasilkan *Decision Tree*. Algoritma C4.5 ini dikembangkan oleh Ross Quinlan. Algoritma C4.5 akan melanjutkan perhitungan Gain Ratio dimana menggunakan Entropy dan Gain yang telah dihitung sebelumnya.

Algoritma C4.5 menggunakan konsep *entropy* untuk menghitung *purity* atau kepastian. Entropy pada data sampel menunjukkan variasi dalam nilai suatu kelas. Secara umum, entropy dapat didefinisikan sebagai berikut :

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

$S$  : Dataset

$c$  : Jumlah partisi  $S$

$p_i$  : Proporsi kelas  $S_i$  dalam dataset  $S$

Algoritma ini menggunakan *entropy* untuk menghitung perubahan pada homogenitas yang menghasilkan *Information Gain*. *Information Gain* mengurangi *entropy* dari dataset asli dengan rata-rata tertimbang *entropy* dari subset-subset yang dihasilkan dari pemisahan berdasarkan atribut  $A$ . Atribut yang mengurangi *entropy* paling banyak adalah yang paling informatif.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^c \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

$S$  : Dataset

$A$  : Atribut

$c$  : Banyaknya partisi  $A$

$|S_i|$  : Banyaknya kasus pada partisi ke  $-i$

$|S|$  : Banyaknya kasus dalam  $S$

*Gain Ratio* adalah modifikasi dari perhitungan *Information Gain* yang mengatasi bias *Information Gain* terhadap atribut dengan banyak nilai unik. *Gain Ratio* mempertimbangkan jumlah dan ukuran subset yang dihasilkan oleh atribut.

$$GainRatio(S, A) = \frac{Gain(S, A)}{Split\ Information(S, A)} \quad (3)$$

Dimana *Split Information* dihitung dengan rumus sebagai berikut :

$$Split\ Information(S, A) = - \sum_{i=1}^c \frac{|S_i|}{|S|} \times \log_2\left(\frac{|S_i|}{|S|}\right) \quad (4)$$

*Split Information* mengukur *entropy* dari pemisahan dataset berdasarkan atribut  $A$ . *Gain Ratio* kemudian menghitung rasio antara *Information Gain* dan *Split Information* untuk mengurangi bias terhadap atribut dengan banyak nilai unik. Atribut dengan *Gain Ratio* tertinggi kemudian akan dipilih untuk pemisahan.

Algoritma C4.5 mempunyai fitur penting yang menjadikannya lebih baik daripada algoritma terdahulunya, yaitu C4.5. Adapun yang menjadi fitur utamanya adalah : (Quinlan, 2004)

1. Program C4.5 dikembangkan guna menganalisis *database* besar dengan puluhan hingga ratusan data numerik maupun nominal.
2. Untuk memaksimalkan tingkat interpretasi pengguna dari hasil yang disajikan, klasifikasi C4.5 ditampilkan dalam dua bentuk, yaitu dalam bentuk pohon keputusan dan seperangkat aturan *IF-then* yang lebih udah dipahami daripada jaringan saraf (*neural network*).
3. Algoritma C4.5 dapat digunakan dengan mudah dan tidak memerlukan pengetahuan yang tinggi mengenai statistik atau *machine learning*.

### 2.3. Kemiskinan

Kemiskinan merupakan suatu kondisi di mana seseorang ataupun sekelompok orang tidak memiliki kemampuan untuk memenuhi hak-hak dasar yang diperlukan guna mempertahankan dan meningkatkan kualitas hidup mereka. BPS dan beberapa negara lain menggunakan konsep pencukupan kebutuhan dasar (*basic needs approach*), yang mendefinisikan kemiskinan sebagai ketidaksanggupan dari sisi ekonomi guna mencukupi kebutuhan dasar makanan dan non-makanan (berdasarkan sisi pengeluaran).

Kemiskinan menjadi masalah yang selalu ada dalam kehidupan masyarakat. Adapun masalah kemiskinan akan selalu terjadi dalam waktu yang panjang, yang dapat disamakan dengan usia manusia yang mengalaminya. Sehingga dapat dikatakan bahwa kemiskinan adalah masalah yang bersifat global dan menjadi perhatian semua negara, walaupun dampak yang dirasakan dari kemiskinan di setiap negara tidak selalu sama (Nurwati, 2018).

Lebih lanjut, menurut penelitian yang dilakukan Ayus (2012), konsep kemiskinan dibagi ke dalam dua jenis yakni kemiskinan relatif dan kemiskinan absolut. Kemiskinan absolut merupakan kemiskinan yang di ukur berdasarkan pendapatan yang dipakai untuk memenuhi kebutuhan dasar dan fasilitas umum lainnya seperti pakaian, Kesehatan, Pendidikan dan tempat tinggal dan lainnya. Sedangkan kemiskinan relatif merupakan kemiskinan yang dikaitkan dengan pendistribusian pendapatan atau ditentukan berdasarkan pendapatan yang lebih rendah di lingkungan sekitar.

## 2.4. Faktor Penyebab Kemiskinan

Menurut Badan Pusat Statistik (2020), kemiskinan diukur menggunakan *Head Count index*, yaitu dengan melihat persentase penduduk miskin yang berada di bawah Garis Kemiskinan.

1. Indeks Kedalaman Kemiskinan (p1) mengukur seberapa besar pengeluaran penduduk miskin dibandingkan dengan tingkat kemiskinan. Semakin besar nilai indeks ini, semakin jauh rata-rata pengeluaran penduduk miskin dari Garis Kemiskinan.
2. Indeks Keparahan Kemiskinan (p2) mengukur distribusi pengeluaran di antara penduduk miskin. Semakin besar nilai indeks ini, semakin tinggi variasi pengeluaran di antara penduduk miskin.

Menurut analisis multivariat yang dilakukan oleh World Bank untuk menentukan faktor penyebab kemiskinan, aset dan akses utama pada rumah berkorelasi dengan kemiskinan di Indonesia (*World Bank*, 2006). Berdasarkan hasil analisis, disimpulkan bahwa terdapat lima faktor penyebab kemiskinan di Indonesia, yaitu:

### 2.4.1. Faktor Pekerjaan

Jumlah penduduk yang selalu meningkat memiliki keterkaitan dengan meningkatnya kebutuhan tenaga kerja. Semakin tinggi jumlah penduduk, meningkat pula jumlah tenaga kerja yang dibutuhkan. Namun, tingkat pendudukan yang tinggi namun tidak diimbangi dengan lapangan pekerjaan yang sejalan dapatlah meningkatkan jumlah pengangguran yang akan berakhir dengan meningkatnya jumlah kemiskinan di tingkat tertentu.

Menurut penelitian yang dilakukan oleh *World Bank* (2006), kepala rumah tangga dengan pekerjaan pada sektor pertanian menghabiskan konsumsi rumah tangga yang sangat rendah, yang diakibatkan oleh upah minimum yang sangat rendah jika dibandingkan dengan sektor pekerjaan lainnya. Lebih lanjut, diketahui bahwasanya banyak sekali kepala rumah tangga di Indonesia yang menggeluti bidang pertanian ini, sehingga sangat berdampak pada kemiskinan di Indonesia.

### 2.4.2. Faktor Pendidikan

Pendidikan menjadi tonggak utama dalam pembangunan masa depan suatu bangsa. Rendahnya tingkat pendidikan juga dinilai dapat menyebabkan kehancuran dari suatu

bangsa, hal ini disebabkan oleh Pendidikan menjadi sarana dalam peningkatan karakter maupun jati diri masyarakat yang ada di suatu bangsa.

Banyak orang miskin cenderung mengalami kebodohan secara sistematis. Hal ini mengingatkan banyak orang bahwa kemiskinan dapat menyebabkan kebodohan, dan kebodohan sering kali terkait erat dengan kemiskinan. Oleh karena itu, pendidikan menjadi alat penting untuk mengurangi tingkat kebodohan sekaligus kemiskinan.

Berdasarkan penelitian yang dilakukan oleh *World Bank*, dinilai terdapat 3 faktor utama yang sangat berkaitan pada dinilainya Pendidikan menjadi salah satu penyebab kemiskinan, yaitu :

1. Kemiskinan berhubungan dengan tingkat pendidikan yang tidak melampaui standarisasi.
2. Telah melalui pendidikan pada tingkat dasar dapat dinilai mampu meningkatkan kesejahteraan secara signifikan.
3. Meningkatnya pencapaian pendidikan di suatu daerah berhubungan langsung dengan baiknya pengentasan kemiskinan di daerah tersebut.

#### 2.4.3. Faktor Gender

Gender merupakan pengelompokan peran, kasta, tingkatan tanggung jawab dan pengelompokan pekerjaan antara pria dan Wanita yang sudah ditetapkan oleh kepercayaan dan juga adat istiadat pada masyarakat. Menurut Jomtien (1991), Pendidikan pada perempuan telah terbukti menjadi cara terbaik dalam pertumbuhan ekonomi yang cepat dan juga mewujudkan pembangunan berkelanjutan.

Pendapat ini juga didukung oleh *World Bank* (2000), yang mengatakan bahwasanya negadar dengan penduduk Wanita yang lebih berpendidikan, memiliki kehidupan yang bukan hanya lebih sehat, namun juga lebih kaya. Lebih lanjut, partisipasi kerja perempuan menjadi determinan penting dalam pembangunan sosial ekonomi dan pengentasan kemiskinan. Sehingga, gender menjadi salah satu faktor penting dalam nilai kemiskinan di suatu daerah.

#### 2.4.4. Faktor Akses terhadap Pelayanan dan Infrastruktur Dasar

Infrastruktur menjadi salah satu penanda dan roda penggerak pertumbuhan ekonomi di suatu negara. Rendahnya peningkatan infrastruktur menjadi salah satu faktor utama dalam pertumbuhan ekonomi yang terhambat (Ndulu et al., 2005). Dalam hal ini,



peningkatan pembangunan akses dan juga infrastruktur menjadi roda dan juga kunci dalam pergerakan kegiatan ekonomi di Indonesia.

Faktanya, infrastruktur di Indonesia saat ini dinilai belum merata, menyebabkan tingginya kesenjangan antar kabupaten dan provinsi dalam aspek ekonomi dan infrastruktur (Sukwika, 2018). Menurut *World Bank*, terdapat beberapa jenis akses infrastruktur yang sangat penting untuk menurunkan tingkat kemiskinan di Indonesia, yaitu:

1. Kemiskinan akan rendahnya akses pada fasilitas dan infrastruktur dasar.
2. Akses menuju fasilitas Sekolah Menengah.
3. Akses kepada Pendidikan informal. Semakin tinggi akses, maka pergerakan ekonomi semakin tinggi.
4. Akses kepada Lembaga Perkreditan yang berpengaruh pada pengeluaran dan penurunan rumah tangga miskin.
5. Akses kepada jalan yang berhubungan dengan peningkatan konsumsi.
6. Akses komunikasi yang berpengaruh secara signifikan pada beberapa daerah.

#### 2.4.5. Faktor Lokasi Geografis

Indonesia adalah salah satu negara maju berkembang yang selalu berupaya dalam meningkatkan nilai pertumbuhan ekonomi dengan berbagai program dan kebijakan. Menjadi sebuah negara kepulauan dengan Sumber Daya Alam yang berbeda-beda menjadikan tingkat ekonomi yang ada di setiap wilayah menjadi berbeda pula. Hal ini menjadikan daerah yang relatif kaya pastinya memiliki nilai tingkat pertumbuhan ekonomi yang lebih tinggi dibandingkan daerah lainnya. Hal ini sejalan dengan tingkat pertumbuhan ekonomi juga berdampak pada nilai tingkat kesejahteraan dan kemiskinan masyarakat (Mahardiki & Santoso, 2013).

Berdasarkan penelitian yang dilakukan oleh *World Bank* (2006), pengelompokan kemiskinan di Indonesia dapat dikelompokkan berdasarkan 6 pulau besar di Indonesia, yakni pulau Sumatera, pulau Jawa/Bali, pulau Kalimantan, pulau Sulawesi, pulau Nusa Tenggara/Maluku dan pulau Papua. Hal ini didasarkan pada :

1. Jawa/Bali : Kelebihan utama yang dimiliki adalah tingginya Pendidikan, akses dan juga asset. Selain itu infrastruktur (terutama akses ke jalan), alur kredit dan mobilitas menjadi keunggulan pada wilayah geografis ini.

2. Sumatera : Kelebihan Sumatera jika dibandingkan dengan Jawa/Bali adalah pengalaman kerja, pilihan jenis pekerjaan dan juga komunikasi yang tertinggi dari semua wilayah. Dalam hal Pendidikan, wilayah ini berada di posisi yang kurang menguntungkan dibandingkan Jawa/Bali.
3. Kalimantan : Wilayah ini memiliki posisi yang rendah pada sektor Pendidikan. Akses dan asset serta tingkat pekerjaan.
4. Sulawesi : Sama hal nya dengan Kalimantan, wilayah ini berada pada posisi yang kurang menguntungkan dalam sisi Pendidikan dan juga tingkat pengangguran
5. Nusa Tenggara/Maluku: Wilayah ini memilki keunggulan pada pilihan pekerjaan dan juga pengalaman pekerjaan. Namun jauh lebih rendah dalam hal akses dan juga Pendidikan. Kondisi tanah dan layanan menuju infrastruktur dasar menghambat pergerakan ekonomi di wilayah ini.
6. Papua : Wilayah ini memiliki nilai yang jauh lebih rendah dalam hal pendidkan, akses infrastruktur dan juga kualitas tenaga kerja.

Lebih lanjut, berdasarkan beberapa penelitian terbaru terkait penyebab kemiskinan di Indonesia, didapati beberapa faktor tambahan yang dinilai dapat memberikan pengaruh akan penyebab kemiskinan di Indonesia, antara lain :

#### 2.4.6. Faktor Kesehatan

Angka Harapan Hidup (AHH) menjadi salah satu faktor dalam mengevaluasi kinerja dan upaya pemerintah dalam meningkatkan kesejahteraan penduduk, khususnya di bidang kesehatan. Di berbagai negara, tentunya tingkat Kesehatan yang baik umumnya dikaitkan dengan lamanya rata-rata hidup masyarakatnya. Semakin tinggi Angka Harapan Hidup tentunya berpengaruh pada meningkatnya pendapatan secara ekonomis.

Memperbaiki layanan kesehatan oleh pemerintah di Indonesia juga menjadi kebijakan penting dalam mengurangi kemiskinan. Salah satu faktor dasar dari kebijakan ini adalah peningkatan tingkat kesehatan, yang berdampak pada peningkatan produktivitas masyarakat di daerah miskin. Kesehatan berkualitas baik dapat meningkatkan kemampuan kerja, meningkatkan output energi, dan mengurangi jumlah hari tidak bekerja. (Anggit & Arianti, 2012).

#### 2.4.7. Faktor Pendapatan Daerah

Pendapatan Daerah merupakan hak yang dimiliki pemerintah dan diakui sebagai faktor pendukung dalam meningkatkan nilai kekayaan bersih (UU No 33 tahun 2004). Pendapatan daerah mencakup semua penerimaan melalui Rekening Kas Umum Daerah, yang menjadi hak suatu daerah dalam satu tahun anggaran dan tidak perlu dibayar kembali oleh daerah (PP No. 58 tahun 2005 tentang Pengelolaan Keuangan). Tingkat kemiskinan di Indonesia juga dipengaruhi oleh pendapatan daerah melalui tingkat pertumbuhan Produk Domestik Bruto (PDB) riil, penyerapan tenaga kerja, dan faktor pendukung lainnya (Permana & Arianti, 2012).

### 2.5. Penelitian Terdahulu

Berbagai penelitian terdahulu telah dilakukan untuk menyempurnakan, menggabungkan, dan menemukan teknik serta metode baru dalam menangani kasus kemiskinan di Indonesia. Beberapa penelitian tersebut akan dijadikan acuan dan contoh dalam penyempurnaan penelitian ini. Penelitian pertama oleh Ramadhani et al. (2022) melibatkan penggunaan Data Mining dengan Algoritma C4.5 untuk menentukan kelayakan penerima bantuan sosial di Kantor Kelurahan Martoba. Studi ini bertujuan menerapkan Data Mining untuk merancang pola penyusunan berkas di Kantor Kelurahan Martoba, Pematang Siantar. Proses penerapan Algoritma C4.5 dibagi menjadi dua tahap: pertama, perhitungan semi-manual menggunakan Microsoft Excel; dan kedua, penyesuaian hasil perhitungan melalui pengujian data menggunakan perangkat lunak RapidMiner 5.3. Hasil penelitian menunjukkan bahwa teknik Data Mining dengan Algoritma C4.5 efektif dalam menyelesaikan masalah penentuan kelayakan penerima bantuan sosial di Pemkot.

Penelitian kedua oleh Abu et al. (2020), pada penelitian yang berjudul *Ensemble Learning for Multidimensional Poverty Classification* yang melakukan upaya klasifikasi kemiskinan multidimensi di Malaysia dengan menggunakan Algoritma Random Forest dan Decision Tree C4.5. Berdasarkan hasilnya, Random Forest menunjukkan performa terbaik dengan tingkat akurasi mencapai 99 persen. Namun, dilakukan pula perbandingan kedua algoritma tersebut yang menunjukkan bahwa akurasi Decision Tree C4.5 hanya berbeda satu persen saja. Selain itu, dinilai dari segi durasi proses, Decision Tree jauh lebih cepat dengan perbedaan yang mendekati sepuluh kali lipat dibandingkan dengan Random Forest. Dengan mempertimbangkan

bahwa Decision Tree menggunakan algoritma C4.5, maka penggunaan algoritma C4.5, yang merupakan pengembangan dari C4.5, berpotensi untuk meningkatkan performa lebih lanjut.

Penelitian ketiga, Wenas (2020), melakukan perbandingan antar penggunaan Algoritma C5.0 dan *Variable Importance Measure*. Berdasarkan metode *Variable Importance Measure*, faktor penyebab kemiskinan adalah kesehatan, pendidikan dan pekerjaan yang digambarkan dengan indikator HH, SD dan tani. Selain itu, indikator-indikator yang tergabung dalam satu pohon belum tentu memiliki hubungan yang kuat antara satu dengan yang lain.

Penelitian keempat, Xuanyuan, et al (2022), pada penelitian yang berjudul “*Application of C4.5 Algorithm in Insurance and Financial Services Using Data Mining Methods*”. Penerapan teknik data mining, khususnya menggunakan algoritma pohon keputusan seperti C4.5, memiliki potensi besar untuk memberikan keunggulan kompetitif bagi perusahaan di berbagai industri, termasuk asuransi. Melalui pelatihan algoritma C4.5, didapatkan pohon keputusan klasifikasi dimana pada penerpan model di set pelatihan dan uji untuk menguji akurasi, hasil pengujian menunjukkan tingkat identifikasi yang benar untuk keuangan asuransi sebesar 96.25%.

Penelitian kelima, Garg (2021), pada penelitian dengan judul *Prediction of Credit Card Using the Naïve Bayes Method and C4.5 Algorithm* melakukan penelitian dalam mengevaluasi dua metode dalam memprediksi pengajuan kartu kredit dengan kriteria yang menjadi dasar pengambilan keputusan antara lain usia, jenis kelamin, terkini, pendidikan, jumlah tanggungan, jenis perusahaan, pendapatan bulanan, dan slip gaji. Disimpulkan bahwa penggunaan metode naïve bayes dan algoritma c45 pada aplikasi Prediksi Pengajuan Kartu Kredit keduanya sudah tepat hasil.

**Tabel 2. 1** Penelitian Terdahulu

No	Penulis	Judul	Tahun	Keterangan
1.	Winda Lidysari	Penerapan Data Mining Dalam Menentukan Kelayakan Penerima Bantuan Sosial Pemko Dengan Algoritma	2020	Masalah dalam menentukan kelayakan penerima bantuan sosial di Pemko dapat diatasi menggunakan teknik Data Mining, yaitu dengan Algoritma C4.5.

	<p>C4.5 (Kasus Kantor Kelurahan Martoba)</p> <p>Berdasarkan hasil penghitungan menggunakan Algoritma C4.5, ditemukan bahwa variabel Penghasilan merupakan faktor yang paling berpengaruh, dengan nilai gain sebesar 0,8474359 dan tingkat akurasi mencapai 100%.</p>
<p>2. Abu Bakar, <i>Ensemble Learning for</i> 2020 A., Hamdan, <i>Multidimensional</i> R., &amp; Sani, N. <i>Poverty Classification</i> S. <i>using Random Forest</i> <i>and Decision Tree</i> <i>C4.5.</i></p>	<p>Penelitian ini membandingkan hasil klasifikasi kemiskinan di Malaysia menggunakan Algoritma <i>Random Forest</i> dan Algoritma <i>Decision Tree</i> C4.5. Dari segi waktu proses, <i>Decision Tree</i> menonjol dengan perbedaan hampir sepuluh kali lipat lebih cepat dibandingkan <i>Random Forest</i>. Mengingat penggunaan algoritma C4.5, maka ada peluang bahwa algoritma C4.5, yang merupakan evolusi dari C4.5, dapat menghasilkan performa yang lebih baik. Akurasi penelitian ini mencapai 98%.</p>

3.	Frenaldy Wahyudi Wenas	Pengklasifikasian Kemiskinan di Indonesia dengan menggunakan Metode Algoritma <i>Decision Tree C5.0</i>	2020	Berdasarkan pemodelan yang dilakukan ditemukan tiga model unik terbaik. faktor penyebab kemiskinan terbaik yang ditemukan adalah kesehatan, pendidikan dan pekerjaan yang digambarkan dengan indikator HH, SD dan tani. Nilai akurasi dari penelitian ini mencapai 97.1%.
4.	Xuanyuan, et al	<i>Application of C4.5 Algorithm in Insurance and Financial Services Using Data Mining Methods</i>	2022	Penerapan algoritma pohon keputusan seperti C4.5 dalam industri, termasuk asuransi, dapat memberikan keunggulan kompetitif. Melalui pelatihan, pohon keputusan menghasilkan tingkat identifikasi yang benar untuk keuangan asuransi sebesar 96.25%.
5.	Ketjie, et al	<i>Prediction of Credit Card Using the Naïve Bayes Method and C4.5 Algorithm</i>	2021	Penelitian ini mengevaluasi dua metode dalam memprediksi pengajuan kartu kredit dengan kriteria yang menjadi dasar pengambilan keputusan antara lain usia, jenis kelamin, terkini, pendidikan, jumlah tanggungan, jenis perusahaan, pendapatan

---

bulanan, dan slip gaji.  
Disimpulkan bahwa  
penggunaan metode naïve  
bayes dan algoritma c45  
keduanya sudah tepat hasil.

---

## 2.6. Perbedaan Penelitian

Terdapat beberapa pembeda penelitian ini dibandingkan penelitian sebelumnya. Penelitian sebelumnya telah dilakukan pengklasifikasian Kemiskinan di Indonesia menggunakan C4.5 menggunakan bahasa pemrograman R. Pada penelitian ini, penulis menggunakan Algoritma *Decision Tree* C4.5 dengan menggunakan bahasa pemrograman Python. Lebih utama, dataset yang akan digunakan pada penelitian ini jauh lebih dalam dan detail yang dimana data yang digunakan adalah seluruh indikator per Kabupaten/Kota di Indonesia, sedangkan pada penelitian sebelumnya hanya menggunakan data per Provinsi. Lebih lanjut, pada penelitian ini penulis juga melakukan tahapan lebih lanjut yakni untuk menentukan indikator utama penyebab kemiskinan dari setiap Provinsi yang dimana tahapan ini tidak ada pada penelitian sebelumnya. Pada penelitian ini penulis juga melakukan empat tahapan dalam *preprocessing* yaitu *missing value*, *data transformation*, *data binning* dan *encoding categorical variables*.

### BAB 3

#### ANALISIS DAN PERANCANGAN SISTEM

##### 3.1. Dataset

Dalam penelitian ini, data yang akan digunakan adalah nilai dari beberapa indikator yang diprediksi mampu menggambarkan penyebab kemiskinan yang ada di setiap Kabupaten/Kota di Indonesia dan telah melewati proses *scrapping*, dan disimpan ke dalam bentuk ekstensi *.xlsx*. Data dari hasil *scrapping* berjumlah sebanyak 1028 data dengan 12 variabel per data bersumber dari Badan Pusat Statistik, SUSENAS dan juga Potensi Desa, yang kemudian akan dibagi menjadi *data training* yang merupakan data tahun 2021 dan juga *data testing* yang merupakan data tahun 2022. Di bawah ini dalam tabel 3.1 merupakan data indikator penyebab kemiskinan.

**Tabel 3. 1** Data hasil *scrapping* pada indikator penyebab kemiskinan

Kabupaten	a	b	c	d	e	f	g	h	i	j	k	l
SIMEULUE	9.48	5.71	7.15	26.03	65.28	87.45	98.97	42.28	83.33	19.57	76.81	44.50
ACEH SINGKIL	8.68	8.36	8.78	21.05	67.43	78.58	99.98	74.36	23.53	66.39	78.15	29.47
ACEH SELATAN	8.88	6.46	8.18	25.41	64.4	79.65	99.88	57.90	94.23	38.46	72.31	32.24
ACEH TENGGARA	9.67	6.43	8.03	24.1	68.22	86.71	100	67.40	30.65	37.92	44.94	35.75
ACEH TIMUR	8.21	7.13	8.58	26.23	68.74	83.16	99.60	64.50	55.73	20.00	49.32	32.30
ACEH TENGAH	9.86	2.61	10.78	36.58	68.86	90.10	99.76	84.42	8.81	74.58	65.42	55.63



**Tabel 3. 2** Deskripsi Indikator Penyebab Kemiskinan

<b>Faktor</b>	<b>Indikator</b>
Faktor Pekerjaan	Tingkat Pengangguran Terbuka Pekerja Tani Sektor Informal
Faktor Pendidikan	Rata-rata Lama Sekolah Jumlah Desa yang Memiliki Sekolah Dasar
Faktor Akses terhadap Pelayanan dan Infrastruktur Dasar	Rumah Tangga dengan Penggunaan Listrik PLN Rumah Tangga dengan Air Minum Layak Rumah Tangga dengan Sanitasi layak Jumlah Desa yang Memiliki Lembaga Keterampilan Jumlah Desa yang Memiliki Lembaga Perkreditan
Faktor Kesehatan	Angka Harapan Hidup
Faktor Pendapatan Daerah	Pengeluaran per Kapita PDRB per Kapita

a. Rata-rata Lama Sekolah

Rata-rata Lama Sekolah adalah rata-rata durasi waktu (tahun) yang dihabiskan oleh penduduk dewasa muda, dalam hal ini berumur lebih dari 15 tahun untuk menjalani seluruh jenjang Pendidikan formal.

b. Tingkat Pengangguran Terbuka

Persentase Pengangguran Terbuka adalah perbandingan antara jumlah pengangguran dengan jumlah total angkatan kerja.

c. Pengeluaran per Kapita

Pengeluaran per kapita adalah rata-rata uang yang dihabiskan oleh setiap individu dalam suatu populasi dalam periode tertentu, mencakup semua jenis pengeluaran seperti makanan, perumahan, dan kesehatan. Ini digunakan sebagai indikator kesejahteraan ekonomi dan standar hidup.

d. PDRB per Kapita

PDRB per kapita merupakan nilai pendapatan rata-rata penduduk yang dihitung dengan membagi PDRB dengan jumlah populasi masyarakat di suatu wilayah pada

kurun 1 tahun tertentu. Disimpulkan bahwa semakin tinggi nilai pendapatan per kapita, maka semakin makmur wilayah tersebut.

e. Angka Harapan Hidup

Angka Harapan Hidup (AHH) digunakan sebagai indikator untuk menilai efektivitas kinerja pemerintah dalam meningkatkan kesejahteraan masyarakat secara keseluruhan, khususnya dalam meningkatkan tingkat kesehatan.

f. Rumah Tangga dengan Air Minum Layak

Air Minum Layak mencakup berbagai sumber air seperti air keran, air keran umum, terminal air, penampungan air hujan atau mata air, sumur yang terlindungi, dan sumur bor yang berlokasi minimal 10 meter dari tempat pembuangan limbah. Persentase indikator ini dihitung dengan membandingkan jumlah rumah tangga yang memiliki akses ke sumber air minum layak dengan total jumlah rumah tangga.

g. Rumah Tangga dengan Penggunaan Listrik PLN

Sumber penerangan yang sah dan dilindungi oleh pemerintah bersumber dari PLN, dalam hal ini penggunaan listrik yang dioperasikan oleh instansi lain disebut listrik non-PLN.

h. Rumah Tangga dengan Sanitasi layak

Sanitasi layak adalah kondisi di mana fasilitas sanitasi (seperti toilet, sistem pembuangan air limbah, dan tempat cuci tangan) dikelola dengan cara yang aman dan higienis guna menjaga kesehatan Masyarakat dan mencegah penyebaran penyakit.

i. Jumlah Desa yang Memiliki Lembaga Keterampilan

Jumlah desa yang memiliki lembaga keterampilan menunjukkan berapa banyak desa yang menyediakan pelatihan dan pengembangan keterampilan bagi warganya, membantu meningkatkan potensi ekonomi dan sumber daya manusia di desa.

j. Jumlah Desa yang Memiliki Lembaga Perkreditan

Jumlah desa yang memiliki lembaga perkreditan mencerminkan berapa banyak desa yang memiliki akses ke fasilitas kredit dan keuangan, mendukung usaha kecil dan menengah serta meningkatkan kesejahteraan ekonomi masyarakat desa.

k. Jumlah Desa yang Memiliki Sekolah Dasar

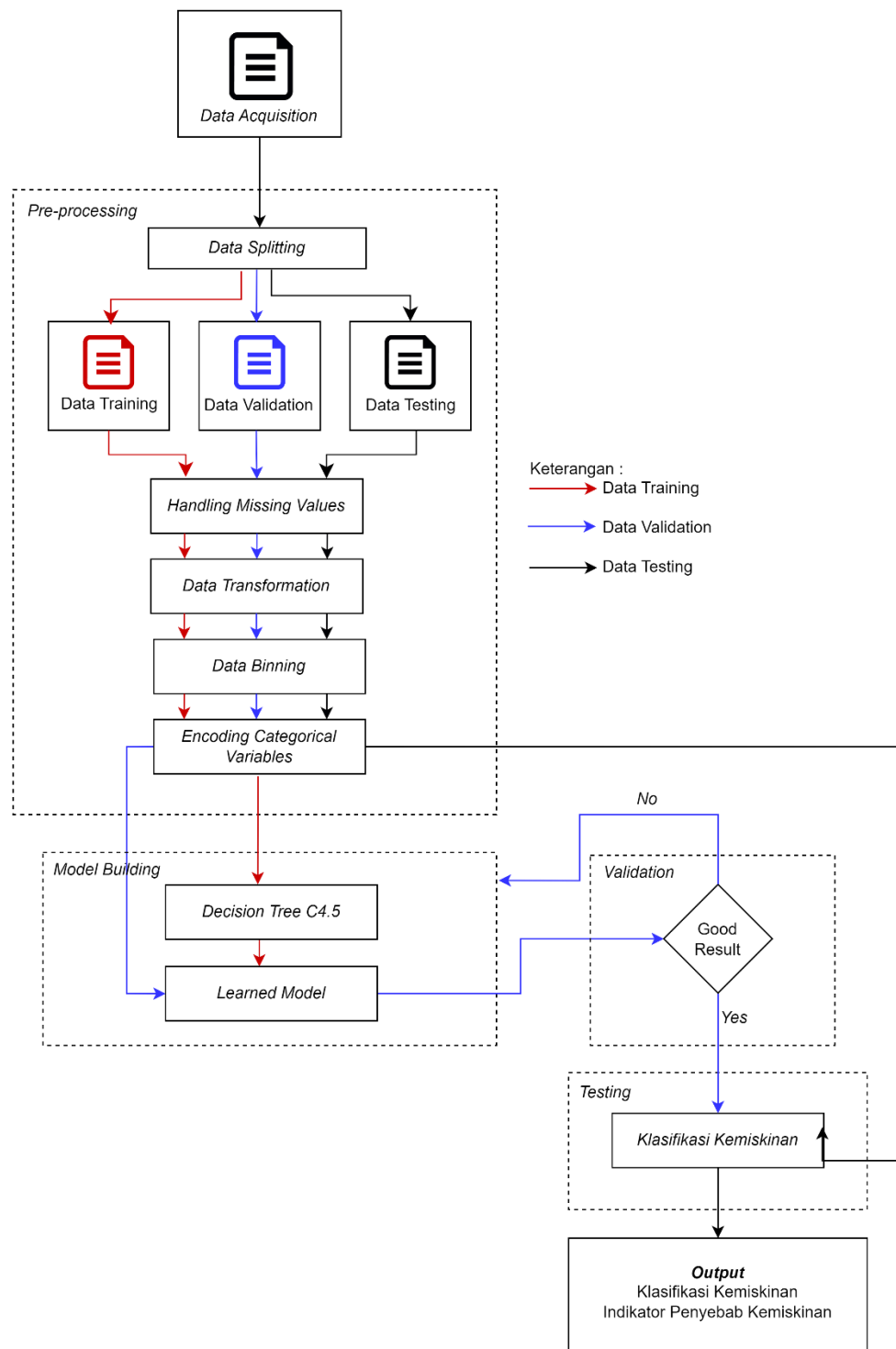
Kehadiran sekolah dasar di setiap desa sangat penting untuk mendukung pembangunan sumber daya manusia yang berkualitas, mengurangi kesenjangan pendidikan, dan mendorong pertumbuhan nilai ekonomi dan juga kesejahteraan masyarakat di pedesaan. Akses pendidikan dasar yang dekat dan terjangkau juga mengurangi angka putus sekolah dan memastikan bahwa anak-anak mendapatkan pendidikan dasar yang mereka butuhkan untuk Pendidikan ke jenjang yang lebih tinggi.

l. Pekerja Tani Sektor Informal

Merupakan indikator yang berisi persentase penduduk yang memiliki pekerjaan di bidang sektor pertanian baik yang merupakan usaha sendiri, pekerja buruh/tidak tetap maupun pekerja bebas.

### 3.2. Arsitektur Umum

Pada penelitian ini, terdapat beberapa tahap yang dilakukan, yaitu tahapan untuk mengumpulkan data, yang dimana data diperoleh dari website <https://www.bps.go.id/id> dengan proses scrapping dan disimpan dalam format *.xlsx*. Setelah itu, data melewati proses *preprocessing* yang terdiri dari *data splitting*, *handling missing value*, *data transformation*, *data binning* dan *encoding categorical variables*. Hasil dari model akan melewati proses pengujian kemampuan dengan menggunakan *data testing* untuk mengidentifikasi aspek dan sentimen. Arsitektur umum dalam penelitian ini digambarkan dalam gambar 3.1 sebagai berikut.



**Gambar 3. 1** Arsitektur Umum

### 3.2.1. Input Data

Dalam penelitian ini, data input berisi nilai-nilai numerik dari variabel yang diperkirakan mampu menggambarkan indikator-indikator penyebab kemiskinan secara optimal di seluruh Kabupaten/Kota di Indonesia. Sebanyak 12 variabel diproses melalui

transformasi data, yang diubah menjadi bentuk kategori yang sesuai untuk data mining, sehingga siap dihitung menggunakan Algoritma Decision Tree C4.5. Berikut ini adalah nilai variabel yang telah dikategorikan berdasarkan nilai setiap atribut.

**Tabel 3. 3** Variabel Rata-rata Lama Sekolah

Atribut Variabel	Nilai Atribut (Tahun)
tinggi	$> 8,9$
rendah	$\leq 8,9$

**Tabel 3. 4** Variabel Tingkat Pengangguran Terbuka

Atribut Variabel	Nilai Atribut (%)
$>10\%$	$> 10\%$
5%-10%	5% - 10%
$<5\%$	$< 5\%$

**Tabel 3. 5** Variabel Pengeluaran per Kapita

Atribut Variabel	Nilai Atribut (Juta)
tinggi	$> 10$
menengah	$\geq 8$
rendah	$< 8$

**Tabel 3. 6** Variabel PDRB per Kapita

Atribut Variabel	Nilai Atribut (Juta)
sangat sejahtera	$> 100$
sejahtera	$\geq 50$
kurang sejahtera	$\geq 40$
tidak sejahtera	$< 40$

**Tabel 3. 7** Variabel Angka Harapan Hidup

Atribut Variabel	Nilai Atribut (Tahun)
umur panjang	$\geq 73,5$
umur singkat	$< 73,5$

**Tabel 3. 8** Variabel Rumah Tangga dengan Air Minum Layak

Atribut Variabel	Nilai Atribut (%)
terjamin	$\geq 90,78\%$
kurang terjamin	$< 90,78\%$

**Tabel 3. 9** Variabel Rumah Tangga dengan Penggunaan Listrik PLN

Atribut Variabel	Nilai Atribut (%)
terdistribusi baik	$\geq 98,76\%$
terdistribusi kurang baik	$< 98,76\%$

**Tabel 3. 10** Variabel Rumah Tangga dengan Sanitasi layak

Atribut Variabel	Nilai Atribut (%)
optimal	$\geq 80,29\%$
tidak optimal	$< 80,29\%$

**Tabel 3. 11** Variabel Jumlah Desa yang Memiliki Lembaga Keterampilan

Atribut Variabel	Nilai Atribut (%)
unggul	$\geq 31,38\%$
berkembang	$< 31,38\%$

**Tabel 3. 12** Variabel Jumlah Desa yang Memiliki Lembaga Perkreditan

Atribut Variabel	Nilai Atribut (%)
optimal	$> 70\%$
terpenuhi	$\geq 50\%$
tidak terpenuhi	$< 50\%$

**Tabel 3. 13** Variabel Jumlah Desa yang Memiliki Sekolah Dasar

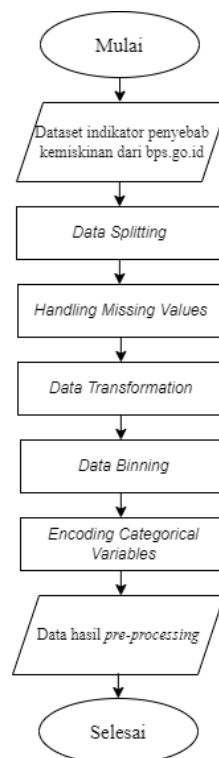
Atribut Variabel	Nilai Atribut (%)
optimal	$> 75\%$
terpenuhi	$\geq 50\%$
tidak terpenuhi	$< 50\%$

**Tabel 3. 14** Variabel Pekerja Tani Sektor Informal

Atribut Variabel	Nilai Atribut (%)
mayoritas	> 60%
sebagian	$\geq$ 40%
sedikit	< 40%

### 3.2.2. *Pre-processing*

Tahap *Pre-processing* adalah tahap awal yang dilakukan dalam proses *Data Mining* guna mengolah data mentah menjadi data yang lebih terstruktur dan siap untuk diolah lebih lanjut. Berikut penjelasan lebih lanjut dalam *flowchart* mengenai proses dalam tahapan *pre-processing* yang akan dilakukan.

**Gambar 3. 2** *Flowchart* tahapan *Pre-processing*

#### 1) *Data Splitting*

Dataset indikator penyebab kemiskinan dengan total 1028 data dibagi menjadi 2 bagian, yakni *data training* merupakan data tahun 2021, sedangkan *data testing* merupakan data tahun 2022. Adapun pembagian data tersebut dapat dilihat pada Tabel 3.14. di bawah ini.

**Tabel 3. 15 Data Splitting**

Data	Jumlah Data
<i>Training</i> dan Validasi	514
<i>Testing</i>	514
Total	1028

## 2) *Handling Missing Values*

Merupakan salah satu tahap yang paling penting saat melakukan *Pre-processing*. Pada proses ini, dataset akan melakukan pengisian / mengganti data yang tidak terisi menggunakan nilai lain yang diambil dari Mean data yang ada. Berikut merupakan *pseudo code* dari tahapan *Handling Missing Values*.

```
Function missing_data(data):
FOR each column in miss_cols:
    IF column type is 'object':
        Remove '%' characters
        Convert type to 'float'
    Fill missing values with column mean
RETURN modified data
```

Tahap *Handling Missing Values* dapat dilihat dalam Tabel 3.15.

**Tabel 3. 16 Tahap *Handling Missing Values***

Dataset Sebelum Proses	Dataset Setelah Proses
34.23%	34.23
51.11%	51.11
	66.52
54.55%	54.55
69.12%	69.12

## 3) *Data Transformation*

*Data transformation* merupakan proses mengubah struktur, format atau nilai data untuk membuatnya lebih cocok untuk analisis atau pemodelan. Dalam penelitian ini, *Data Transformation* dilakukan untuk mengubah seluruh tipe data yang berbeda-beda seperti persen dan *string* menjadi *float*. Berikut merupakan *pseudo code* dari tahapan *Data Transformation*.

```
Function preprocess data(self, data):
```



```
# Drop unnecessary columns
Remove column "Kabupaten/Kota" from data

# Map 'miskin' column to binary values
Convert values in "miskin" column:
    If value is "no", change to 0
    If value is "yes", change to 1

# Convert percentage columns to float
Define percentage_columns as list containing "tpt",
"air_minum_layak", "listrik_pln", "sanitasi",
"lembaga_keterampilan", "lembaga_kredit", "sekolah_dasar",
"sektor_tani"

For each column in percentage_columns:
    Convert column values to string
    Remove '%' character from column values
    Convert column values to float
```

Tahap *Data Transformation* dapat dilihat dalam tabel 3.16.

**Tabel 3. 17 Tahap Data Transformation**

Dataset Sebelum Proses	Dataset Setelah Proses
34.23	34.23
51.11	51.11
66.52	66.52
54.55	54.55
69.12	69.12

#### 4) Data Binning

*Data Binning* adalah teknik prapemrosesan data yang digunakan untuk mengelompokkan nilai-nilai numerik dari suatu variabel ke dalam beberapa interval atau kategori yang disebut "bins". Proses ini membantu dalam mengurangi efek dari outlier dan membuat data lebih mudah dianalisis dengan cara mengubah data kontinu menjadi data diskrit. Berikut merupakan *pseudo code* dari tahap *Data Binning*.

```
Function apply_binning(data):
    # Bin 'lama_sekolah_tahun' into categories
    Bin data['lama_sekolah_tahun'] into:
        'rendah' for values <= 8.9
        'tinggi' for values > 8.9
```

```

# Bin 'tpt' into categories
Bin data['tpt'] into:
    '<5%' for values <= 5
    '5%-10%' for values > 5 and <= 10
    '>10%' for values > 10

# Bin 'pengeluaran_kapita_juta' into categories
Bin data['pengeluaran_kapita_juta'] into:
    'rendah' for values <= 8
    'menengah' for values > 8 and <= 10
    'tinggi' for values > 10

# Bin 'pdrb_kapita_juta' into categories
Bin data['pdrb_kapita_juta'] into:
    'tidak sejahtera' for values <= 40
    'kurang sejahtera' for values > 40 and <= 50
    'sejahtera' for values > 50 and <= 100
    'sangat sejahtera' for values > 100

```

Tahap *Data Binning* dapat dilihat dalam tabel 3.17.

**Tabel 3. 18 Tahap Data Binning**

Dataset Sebelum Proses	Dataset Setelah Proses
34.23	tidak terpenuhi
51.11	terpenuhi
66.52	terpenuhi
54.55	terpenuhi
69.12	terpenuhi

##### 5) *Encoding Categorical Variabels*

*Encoding Categorical Variabels* adalah proses mengonversi data kategorikal (seperti teks atau label) menjadi format numerik yang dapat digunakan oleh algoritma machine learning. Proses ini penting karena banyak algoritma, termasuk C4.5, tidak dapat bekerja dengan data kategorikal secara langsung. Berikut merupakan *pseudo code* dari tahap *Encoding Categorical Variabels*.

```

Function encode_categorical_columns(data):
    Initialize empty dictionary label_encoders

    Define categorical_columns as list containing:

```

```

        'lama_sekolah_tahun', 'tpt',
        'pengeluaran_kapita_juta', 'pdrb_kapita_juta', 'ahh_tahun',
        'air_minum_layak', 'listrik_pln', 'sanitasi',
        'lembaga_keterampilan', 'lembaga_kredit',
        'sekolah_dasar', 'sektor_tani'

    For each column in categorical_columns:
        Initialize LabelEncoder le
        Fit and transform data[column] using le
        Store le in label_encoders with key as column name

    Return data and label_encoders

```

Tahap *Encoding Categorical Variabels* dapat dilihat dalam tabel 3.18.

**Tabel 3. 19 Tahap Encoding Categorical Variabels**

Dataset Sebelum Proses	Dataset Setelah Proses
tidak terpenuhi	2
terpenuhi	1
terpenuhi	1
terpenuhi	1
terpenuhi	1

### 3.2.3. Training Model Building

*Training Model Building* merupakan tahapan yang akan dilakukan setelah melalui tahap *Pre-processing* yang bertujuan untuk membentuk dan melatih model. Dalam proses pembentukan *Decision Tree C4.5*, dimulai dengan menghitung nilai node akar. Berikut adalah Langkah-langkah dalam perhitungan *Decision Tree C4.5* :

#### 1) Menghitung Jumlah Kasus

Dalam membangun model, sebagai langkah awal perlu dilakukannya perhitungan jumlah kasus dalam hal ini yakni variabel dari indikator penyebab kemiskinan atau dengan kata lain dilakukannya agresi pada dataset seperti pada tabel 3.18 :

**Tabel 3. 20 Agresi Dataset Penyebab Kemiskinan**

Atribut	Nilai	Jumlah Kasus	Miskin	Tidak Miskin
		514	318	196

Lama Sekolah	tinggi	163	57	106
	rendah	351	261	90
Tingkat Pengangguran Terbuka	<5%	293	213	80
	5%-10%	192	95	97
	>10%	29	10	19
Pengeluaran per Kapita	rendah	95	64	31
	menengah	140	104	36
	tinggi	279	150	129
PDRB per Kapita	sangat sejahtera	56	10	46
	sejahtera	119	52	67
	kurang sejahtera	87	58	29
	tidak sejahtera	252	198	54
Angka Harapan Hidup	umur panjang	76	43	33
	umur singkat	438	275	163
Rumah Tangga dengan Air	terjamin	243	145	98
Minum Layak	tidak terjamin	271	173	98
Rumah Tangga dengan Penggunaan Listrik PLN	terdistribusi baik	264	159	105
	terdistribusi kurang baik	250	159	91
Rumah Tangga dengan Sanitasi layak	optimal	310	175	135
	tidak optimal	204	143	61
Jumlah Desa yang Memiliki Lembaga Keterampilan	unggul	135	50	85
	berkembang	379	268	111
Jumlah Desa yang Memiliki Lembaga Perkreditan	optimal	239	162	77
	terpenuhi	71	33	38
	tidak terpenuhi	204	123	81
Jumlah Desa yang Memiliki Sekolah Dasar	optimal	438	280	158
	terpenuhi	47	26	21
	tidak terpenuhi	29	12	17
	mayoritas	35	22	13

Pekerja Tani Sektor	sebagian	149	112	37
Informal	sedikit	330	184	146

---

2) Perhitungan Nilai *Entropy* Total, *Entropy* dan *Gain*. Sebagaimana persamaan (1) yang telah dijelaskan pada Bab 2, maka ditentukan :

$$\begin{aligned}
 Entropy(Total) &= \sum_{i=1}^c -p_i \times \log_2(p_i) \\
 &= \left( -\frac{318}{514} \times \log_2\left(\frac{318}{514}\right) \right) + \left( -\frac{196}{514} \times \log_2\left(\frac{196}{514}\right) \right) \\
 &= (-0,619 \times \log_2(0,619)) + (-0,381 \times \log_2(0,381)) \\
 &= (-0,619 \times -0,693) + (-0,381 \times -0,419) \\
 &= 0,429 + 0,530 \\
 &= 0,959
 \end{aligned}$$

Berikut merupakan contoh perhitungan *entropy* dari nilai-nilai atribut variabel Rata-rata Lama Sekolah, yaitu Tinggi dan Rendah.

$$\begin{aligned}
 Entropy(Lama Sekolah, Tinggi) &= \left( -\frac{57}{163} \times \log_2\left(\frac{57}{163}\right) \right) + \left( -\frac{106}{163} \times \log_2\left(\frac{106}{163}\right) \right) \\
 &= (-0,350 \times -0,516) + (-0,650 \times -0,620) \\
 &= 0,530 + 0,404 \\
 &= 0,934
 \end{aligned}$$

$$\begin{aligned}
 Entropy(Lama Sekolah, Rendah) &= \left( -\frac{261}{351} \times \log_2\left(\frac{261}{351}\right) \right) + \left( -\frac{90}{351} \times \log_2\left(\frac{90}{351}\right) \right) \\
 &= (-0,744 \times -0,427) + (-0,256 \times -0,193) \\
 &= 0,317 + 0,504 \\
 &= 0,821
 \end{aligned}$$

Berikut merupakan tabel hasil perhitungan *Entropy* nilai-nilai atribut Rata-rata Lama Sekolah.

**Tabel 3. 21** Sampel Perhitungan *Entropy* Rata-rata Lama Sekolah

Atribut	Nilai	Jumlah Kasus	Miskin	Tidak Miskin	Entropy
Lama Sekolah	tinggi	163	57	106	0,934
	rendah	351	261	90	0,821

Perhitungan serupa juga perlu dilakukan pada seluruh Atribut yang ada di Dataset. Lebih lanjut, dilakukan perhitungan nilai *Gain* pada setiap Atribut menggunakan persamaan (2) sebagaimana yang telah dijelaskan pada Bab 2. Berikut merupakan contoh perhitungan *Gain* dari variabel Rata-rata Lama Sekolah.

$$\begin{aligned}
 \text{Gain}(\text{Total}, \text{Lama Sekolah}) &= \text{Entropy}(\text{Total}) - \sum_{i=1}^c \frac{|S_i|}{|S|} \times \text{Entropy}(S_i) \\
 &= 0,959 - \left( \left( \frac{163}{514} \times 0,934 \right) + \left( \frac{351}{514} \times 0,821 \right) \right) \\
 &= 0,959 - ((0,296) + (0,561)) \\
 &= 0,959 - (0,857) \\
 &= 0,102
 \end{aligned}$$

Sebagai tahap lebih lanjut yang wajib dilakukan pada *Algoritma Decision Tree* C4.5, maka akan dilakukan perhitungan *Gain Ratio* sebagaimana akan dilampirkan pada variabel Rata-rata Lama Sekolah berikut.

$$\begin{aligned}
 \text{GainRatio}(\text{Total}, \text{Lama Sekolah}) &= \frac{\text{Gain}(\text{Total}, \text{Lama Sekolah})}{\text{Split Information}(\text{Total}, \text{Lama Sekolah})} \\
 &= \frac{0,102}{-\left(\frac{163}{514}\right) \times \log_2\left(\frac{163}{514}\right) - \left(\frac{351}{514}\right) \times \log_2\left(\frac{351}{514}\right)} \\
 &= \frac{0,102}{0,901} \\
 &= \frac{0,102}{0,113}
 \end{aligned}$$

Perhitungan *Entropy* dari seluruh atribut pada Tabel 3.20 beserta nilai *Gain* dan *Ratio Gain* akan dilampirkan pada Tabel berikut.

**Tabel 3. 22** Perhitungan *Entropy* dan *Gain* untuk Node Akar

Atribut	Nilai	Jumlah Kasus	Miskin	Tidak Miskin	Entropy	Gain	Gain Ratio
		514	318	196	0,957		

Lama Sekolah	tinggi	163	57	106	0,934		
	rendah	351	261	90	0,821	0,102	0,113
Tingkat Pengangguran Terbuka	<5%	293	213	80	0,846		
	5%-10%	192	95	97	1,000		
	>10%	29	10	19	0,929	0,051	0,041
Pengeluaran per Kapita	rendah	95	64	31	0,911		
	menengah	140	104	36	0,822	0,026	0,018
	tinggi	279	150	129	0,996		
PDRB per Kapita	sangat sejahtera	56	56	10	0,677		
	sejahtera						0,075
	sejahtera	119	52	67	0,989		
	kurang sejahtera	87	58	29	0,918	0,133	
	tidak sejahtera	252	198	54	0,750		
Angka Harapan Hidup	umur	76	43	33	0,987		
	panjang						
	umur singkat	438	275	163	0,952	0,001	0,002
Rumah Tangga dengan Air Minum Layak	terjamin	243	145	98	0,973		
	tidak terjamin	271	173	98	0,944	0,001	0,001
Rumah Tangga dengan Penggunaan Listrik PLN	terdistribusi baik	264	159	105	0,970		
	terdistribusi kurang baik	250	159	91	0,946	0,001	0,001
Rumah Tangga	optimal	310	175	135	0,988		
	tidak optimal	204	143	61	0,880		

dengan						0,014	0,014
Sanitasi layak							
Jumlah Desa	unggul	135	50	85	0,951		
yang	berkembang	379	268	111	0,872		
Memiliki						0,066	0,079
Lembaga							
Keterampilan							
Jumlah Desa	optimal	239	162	77	0,907		
yang	terpenuhi	71	33	38	0,996		
Memiliki	tidak	204	123	81	0,969	0,345	0,240
Lembaga	terpenuhi						
Perkreditan							
Jumlah Desa	optimal	438	280	158	0,943		
yang	terpenuhi	47	26	21	0,992	0,009	0,012
Memiliki	tidak	29	12	17	0,978		
Sekolah	terpenuhi						
Dasar							
Pekerja Tani	mayoritas	35	22	13	0,952		
Sektor	sebagian	149	112	37	0,809	0,024	0,031
Informal	sedikit	330	184	146	0,990		

Kemudian terpilih Atribut “Jumlah Desa yang Memiliki Lembaga Perkreditan” dengan Nilai *Gain Ratio* tertinggi, yang akan digunakan untuk memisahkan data dalam Langkah pertama. Lalu proses akan dilakukan berulang hingga semua data termasuk dalam satu kelas atau tidak ada atribut lagi yang bisa dipilih.

### 3) Model *Decision Tree C4.5*

Algoritma C4.5 bekerja dengan memilih atribut terbaik untuk memisahkan data ke dalam kelas yang berbeda menggunakan *entropy*, *gain*, dan *gain ratio* seperti yang sudah dihitung pada Tabel 3.20. Pada tahapan ini model akan mulai dibentuk menggunakan data training dan data validation. Dari pembentukan model ini, akan



dijabarkan hasil dari nilai *Accuracy*, *Precision*, *F1-Score* dan *Cross Validation* untuk melihat Gambaran tentang seberapa baik model bekerja pada data yang tidak digunakan selama pelatihan. Model ini menggunakan 2 parameter yakni *max\_depth* dengan nilai 5 dan *min\_sample\_leaf* dengan nilai 2. Berikut merupakan *pseudo code* dari *Decision Tree C4.5*.

```
Function ModelC45(X_train, y_train):
    Initialize C45Classifier as C45
    Fit C45 with X_train and y_train
    Return C45

model = Call ModelC45 with parameters X and y
```

#### 3.2.4. Testing Model Building

Pada tahapan ini, dilakukan klasifikasi data dengan Model yang sudah dibangun, dengan data yang digunakan adalah data *testing*. Tahap ini menjadi tahap guna menguji model yang telah dievaluasi pada proses training dan mengetahui apakah model ini dapat melakukan klasifikasi status kemiskinan di suatu daerah berdasarkan fitur ataupun indikator dan juga nilai yang ada pada dataset test. Dalam menilai seberapa baik model ini dalam melakukan klasifikasi, akan ditampilkan nilai dari *F1-score* pada tahap *testing* ini.

#### 3.2.5. Output

Adapun keluaran dari program pada penelitian ini adalah klasifikasi status kemiskinan yakni Miskin dan Tidak Miskin dari setiap data, beserta indikator-indikator yang memberikan pengaruh paling baik dalam penentuan klasifikasi tersebut, yang diharapkan dapat digunakan sebagai bahan evaluasi oleh pihak terkait.

### 3.3. Perancangan Aplikasi Sistem

Perancangan aplikasi sistem merupakan proses sistematis dan terstruktur guna mengembangkan aplikasi yang sesuai dengan kebutuhan pengguna. Proses ini mencakup penentuan persyaratan, analisis, desain, implementasi, pengujian, dan pemeliharaan aplikasi sistem. Tujuan dari perancangan Aplikasi Sistem ini adalah untuk menggambarkan operasi sistem secara umum.

### 3.3.1. Gambaran tampilan *homepage*

Tampilan halaman depan atau *homepage* merupakan tampilan utama yang pertama kali dilihat oleh pengguna saat mengakses sistem. Pada halaman ini, informasi yang ditampilkan mencakup judul penelitian, identitas peneliti, serta menu tombol yang berfungsi untuk menghubungkan ke halaman-halaman lainnya. Berikut adalah gambaran tampilan halaman depan web aplikasi yang dapat dilihat pada gambar 3.3.



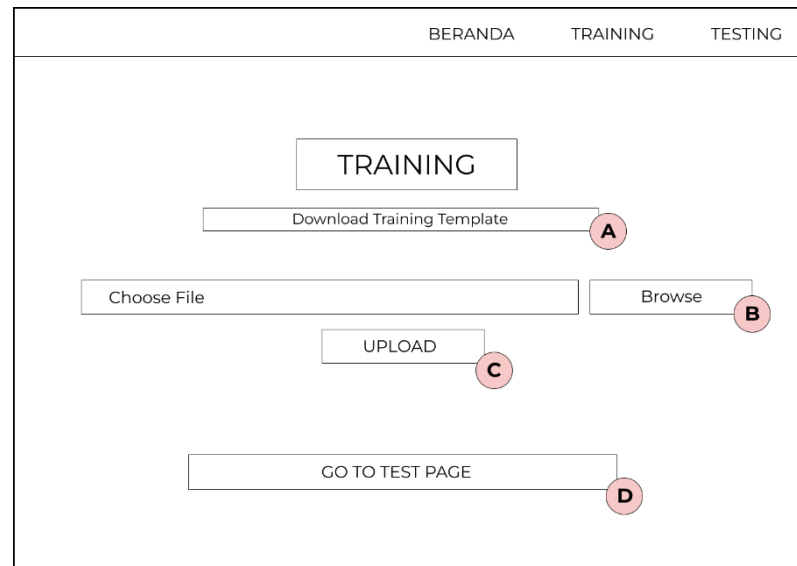
**Gambar 3. 3** Gambaran *Homepage*

Keterangan:

1. Fungsi tombol A untuk menampilkan halaman depan atau beranda.
2. Fungsi tombol B untuk memberikan tampilan halaman training.
3. Fungsi tombol C untuk menampilkan halaman testing.
4. Fungsi tombol D untuk menampilkan informasi seputar kemiskinan dan data kemiskinan di Indonesia.

### 3.3.2. Gambaran tampilan halaman *training*

Perancangan tampilan *training* bertujuan untuk memberikan visualisasi yang jelas dan informatif mengenai proses pelatihan model. Tampilan ini dirancang agar pengguna dapat dengan mudah memantau perkembangan kinerja model selama proses *training*. Berikut adalah gambaran tampilan training data.



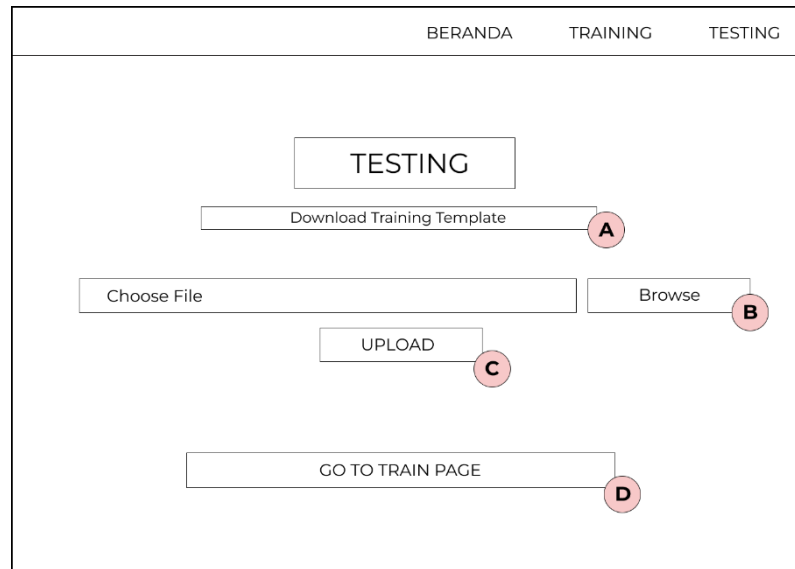
**Gambar 3. 4** Gambaran tampilan *training data*

Keterangan:

1. Fungsi tombol A untuk mengunduh template yang digunakan dalam *training data*.
2. Fungsi tombol B untuk mencari file dari direktori dengan ekstensi *.csv* dan *.xlsx*.
3. Fungsi tombol C untuk mengunggah *data training*.
4. Fungsi tombol D untuk menuju ke halaman *data testing*.

### 3.3.3. Gambaran tampilan halaman *testing*

Tampilan ini merupakan gambaran proses *testing* data dengan hasil output berupa nilai akurasi dan hasil prediksi dari model yang telah dibuat sebelumnya. Berikut ini adalah gambaran dari tampilan *testing data*.



**Gambar 3. 5** Gambaran tampilan *testing data*

Keterangan:

1. Fungsi tombol A untuk mengunduh *template* yang digunakan dalam *testing data*.
2. Fungsi tombol B untuk mencari file dari direktori dengan ekstensi *.csv* dan *.xlsx*.
3. Fungsi tombol C untuk mengunggah *data testing*.
4. Fungsi tombol D untuk menuju ke halaman *data training*.

### 3.4. Metode Evaluasi

Metode Evaluasi dilakukan untuk mengidentifikasi hasil dari pemrosesan model yang dilakukan selama *data training*. Adapun metode yang digunakan dalam penelitian ini yakni metode *K-Fold Cross Validation*, *Confusion Matrix* dan *Permutation Importance*.

#### 3.4.1. *K-Fold Cross Validation*

*K-Fold Cross Validation* adalah metode validasi yang digunakan untuk menilai performa model dengan cara membagi dataset menjadi K subset (folds) yang kurang lebih sama besar. Proses ini diulang K kali, setiap kali menggunakan satu subset yang berbeda sebagai data validasi dan sisanya sebagai data pelatihan. Akurasi atau metrik evaluasi lainnya kemudian dirata-rata dari semua iterasi untuk memberikan gambaran yang lebih stabil tentang kinerja model. Dengan menggunakan metode ini, dapat memastikan bahwa model *Decision Tree C4.5* dengan dataset terbatas yang dibangun untuk mengidentifikasi indikator kemiskinan di Indonesia memiliki kinerja yang andal dan memberikan hasil yang lebih akurat. Berikut merupakan *pseudo code* dari *K-Fold Cross Validation*.

```

Initialize k to 5
Set random_state to 42

For each train_index, val_index in KFold(n_splits=k,
shuffle=True, random_state=42).split(X_train):
    Split X_train and y_train into X_train_fold, X_val_fold,
y_train_fold, y_val_fold using train_index and val_index
    Predict y_val_pred using X_val_fold
    Calculate val_accuracy as accuracy_score(y_val_fold,
y_val_pred)

Output validation_scores

```

### 3.4.2. Confusion Matrix

*Confusion Matrix* adalah alat evaluasi yang digunakan guna mengukur kinerja model klasifikasi. *Confusion Matrix* memberikan gambaran tentang bagaimana model klasifikasi bekerja dalam bentuk tabel yang menunjukkan perbandingan antara prediksi model dengan hasil aktual. Matriks ini terdiri dari empat komponen utama:

**Tabel 3. 23** Implementasi Evaluasi *Confusion Matrix*

		Label Data	
		<i>Positive</i>	<i>Negative</i>
Hasil Identifikasi	Positive (1)	(TP) <i>True Positive</i>	(FP) <i>False Positive</i>
	Negative (-1)	(FN) <i>False Negative</i>	(TN) <i>True Negative</i>

Keterangan:

1. (TP) *True Positive* : Banyaknya nilai aktual yang memiliki nilai positif dan data prediksi memiliki nilai positif.
2. (FP) *False Positive* : Banyaknya nilai aktual yang memiliki nilai negatif dan data prediksi memiliki nilai positif.
3. (TN) *True Negative* : Banyaknya nilai aktual yang memiliki nilai negatif dan data prediksi memiliki nilai negatif.

4. (FN) *False Negative* : Banyaknya nilai aktual yang memiliki nilai positif dan data prediksi memiliki nilai negatif.

### 3.4.3. *Permutation Importance*

*Permutation Importance* adalah teknik dalam *machine learning* yang digunakan untuk mengukur pengaruh setiap variabel input (fitur) terhadap kinerja model prediksi. Metode ini bekerja dengan cara mengacak nilai-nilai dari setiap fitur secara individual dan mengamati seberapa besar penurunan kinerja model sebagai akibatnya. Dengan demikian, *permutation importance* menunjukkan seberapa kritis setiap fitur dalam membantu model membuat prediksi yang akurat. Dalam konteks penelitian ini, *permutation importance* membantu mengidentifikasi indikator-indikator kunci yang mempengaruhi kemiskinan, mendukung pengambilan keputusan yang lebih berbasis data, serta memberikan wawasan yang berharga untuk pengembangan kebijakan dan strategi penanggulangan kemiskinan yang lebih efektif. Berikut merupakan *pseudo code* dari *Permutation Importance*.

```
Calculate permutation importance using
permutation_importance(model, X_train, y_train)
Create DataFrame importance_df with 'Feature' as X.columns and
'Importance' as importances
Sort importance_df by 'Importance' in descending order
Output importance_df
```

## BAB 4

### IMPLEMENTASI DAN PENGUJIAN SISTEM

#### 4.1. Implementasi Sistem

Rancangan sistem yang telah dijelaskan pada Bab 3 akan diimplementasikan menggunakan perangkat keras dan perangkat lunak sebagai berikut:

##### 4.1.1. Spesifikasi Perangkat Keras dan Perangkat Lunak

Adapun spesifikasi perangkat keras (*Hardware*) yang digunakan pada implementasi sistem yang dirancang sebagai berikut :

1. Laptop Lenovo IdeaPad Slim 5 14ABR8
2. AMD Ryzen 5 7530U with Radeon Graphics 2.00 GHz
3. Installed RAM 16.00 GB
4. SSD 256 GB

Adapun spesifikasi perangkat lunak (*Software*) yang digunakan pada implementasi sistem yang dirancang sebagai berikut :

1. Windows 11 *Home Single Language*
2. Bahasa pemrograman *Python* 3.8.6 dengan library *pandas*, *sklearn*, *c45-decision-tree*, *numpy*, *scikit-learn*, *seaborn* dan *matplotlib*.
3. *Microsoft Visual Studio Code*

##### 4.1.2. Penerapan Perancangan Tampilan Antarmuka

Adapun implementasi perancangan tampilan antarmuka yang telah dibuat pada Bab 3 yakni :

1. Tampilan *Homepage*

Tampilan ini merupakan yang pertama kali dilihat oleh User saat mengakses ataupun menjalankan system. Pada tampilan ini terdapat Judul Aplikasi (Penelitian), nama penulis dan juga *navigation bar* guna menuju halamn *training* dan juga *testing*. Lebih lanjut, pada halaman depan ini juga disediakan informasi terkait kemiskinan dan dataset kemiskinan di Indonesia. Tampilan Homepage dapat dilihat pada gambar 4.1.

Homepage Training Testing

## IDENTIFIKASI INDIKATOR PENYEBAB KEMISKINAN DI INDONESIA MENGGUNAKAN ALGORITMA DECISION TREE C4.5

Bella Fransiska Rejeki Simamora (191402036)  
Program Studi S1 Teknologi Informasi  
Fakultas Ilmu Komputer dan Teknologi Informasi  
2024

### Sumber Informasi

Download Dataset

No	Nama	Sumber (Filter Tahun dan Wilayah Menurut Provinsi)
1	Informasi Kabupaten/Kota dan Provinsi di Indonesia	<a href="#">Kode Relasi - SIG BPS</a>
2	Website Badan Pusat Statistik Provinsi	<a href="#">BPS Per Provinsi</a>
3	Website Badan Pusat Statistik Kabupaten/Kota	<a href="#">BPS Per Kabupaten/Kota</a>
4	Tingkat Pengangguran Terbuka	<a href="#">TPT</a>
5	Rata-Rata Lama Sekolah	<a href="#">Rata-rata Lama Sekolah</a>
6	PDRB ADHK Menurut Pengeluaran Per Kapita	<a href="#">Pengeluaran Per Kapita</a>
7	Produk Domestik Regional Bruto ADHK Per Kapita	<a href="#">PDRB Per Kapita</a>
8	Angka Harapan Hidup	<a href="#">AHH</a>

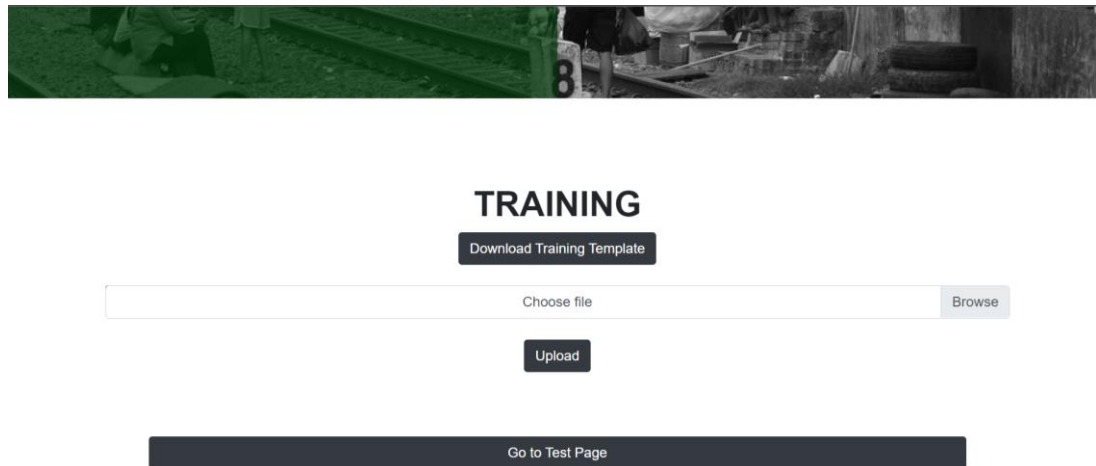
**Gambar 4. 1** Tampilan *Homepage*

## 2. Tampilan Halaman *Training*

Halaman ini akan digunakan oleh user saat akan mengunggah data yang akan dilatih oleh model. Tampilan ini memiliki judul halaman, tombol *untuk Download Training Template* yang dimana dataset yang akan diunggah wajib menyesuaikan dengan isi *template*.

Adapun urutan kolom yang wajib diisi dalam dataset secara berurutan adalah “Kabupaten/Kota, lama\_sekolah, tpt, pengeluaran\_kapita\_juta, pdrb\_kapita\_juta, ahh\_tahun, air\_minum\_layak, listrik\_pln, sanitasi, lembaga\_keterampilan, lembaga\_kredit, sekolah\_dasar, sektor\_tani dan miskin sebagai kolom target”. Setelah menunggah dataset, user dapat menekan tombol “*Upload*” guna menjalankan model. Tampilan halaman *training* dapat dilihat pada gambar 4.2.

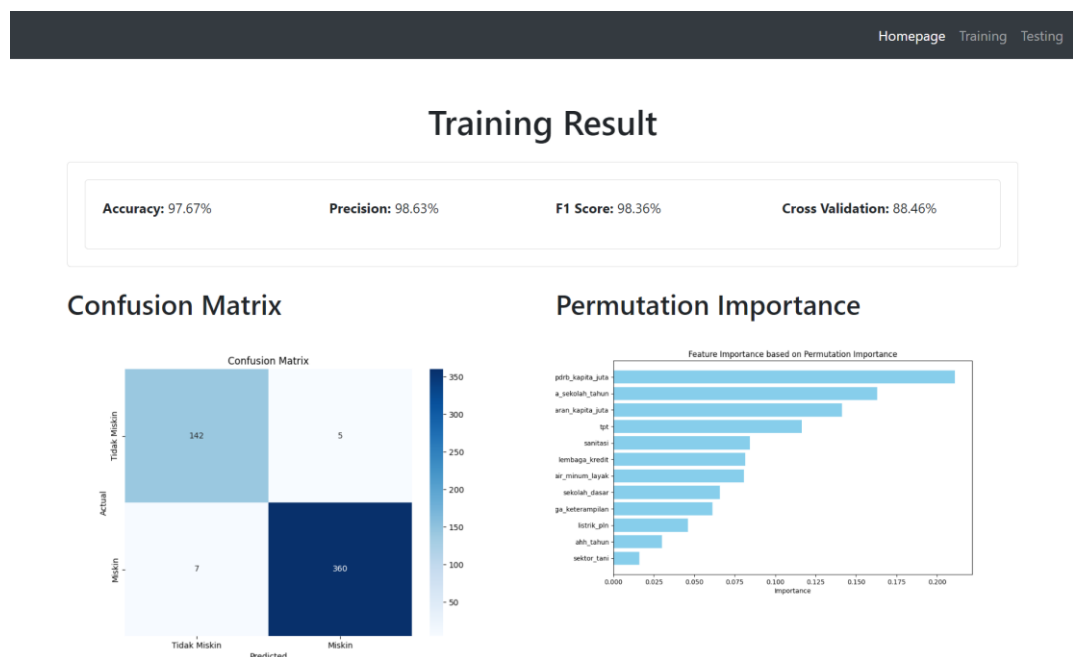




**Gambar 4. 2** Tampilan *Halaman Training*

### 3. Tampilan Halaman Hasil *Training*

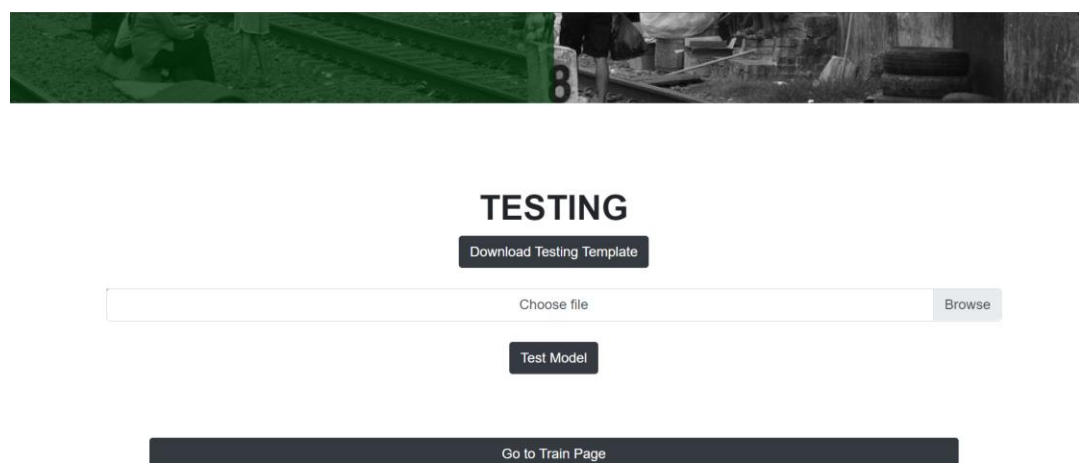
Halaman hasil *training* akan muncul saat user sudah menekan tombol “Upload” pada halaman training setelah menginput data yang akan diolah dan proses olah data oleh model telah selesai. Hasil dari proses tersebut berisi *Classification Report* (nilai *Accuracy*, *Precision*, *F1*), *Cross Validation* dari proses *training* yang telah dilakukan, *Confusion Matrix* dan *Permutation Importance* seperti yang sudah dijelaskan di Bab 3. Tampilan hasil training dapat dilihat pada gambar 4.3.



**Gambar 4. 3** Tampilan Halaman Hasil *Training*

#### 4. Tampilan Halaman *Testing*

Hampir sama dengan halaman training, halaman testing juga mewajibkan user untuk mengunggah dataset yang telah disesuaikan dengan template, namun tanpa kolom target, sesuai dengan fungsi testing itu sendiri. Tampilan ini memiliki judul halaman, tombol untuk *Download Testing Template* dan “*Test Model*” untuk menjalankan model. Tombol “*Test Model*” dapat ditekan setelah user sudah menyematkan dataset yang kemudian proses *testing* akan berjalan dan mengeluarkan *output* terkait hasil klasifikasi.

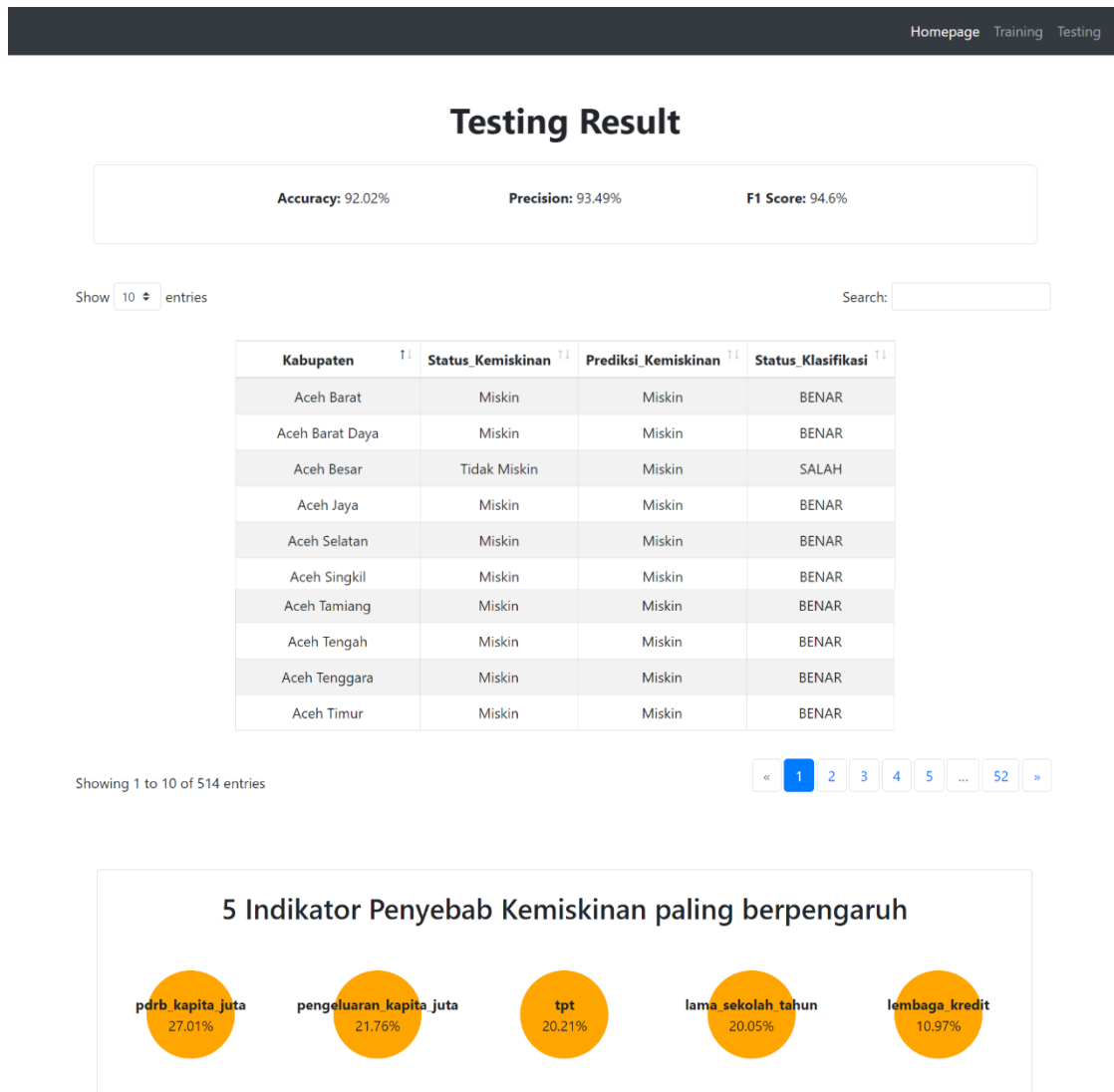


**Gambar 4. 4** Tampilan Halaman *Testing*

#### 5. Tampilan Halaman Hasil *Testing*

Tampilan halaman hasil testing akan muncul ketika proses *testing* telah selesai dijalankan. Halaman ini akan menampilkan tabel berisi nilai prediksi model akan klasifikasi kemiskinan, seluruh variabel yang diprediksi mampu menyebabkan kemiskinan per kabupaten di Indonesia, serta status apakah prediksi tersebut benar atau salah.

Selanjutnya, *user* juga dapat menampilkan dan menentukan entries data dan *pagination*. Fitur *search* juga tersedia di halaman ini untuk memudahkan *user* mencari kata kunci yang diinginkan, contohnya saat *user* ingin melihat status kemiskinan menurut kata kunci Kabupaten.



**Gambar 4. 5** Tampilan Halaman *Hasil Testing*

## 4.2. Evaluasi *Training Model*

*Training Model* dilakukan dengan cara melatih sebanyak 514 data *training* sekaligus data validation yang berisikan fitu indikator penyebab kemiskinan di Indonesia. Pembagian data ini dapat dilihat pada Tabel 3.14.

Pelatihan model dilakukan setelah data melewati tahap *preprocessing* yang terdiri dari *data splitting*, *handling missing value*, *data transformation*, *data binning* dan *encoding categorical variables*. Pelatihan dengan model *Decision Tree C4.5* ini menghasilkan *Confusion Matrix* yang memberikan informasi mengenai perbandingan antara nilai aktual dengan nilai prediksi dengan nilai *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*. Hasil dari *Confusion Matrix* tersebut dapat dilihat pada Tabel 4.2 dibawah ini.

**Tabel 4. 1** Hasil *Confusion Matrix*

	Tidak Miskin	Miskin
Tidak Miskin	142 (TP)	5 (FN)
Miskin	7 (FP)	360 (TN)

Dari tabel tersebut, dapat diketahui bahwa terdapat 142 data dengan kelas yang terprediksi Tidak Miskin dan benar serta 7 data yang terprediksi Tidak Miskin dan salah. Terdapat pula 360 data dengan kelas yang terprediksi Miskin dengan benar serta 5 data yang terprediksi Miskin dan salah. Melalui perhitungan *Confusion Matrix* tersebut, bisa dihitung nilai dari *Accuracy*, *Precision*, *F1-Score* dan *Specitifity* secara manual. Adapun perhitungannya adalah sebagai berikut :

$$Accuracy = \frac{TP+TN+TNet}{TP+TN+FN+FP} \times 100\% = \frac{142+360}{142+360+5+7} \times 100\% = 97.67\%$$

$$Precision = \frac{TP}{TP+FP} \times 100\% = \frac{142}{142+7} \times 100\% = 98.63\%$$

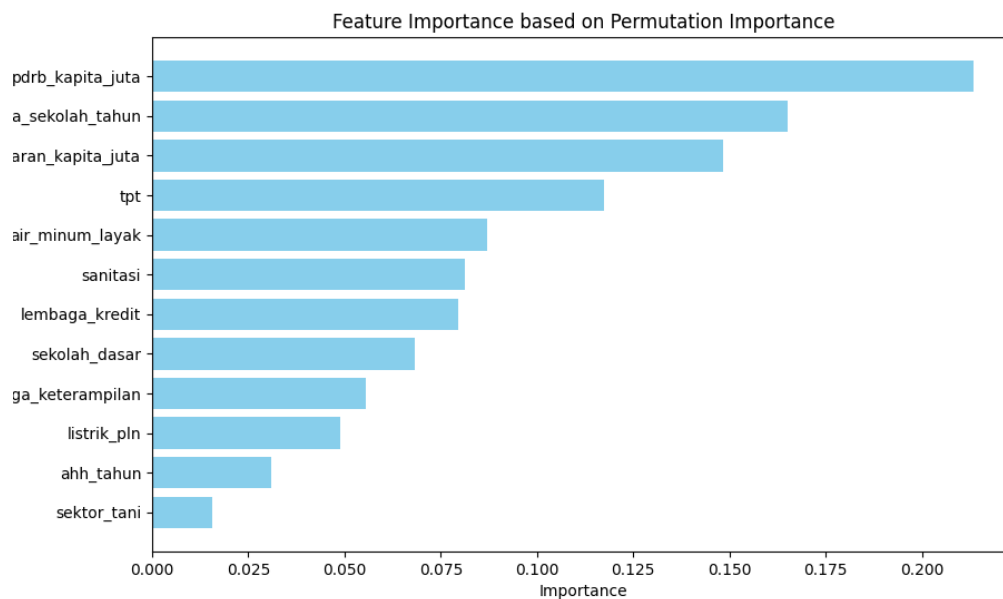
$$Recall = \frac{TP}{TP+FN} \times 100\% = \frac{142}{142+5} \times 100\% = 96.59\%$$

$$F1 - Score = \frac{2 \times recall \times precision}{recall+precision} = \frac{2 \times 0.986 \times 0.966}{0.986+0.966} \times 100\% = 98.36\%$$

Sedangkan, untuk perhitungan *specificity* adalah sebagai berikut :

$$Specificity = \frac{TN}{TN+FP} \times 100\% = \frac{360}{360+7} \times 100\% = 98.03\%$$

Lebih lanjut, dari pelatihan ini dilakukan juga Evaluasi Permutation Matrix pada Model ini guna melihat pengaruh-pengaruh dari setiap variabel terhadap status kemiskinan di suatu daerah. Hasil dari perhitungan pengaruh nilai permutasi terhadap klasifikasi ini dapat dilihat pada Gambar 4.6.



**Gambar 4. 6** *Permutation Importance Training Model*

Dari Gambar 4.6 di atas, secara keseluruhan dapat dilihat bahwasanya faktor ekonomi (PDRB per kapita, lembaga kredit), pendidikan (rata-rata tahun sekolah, sekolah dasar), infrastruktur (sanitasi, listrik, air minum layak), dan indikator sosial-ekonomi lainnya (tingkat pengangguran, pelatihan vokasi, sektor pertanian) semuanya signifikan dan relevan dalam menentukan performa model. Metode permutation importance telah mengidentifikasi fitur-fitur yang paling kritis terhadap akurasi prediksi model, dengan "**pdrb\_kapita\_juta**" sebagai yang paling berpengaruh.

#### **4.3.Evaluasi Testing Model**

Testing Model dilakukan dengan menggunakan 514 data indikator penyebab kemiskinan di Indonesia yang dimana merupakan data tahun 2022. Sistem ini mengklasifikasikan status kemiskinan setiap daerah ke dalam 2 kelas yakni Miskin dan Tidak Miskin. Contoh dari hasil pengujian sistem dapat dilihat pada Tabel 4.3.

**Tabel 4. 2** Contoh Hasil *Testing Model*

Kabupaten / Kota	Status Kemiskinan	Klasifikasi Kemiskinan	Status Klasifikasi
Aceh Barat	Miskin	Miskin	BENAR

Aceh Barat Daya	Miskin	Miskin	BENAR
Aceh Besar	Tidak Miskin	Miskin	SALAH
Aceh Jaya	Miskin	Miskin	BENAR
Aceh Selatan	Miskin	Miskin	BENAR
Aceh Singkil	Miskin	Miskin	BENAR
Aceh Tamiang	Miskin	Miskin	BENAR
Aceh Tengah	Miskin	Miskin	BENAR
Aceh Tenggara	Miskin	Miskin	BENAR
Aceh Timur	Miskin	Miskin	BENAR

Tabel di atas merupakan 10 dari hasil tahap *testing*. Hasil lengkap *Testing Model* tersebut dapat dilihat pada lampiran penelitian. Berdasarkan perhitungan tersebut, didapati nilai dari *Accuracy*: 92.02%, *Precision*: 93.49% dan *F1 Score*: 94.6%, yang dimana *Testing Model* ini berhasil mengklasifikasikan data secara benar sebanyak 473 data, dan 41 data secara salah. Hal ini diperkirakan bisa disebabkan oleh berbagai faktor seperti terdapat beberapa nilai pada variabel yang kosong dan faktor lainnya. Namun, nilai dari *F1 Score* ini dapat dianggap tinggi

Pada tahap *pre-processing* penelitian ini, dataset akan melakukan pengisian data kosong menggunakan nilai lain yang diambil dari *Mean* pada dataset. Sebagai bahan pembnding, dilakukan pengisian data kosong menggunakan teknik KNN Imputer untuk mencari k tetangga terdekat. Adapun potongan *pseudo code* dari proses ini adalah :

```

FUNCTION handle_missing_values_knn(data, miss_cols, k):
    imputer = CREATE INSTANCE OF KNNImputer WITH n_neighbors=k
    data[miss_cols] = imputer.fit_transform(data[miss_cols])
    RETURN data

```

Berdasarkan penggunaan Teknik tersebut, didapati nilai *F1 Score* pada *Testing Model* adalah sebesar 88.33% yang dimana hasil ini berubah dari nilai sebelumnya yakni sebesar 92.02%. Hal ini membuktikan bahwa variabel yang kosong pada data memberikan pengaruh dalam kinerja dari Model ini, dan juga penanganan data kosong menggunakan Teknik sudah cukup baik, namun memerlukan eksplorasi lebih lanjut dalam menentukan teknik *pre-processing* maupun training model guna meningkatkan kinerja model.



**Gambar 4. 7** *Permutation Importance Testing Model*

Dari Gambar 4.7 di atas, dapat disimpulkan bahwa indikator PDRB per Kapita merupakan faktor paling signifikan dalam memprediksi status kemiskinan suatu daerah. Ditemukan bahwa semakin rendah nilai PDRB per Kapita di suatu daerah, maka cenderung memiliki tingkat kemiskinan yang lebih tinggi. Hal ini sejalan dengan beberapa penelitian sebelumnya, seperti yang dilakukan oleh Permana dan Pasaribu (2023) yang menunjukkan bahwa semakin tinggi PDRB regional, semakin rendah tingkat kemiskinan di Pulau Sumatera. Temuan serupa juga didukung oleh Pramesona (2021), Ritonga dan Wulantika (2020) melalui penelitian yang berbeda.

Adapun Tingkat Pengangguran Terbuka menjadi salah satu Indikator yang memberikan pengaruh cukup tinggi dalam memprediksi status kemiskinan suatu daerah. Hal ini sejalan dengan penelitian yang dilakukan oleh Rohmah dan Prakoso (2022) dimana melakukan Analisis Regresi terhadap pengaruh Tingkat Pengangguran Terbuka terhadap Kemiskinan dimana variabel Tingkat Pengangguran Terbuka secara simultan berpengaruh terhadap kemiskinan yang dimana rendahnya Tingkat produktivitas menjadi salah satu penjelasan dalam korelasi antara pengangguran dan kemiskinan.

Lebih lanjut disimpulkan juga bahwa Indikator Jumlah Desa yang Memiliki Lembaga Perkreditan dan Rata-rata Lama Sekolah juga menjadi variabel paling berkontribusi dalam pemodelan ini. Hal ini didukung juga dengan hasil penelitian yang dilakukan oleh Wenas (2021) yang menyimpulkan bahwa dengan pengolahan data menggunakan Variable Importance Measure, memperoleh tiga variabel yang berkontribusi besar pada pemodelan yakni Tingkat Pengangguran Terbuka, Rata-rata lama sekolah dan persentase penduduk pengguna ponsel.

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Dari keseluruhan pelaksanaan penelitian ini, adapun kesimpulan yang didapat dari klasifikasi dan Identifikasi Indikator penyebab Kemiskinan menggunakan Algoritma *Decision Tree* C4.5 adalah sebagai berikut :

1. Penggunaan Algoritma *Decision Tree* C4.5 mampu mengklasifikasikan status kemiskinan dengan menghasilkan performa dengan nilai Akurasi sebesar 92.02% dan nilai F1-Score sebesar 94.6%.
2. Berdasarkan hasil *Permutation Matrix* dalam mengidentifikasi Indikator yang paling berpengaruh dalam penentuan kemiskinan di Indonesia, secara keseluruhan seluruh indikator cukup signifikan dan relevan dalam menentukan performa model. Metode *permutation importance* telah mengidentifikasi fitur-fitur yang paling kritis terhadap akurasi prediksi model, dimana Faktor Pendapatan Daerah (Indikator PDRB per Kapita, Pengeluaran per Kapita), Faktor Pendidikan (Rata-rata Lama Sekolah), Faktor Pekerjaan (Tingkat Pengangguran Terbuka) dan Faktor Akses terhadap Pelayanan dan Infrastruktur Dasar (Jumlah Desa yang memiliki Lembaga Kredit) menjadi 5 Indikator yang memberikan pengaruh paling besar dalam penentuan kemiskinan di Indonesia.

#### 5.2. Saran

Saran dari penulis yang dapat dijadikan sebagai catatan untuk penelitian sebelumnya adalah :

1. Perlu dilakukan penelitian lebih lanjut terkait pilihan algoritma yang optimal untuk mengklasifikasikan kemiskinan di Indonesia.
2. Memperluas dataset dengan indikator-indikator lain yang mungkin memiliki pengaruh lebih signifikan terhadap kemiskinan di Indonesia.
3. Teknik-teknik seperti *boosting*, *resampling*, dan teknik tambahan lainnya pada tahap *pre-processing* dapat dieksplorasi untuk meningkatkan kualitas data sebelum dilakukan pengolahan lebih lanjut.



## DAFTAR PUSTAKA

- Abu, A., Hamdan, R., & Sani, N. S. (2020). Ensemble Learning for Multidimensional Poverty Classification. *Sains Malaysiana*, 49(2), 447–459.
- Alyad Ulya Iman. (2021). *Analisis Penentuan Kelayakan Mustahik Zakat Menggunakan Algoritma C5.0*. <https://repositori.usu.ac.id/handle/123456789/47016>
- Anggit, P. Y., & Arianti, F. (2012). Pengangguran, Pendidikan, dan Kesehatan Terhadap Kemiskinan. *Jurnal Ekonomi Pembangunan*, 1(1), 1–8.
- Ayu, D., & Ayu, A. A. (2012). Sektor Informal, Pengangguran, dan Kemiskinan Di Provinsi Bali Tahun 2004-2010. *Jurnal Ekonomi Pembangunan, Fakultas Ekonomi Universitas Udayana*, 29–38.
- Badan Pusat Statistik. (2020). Kemiskinan dan Ketimpangan. <https://www.bps.go.id/Subject/23/Kemiskinan-Dan-Ketimpangan.html>.
- Frank, E., Wang, Y., Inglis, S., Holmes, G., & Witten, I. H. (1998). Using Model Trees for Classification. *Machine Learning*, 32, 63–76.
- Garg, S. (2021). An evaluation of investor acceptability for physical gold using classification (Decision Tree). *Materials Today: Proceedings*, 37, 950–954.
- Kurniawan, I., & Saputra, R. A. (2017). Penerapan Algoritma C5. 0 Pada Sistem Pendukung Keputusan Kelayakan Penerimaan Beras Masyarakat Miskin. *Jurnal Informatika*, 4(2).
- Mahardiki, D., & Santoso, R. P. (2013). Analisis perubahan ketimpangan pendapatan dan pertumbuhan ekonomi antar propinsi di indonesia 2006-2011. *JEJAK*, 6(2).
- Ndulu, B., Kritzinger-van Niekerk, L., & Reinikka, R. (2005). Infrastructure, regional integration and growth in Sub-Saharan Africa. *Africa in the World Economy*, 101, 101–121.
- Nurkse, R. (1971). The theory of development and the idea of balanced growth. *Developing the Underdeveloped Countries*, 115–128.
- Nurwati, N. (2008). Kemiskinan: Model pengukuran, permasalahan dan alternatif kebijakan. *Jurnal Kependudukan Padjadjaran*, 10(1), 1.
- Permana, H., & Pasaribu, E. (2023). Pengaruh Inflasi, IPM, UMP dan PDRB terhadap Kemiskinan di Pulau Sumatera. *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*, 7(3), 1113-1132.
- Quinlan, J. R. (2004). Data mining tools See5 and C5. 0. <http://www.rulequest.com/See5-Info.html>.

- Ramadhani, W. A., Irawati, N., & Maulana, C. (2022). Penerapan Metode Multifactor Evaluation Process (MFEP) Untuk Menentukan Kelayakan Penerima Bantuan Pinjaman Modal Usaha Kecil Menengah. *Building of Informatics, Technology and Science (BITS)*, 4(1), 50–59.
- Rerung, R. R. (2018). Penerapan data mining dengan memanfaatkan metode association rule untuk promosi produk. *J. Teknol. Rekayasa*, 3(1), 89.
- Rohmah, I. S. A., & Prakoso, J. A. (2022). Pengaruh Ipm, Rls, Tpt, Dan Pengeluaran Perkapita Terhadap Kemiskinan Di Provinsi Jawa Barat. *Transekonomika: Akuntansi, Bisnis Dan Keuangan*, 2(6), 255-266.
- Rojas M. (2015). Poverty and people's wellbeing. *Global Handbook of Quality of Life: Exploration of Well-Being of Nations and Continents*, 317–350.
- Schapire, R. E. (2013). Explaining Adaboost. *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, 37–52.
- Sukwika, T. (2018). Peran pembangunan infrastruktur terhadap ketimpangan ekonomi antarwilayah di Indonesia. *Jurnal Wilayah Dan Lingkungan*, 6(2), 115–130.
- Vulandari, R. T. (2017). Data Mining Teori dan Aplikasi Rapidminer. Yogyakarta: Gava Media.
- Wenas, F. (2020). *Pengklasifikasian Kemiskinan di Indonesia dengan Menggunakan Metode Algoritma Decision Tree C5.0*.
- World Bank. (2000). *World development report 2000/2001: Attacking poverty*. The World Bank.
- World Bank. (2006). *Making the New Indonesia Work for the Poor*. World Bank Office Jakarta. <https://books.google.co.id/books?id=uuzsAAAAMAAJ>
- World Bank. (2018). *Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle*. World Bank Publications. <https://books.google.co.id/books?id=P-uADwAAQBAJ>
- Xuanyuan, S., Xuanyuan, S., & Yue, Y. (2022). Application of C4. 5 Algorithm in Insurance and Financial Services Using Data Mining Methods. *Mobile Information Systems*, 2022(1), 5670784.
- Zhang, X., & Sun, Y. (2018). Breast cancer risk prediction model based on C5.0 algorithm for postmenopausal women. *2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, 321–325. <https://doi.org/10.1109/SPAC46244.2018.8965528>

## LAMPIRAN

Lampiran 1. Tabel hasil *Testing Model*

Kabupaten / Kota	Status Kemiskinan	Klasifikasi Kemiskinan	Status Klasifikasi
Aceh Barat	Miskin	Miskin	BENAR
Aceh Barat Daya	Miskin	Miskin	BENAR
Aceh Besar	Tidak Miskin	Miskin	SALAH
Aceh Jaya	Miskin	Miskin	BENAR
Aceh Selatan	Miskin	Miskin	BENAR
Aceh Singkil	Miskin	Miskin	BENAR
Aceh Tamiang	Miskin	Miskin	BENAR
Aceh Tengah	Miskin	Miskin	BENAR
Aceh Tenggara	Miskin	Miskin	BENAR
Aceh Timur	Miskin	Miskin	BENAR
Aceh Utara	Miskin	Miskin	BENAR
Agam	Miskin	Miskin	BENAR
Alor	Miskin	Miskin	BENAR
Asahan	Miskin	Miskin	BENAR
Asmat	Tidak Miskin	Tidak Miskin	BENAR
Badung	Miskin	Miskin	BENAR
Balangan	Miskin	Miskin	BENAR
Bandung	Miskin	Miskin	BENAR
Bandung Barat	Tidak Miskin	Tidak Miskin	BENAR
Banggai	Miskin	Miskin	BENAR
Banggai Kepulauan	Miskin	Miskin	BENAR
Banggai Laut	Miskin	Tidak Miskin	SALAH
Bangka	Tidak Miskin	Tidak Miskin	BENAR
Bangka Barat	Miskin	Miskin	BENAR
Bangka Selatan	Miskin	Miskin	BENAR
Bangka Tengah	Miskin	Miskin	BENAR
Bangkalan	Miskin	Miskin	BENAR
Bangli	Miskin	Miskin	BENAR
Banjar	Miskin	Miskin	BENAR
Banjarnegara	Miskin	Miskin	BENAR
Bantaeng	Miskin	Miskin	BENAR
Bantul	Miskin	Miskin	BENAR
Banyu Asin	Miskin	Miskin	BENAR
Banyumas	Miskin	Miskin	BENAR
Banyuwangi	Tidak Miskin	Tidak Miskin	BENAR

Barito Kuala	Miskin	Miskin	BENAR
Barito Selatan	Miskin	Miskin	BENAR
Barito Timur	Miskin	Miskin	BENAR
Barito Utara	Miskin	Miskin	BENAR
Barru	Miskin	Miskin	BENAR
Batang	Miskin	Miskin	BENAR
Batang Hari	Miskin	Miskin	BENAR
Batu Bara	Miskin	Miskin	BENAR
Bekasi	Tidak Miskin	Tidak Miskin	BENAR
Belitung	Tidak Miskin	Tidak Miskin	BENAR
Belitung Timur	Miskin	Miskin	BENAR
Belu	Miskin	Miskin	BENAR
Bener Meriah	Miskin	Miskin	BENAR
Bengkalis	Tidak Miskin	Tidak Miskin	BENAR
Bengkayang	Miskin	Miskin	BENAR
Bengkulu Selatan	Miskin	Miskin	BENAR
Bengkulu Tengah	Miskin	Miskin	BENAR
Bengkulu Utara	Miskin	Miskin	BENAR
Berau	Tidak Miskin	Tidak Miskin	BENAR
Biak Numfor	Tidak Miskin	Tidak Miskin	BENAR
Bima	Miskin	Miskin	BENAR
Bintan	Miskin	Miskin	BENAR
Bireuen	Miskin	Miskin	BENAR
Blitar	Miskin	Miskin	BENAR
Blora	Miskin	Miskin	BENAR
Boalemo	Miskin	Miskin	BENAR
Bogor	Tidak Miskin	Tidak Miskin	BENAR
Bojonegoro	Miskin	Miskin	BENAR
Bolaang Mongondow	Miskin	Miskin	BENAR
Bolaang Mongondow Selatan	Miskin	Miskin	BENAR
Bolaang Mongondow Timur	Tidak Miskin	Miskin	SALAH
Bolaang Mongondow Utara	Miskin	Miskin	BENAR
Bombana	Miskin	Miskin	BENAR
Bondowoso	Miskin	Miskin	BENAR
Bone	Miskin	Miskin	BENAR
Bone Bolango	Miskin	Miskin	BENAR
Boven Digoel	Tidak Miskin	Tidak Miskin	BENAR
Boyolali	Miskin	Miskin	BENAR

Brebes	Miskin	Miskin	BENAR
Buleleng	Miskin	Miskin	BENAR
Bulukumba	Miskin	Miskin	BENAR
Bulungan	Miskin	Miskin	BENAR
Bungo	Miskin	Miskin	BENAR
Buol	Tidak Miskin	Tidak Miskin	BENAR
Buru	Miskin	Miskin	BENAR
Buru Selatan	Tidak Miskin	Tidak Miskin	BENAR
Buton	Miskin	Miskin	BENAR
Buton Selatan	Miskin	Miskin	BENAR
Buton Tengah	Miskin	Tidak Miskin	SALAH
Buton Utara	Miskin	Miskin	BENAR
Ciamis	Tidak Miskin	Tidak Miskin	BENAR
Cianjur	Miskin	Miskin	BENAR
Cilacap	Miskin	Miskin	BENAR
Cirebon	Miskin	Miskin	BENAR
Dairi	Miskin	Miskin	BENAR
Deiyai	Tidak Miskin	Tidak Miskin	BENAR
Deli Serdang	Tidak Miskin	Tidak Miskin	BENAR
Demak	Miskin	Miskin	BENAR
Dharmasraya	Miskin	Miskin	BENAR
Dogiyai	Tidak Miskin	Tidak Miskin	BENAR
Dompu	Miskin	Miskin	BENAR
Donggala	Miskin	Miskin	BENAR
Empat Lawang	Miskin	Miskin	BENAR
Ende	Miskin	Miskin	BENAR
Enrekang	Miskin	Miskin	BENAR
Fakfak	Tidak Miskin	Tidak Miskin	BENAR
Flores Timur	Tidak Miskin	Tidak Miskin	BENAR
Garut	Miskin	Miskin	BENAR
Gayo Lues	Miskin	Miskin	BENAR
Gianyar	Tidak Miskin	Tidak Miskin	BENAR
Gorontalo	Miskin	Miskin	BENAR
Gorontalo Utara	Miskin	Miskin	BENAR
Gowa	Miskin	Miskin	BENAR
Gresik	Tidak Miskin	Tidak Miskin	BENAR
Grobogan	Miskin	Miskin	BENAR
Gunung Kidul	Miskin	Miskin	BENAR
Gunung Mas	Miskin	Miskin	BENAR
Halmahera Barat	Tidak Miskin	Tidak Miskin	BENAR
Halmahera Selatan	Miskin	Miskin	BENAR

Halmahera Tengah	Miskin	Miskin	BENAR
Halmahera Timur	Tidak Miskin	Tidak Miskin	BENAR
Halmahera Utara	Miskin	Tidak Miskin	SALAH
Hulu Sungai Selatan	Miskin	Miskin	BENAR
Hulu Sungai Tengah	Miskin	Miskin	BENAR
Hulu Sungai Utara	Miskin	Miskin	BENAR
Humbang Hasundutan	Miskin	Miskin	BENAR
Indragiri Hilir	Tidak Miskin	Miskin	SALAH
Indragiri Hulu	Tidak Miskin	Tidak Miskin	BENAR
Indramayu	Miskin	Miskin	BENAR
Intan Jaya	Tidak Miskin	Tidak Miskin	BENAR
Jayapura	Tidak Miskin	Miskin	SALAH
Jayawijaya	Tidak Miskin	Tidak Miskin	BENAR
Jember	Miskin	Miskin	BENAR
Jembrana	Miskin	Miskin	BENAR
Jeneponto	Miskin	Miskin	BENAR
Jepara	Miskin	Miskin	BENAR
Jombang	Miskin	Miskin	BENAR
Kaimana	Tidak Miskin	Tidak Miskin	BENAR
Kampar	Tidak Miskin	Tidak Miskin	BENAR
Kapuas	Miskin	Miskin	BENAR
Kapuas Hulu	Miskin	Miskin	BENAR
Karang Asem	Miskin	Miskin	BENAR
Karanganyar	Miskin	Miskin	BENAR
Karawang	Miskin	Miskin	BENAR
Karimun	Miskin	Miskin	BENAR
Karo	Miskin	Miskin	BENAR
Katingan	Tidak Miskin	Miskin	SALAH
Kaur	Miskin	Miskin	BENAR
Kayong Utara	Miskin	Miskin	BENAR
Kebumen	Miskin	Miskin	BENAR
Kediri	Miskin	Miskin	BENAR
Keerom	Tidak Miskin	Tidak Miskin	BENAR
Kendal	Miskin	Miskin	BENAR
Kepahiang	Miskin	Miskin	BENAR
Kepulauan Anambas	Miskin	Tidak Miskin	SALAH
Kepulauan Aru	Miskin	Miskin	BENAR
Kepulauan Mentawai	Miskin	Miskin	BENAR
Kepulauan Meranti	Miskin	Miskin	BENAR
Kepulauan Sangihe	Miskin	Miskin	BENAR
Kepulauan Selayar	Miskin	Miskin	BENAR

Kepulauan Seribu	Miskin	Miskin	BENAR
Kepulauan Sula	Miskin	Miskin	BENAR
Kepulauan Talaud	Miskin	Miskin	BENAR
Kepulauan Yapen	Tidak Miskin	Tidak Miskin	BENAR
Kerinci	Miskin	Miskin	BENAR
Ketapang	Miskin	Miskin	BENAR
Klaten	Miskin	Miskin	BENAR
Klungkung	Miskin	Miskin	BENAR
Kolaka	Miskin	Miskin	BENAR
Kolaka Timur	Miskin	Miskin	BENAR
Kolaka Utara	Miskin	Miskin	BENAR
Konawe	Miskin	Miskin	BENAR
Konawe Kepulauan	Tidak Miskin	Tidak Miskin	BENAR
Konawe Selatan	Miskin	Miskin	BENAR
Konawe Utara	Miskin	Miskin	BENAR
Kota Ambon	Tidak Miskin	Tidak Miskin	BENAR
Kota Balikpapan	Tidak Miskin	Tidak Miskin	BENAR
Kota Banda Aceh	Tidak Miskin	Tidak Miskin	BENAR
Kota Bandar Lampung	Miskin	Miskin	BENAR
Kota Bandung	Tidak Miskin	Tidak Miskin	BENAR
Kota Banjar	Tidak Miskin	Tidak Miskin	BENAR
Kota Banjar Baru	Miskin	Tidak Miskin	SALAH
Kota Banjarmasin	Miskin	Miskin	BENAR
Kota Baru	Tidak Miskin	Miskin	SALAH
Kota Batam	Miskin	Miskin	BENAR
Kota Batu	Miskin	Miskin	BENAR
Kota Baubau	Miskin	Miskin	BENAR
Kota Bekasi	Tidak Miskin	Tidak Miskin	BENAR
Kota Bengkulu	Miskin	Miskin	BENAR
Kota Bima	Miskin	Miskin	BENAR
Kota Binjai	Tidak Miskin	Miskin	SALAH
Kota Bitung	Miskin	Miskin	BENAR
Kota Blitar	Tidak Miskin	Tidak Miskin	BENAR
Kota Bogor	Tidak Miskin	Tidak Miskin	BENAR
Kota Bontang	Tidak Miskin	Tidak Miskin	BENAR
Kota Bukittinggi	Miskin	Miskin	BENAR
Kota Cilegon	Miskin	Miskin	BENAR
Kota Cimahi	Miskin	Miskin	BENAR
Kota Cirebon	Miskin	Miskin	BENAR
Kota Denpasar	Tidak Miskin	Miskin	SALAH
Kota Depok	Tidak Miskin	Miskin	SALAH

Kota Dumai	Tidak Miskin	Tidak Miskin	BENAR
Kota Gorontalo	Miskin	Miskin	BENAR
Kota Gunungsitoli	Miskin	Miskin	BENAR
Kota Jakarta Barat	Tidak Miskin	Tidak Miskin	BENAR
Kota Jakarta Pusat	Tidak Miskin	Tidak Miskin	BENAR
Kota Jakarta Selatan	Tidak Miskin	Tidak Miskin	BENAR
Kota Jakarta Timur	Tidak Miskin	Tidak Miskin	BENAR
Kota Jakarta Utara	Tidak Miskin	Tidak Miskin	BENAR
Kota Jambi	Tidak Miskin	Tidak Miskin	BENAR
Kota Jayapura	Tidak Miskin	Miskin	SALAH
Kota Kediri	Tidak Miskin	Tidak Miskin	BENAR
Kota Kendari	Miskin	Miskin	BENAR
Kota Kotamobagu	Tidak Miskin	Tidak Miskin	BENAR
Kota Kupang	Miskin	Miskin	BENAR
Kota Langsa	Tidak Miskin	Tidak Miskin	BENAR
Kota Lhokseumawe	Tidak Miskin	Tidak Miskin	BENAR
Kota Lubuklinggau	Tidak Miskin	Tidak Miskin	BENAR
Kota Madiun	Miskin	Miskin	BENAR
Kota Magelang	Tidak Miskin	Tidak Miskin	BENAR
Kota Makassar	Miskin	Miskin	BENAR
Kota Malang	Miskin	Miskin	BENAR
Kota Manado	Miskin	Miskin	BENAR
Kota Mataram	Miskin	Miskin	BENAR
Kota Medan	Tidak Miskin	Miskin	SALAH
Kota Metro	Miskin	Miskin	BENAR
Kota Mojokerto	Miskin	Miskin	BENAR
Kota Padang	Miskin	Miskin	BENAR
Kota Padang Panjang	Miskin	Miskin	BENAR
Kota Padangsidimpuan	Tidak Miskin	Tidak Miskin	BENAR
Kota Pagar Alam	Miskin	Miskin	BENAR
Kota Palangka Raya	Miskin	Miskin	BENAR
Kota Palembang	Miskin	Miskin	BENAR
Kota Palopo	Miskin	Miskin	BENAR
Kota Palu	Miskin	Miskin	BENAR
Kota Pangkal Pinang	Miskin	Miskin	BENAR
Kota Parepare	Miskin	Miskin	BENAR
Kota Pariaman	Miskin	Miskin	BENAR
Kota Pasuruan	Miskin	Miskin	BENAR
Kota Payakumbuh	Miskin	Miskin	BENAR
Kota Pekalongan	Miskin	Miskin	BENAR
Kota Pekanbaru	Tidak Miskin	Tidak Miskin	BENAR



Kota Pematang Siantar	Miskin	Miskin	BENAR
Kota Pontianak	Miskin	Miskin	BENAR
Kota Prabumulih	Tidak Miskin	Tidak Miskin	BENAR
Kota Probolinggo	Miskin	Miskin	BENAR
Kota Sabang	Miskin	Miskin	BENAR
Kota Salatiga	Tidak Miskin	Tidak Miskin	BENAR
Kota Samarinda	Tidak Miskin	Tidak Miskin	BENAR
Kota Sawah Lunto	Miskin	Miskin	BENAR
Kota Semarang	Tidak Miskin	Tidak Miskin	BENAR
Kota Serang	Tidak Miskin	Tidak Miskin	BENAR
Kota Sibolga	Tidak Miskin	Tidak Miskin	BENAR
Kota Singkawang	Tidak Miskin	Tidak Miskin	BENAR
Kota Solok	Tidak Miskin	Tidak Miskin	BENAR
Kota Sorong	Miskin	Miskin	BENAR
Kota Subulussalam	Miskin	Miskin	BENAR
Kota Sukabumi	Tidak Miskin	Tidak Miskin	BENAR
Kota Sungai Penuh	Miskin	Miskin	BENAR
Kota Surabaya	Tidak Miskin	Tidak Miskin	BENAR
Kota Surakarta	Tidak Miskin	Tidak Miskin	BENAR
Kota Tangerang	Tidak Miskin	Tidak Miskin	BENAR
Kota Tangerang Selatan	Tidak Miskin	Tidak Miskin	BENAR
Kota Tanjung Balai	Tidak Miskin	Tidak Miskin	BENAR
Kota Tanjung Pinang	Tidak Miskin	Tidak Miskin	BENAR
Kota Tarakan	Tidak Miskin	Tidak Miskin	BENAR
Kota Tasikmalaya	Tidak Miskin	Tidak Miskin	BENAR
Kota Tebing Tinggi	Tidak Miskin	Tidak Miskin	BENAR
Kota Tegal	Miskin	Miskin	BENAR
Kota Ternate	Miskin	Miskin	BENAR
Kota Tidore Kepulauan	Miskin	Miskin	BENAR
Kota Tomohon	Miskin	Miskin	BENAR
Kota Tual	Tidak Miskin	Tidak Miskin	BENAR
Kota Yogyakarta	Tidak Miskin	Tidak Miskin	BENAR
Kotawaringin Barat	Tidak Miskin	Tidak Miskin	BENAR
Kotawaringin Timur	Miskin	Miskin	BENAR
Kuantan Singingi	Tidak Miskin	Tidak Miskin	BENAR
Kubu Raya	Tidak Miskin	Miskin	SALAH
Kudus	Miskin	Miskin	BENAR
Kulon Progo	Miskin	Miskin	BENAR
Kuningan	Tidak Miskin	Tidak Miskin	BENAR
Kupang	Miskin	Miskin	BENAR
Kutai Barat	Miskin	Miskin	BENAR

Kutai Kartanegara	Tidak Miskin	Tidak Miskin	BENAR
Kutai Timur	Tidak Miskin	Tidak Miskin	BENAR
Labuhan Batu	Tidak Miskin	Tidak Miskin	BENAR
Labuhan Batu Selatan	Miskin	Miskin	BENAR
Labuhan Batu Utara	Miskin	Miskin	BENAR
Lahat	Miskin	Tidak Miskin	SALAH
Lamandau	Miskin	Miskin	BENAR
Lamongan	Miskin	Miskin	BENAR
Lampung Barat	Miskin	Miskin	BENAR
Lampung Selatan	Miskin	Tidak Miskin	SALAH
Lampung Tengah	Miskin	Miskin	BENAR
Lampung Timur	Miskin	Miskin	BENAR
Lampung Utara	Miskin	Tidak Miskin	SALAH
Landak	Miskin	Miskin	BENAR
Langkat	Miskin	Miskin	BENAR
Lanny Jaya	Tidak Miskin	Tidak Miskin	BENAR
Lebak	Miskin	Miskin	BENAR
Lebong	Miskin	Miskin	BENAR
Lembata	Miskin	Miskin	BENAR
Lima Puluh Kota	Miskin	Miskin	BENAR
Lingga	Miskin	Miskin	BENAR
Lombok Barat	Miskin	Miskin	BENAR
Lombok Tengah	Tidak Miskin	Miskin	SALAH
Lombok Timur	Miskin	Miskin	BENAR
Lombok Utara	Miskin	Miskin	BENAR
Lumajang	Miskin	Miskin	BENAR
Luwu	Miskin	Miskin	BENAR
Luwu Timur	Miskin	Miskin	BENAR
Luwu Utara	Miskin	Miskin	BENAR
Madiun	Miskin	Miskin	BENAR
Magelang	Miskin	Miskin	BENAR
Magetan	Miskin	Miskin	BENAR
Mahakam Hulu	Tidak Miskin	Miskin	SALAH
Majalengka	Miskin	Miskin	BENAR
Majene	Miskin	Miskin	BENAR
Malaka	Miskin	Miskin	BENAR
Malang	Tidak Miskin	Miskin	SALAH
Malinau	Tidak Miskin	Miskin	SALAH
Maluku Barat Daya	Tidak Miskin	Tidak Miskin	BENAR
Maluku Tengah	Miskin	Miskin	BENAR
Maluku Tenggara	Tidak Miskin	Tidak Miskin	BENAR

Maluku Tenggara Barat	Tidak Miskin	Tidak Miskin	BENAR
Mamasa	Miskin	Miskin	BENAR
Mamberamo Raya	Miskin	Miskin	BENAR
Mamberamo Tengah	Tidak Miskin	Tidak Miskin	BENAR
Mamuju	Miskin	Miskin	BENAR
Mamuju Tengah	Miskin	Miskin	BENAR
Mamuju Utara	Miskin	Miskin	BENAR
Mandailing Natal	Miskin	Miskin	BENAR
Manggarai	Tidak Miskin	Tidak Miskin	BENAR
Manggarai Barat	Miskin	Miskin	BENAR
Manggarai Timur	Miskin	Miskin	BENAR
Mappi	Tidak Miskin	Tidak Miskin	BENAR
Maros	Miskin	Miskin	BENAR
Maybrat	Tidak Miskin	Miskin	SALAH
MaYeskwari	Miskin	Miskin	BENAR
MaYeskwari Selatan	Miskin	Miskin	BENAR
Melawi	Tidak Miskin	Tidak Miskin	BENAR
Merangin	Miskin	Miskin	BENAR
Merauke	Miskin	Miskin	BENAR
Mesuji	Miskin	Miskin	BENAR
Mimika	Miskin	Miskin	BENAR
Minahasa	Miskin	Miskin	BENAR
Minahasa Selatan	Miskin	Miskin	BENAR
Minahasa Tenggara	Miskin	Miskin	BENAR
Minahasa Utara	Tidak Miskin	Tidak Miskin	BENAR
Mojokerto	Miskin	Miskin	BENAR
Morowali	Miskin	Tidak Miskin	SALAH
Morowali Utara	Miskin	Miskin	BENAR
Muara Enim	Miskin	Miskin	BENAR
Muaro Jambi	Miskin	Miskin	BENAR
Mukomuko	Miskin	Miskin	BENAR
Muna	Miskin	Miskin	BENAR
Muna Barat	Miskin	Miskin	BENAR
Murung Raya	Tidak Miskin	Miskin	SALAH
Musi Banyuasin	Tidak Miskin	Tidak Miskin	BENAR
Musi Rawas	Miskin	Tidak Miskin	SALAH
Musi Rawas Utara	Tidak Miskin	Tidak Miskin	BENAR
Nabire	Tidak Miskin	Miskin	SALAH
Nagan Raya	Tidak Miskin	Miskin	SALAH
Nagekeo	Miskin	Miskin	BENAR
Natuna	Miskin	Tidak Miskin	SALAH

Nduga *	Tidak Miskin	Tidak Miskin	BENAR
Ngada	Miskin	Miskin	BENAR
Nganjuk	Miskin	Miskin	BENAR
Ngawi	Miskin	Miskin	BENAR
Nias	Miskin	Miskin	BENAR
Nias Barat	Miskin	Miskin	BENAR
Nias Selatan	Miskin	Miskin	BENAR
Nias Utara	Miskin	Miskin	BENAR
Nunukan	Miskin	Miskin	BENAR
Ogan Ilir	Miskin	Miskin	BENAR
Ogan Komering Ilir	Miskin	Miskin	BENAR
Ogan Komering Ulu	Tidak Miskin	Tidak Miskin	BENAR
Ogan Komering Ulu Selatan	Miskin	Miskin	BENAR
Ogan Komering Ulu Timur	Miskin	Miskin	BENAR
Pacitan	Miskin	Miskin	BENAR
Padang Lawas	Tidak Miskin	Tidak Miskin	BENAR
Padang Lawas Utara	Miskin	Miskin	BENAR
Padang Pariaman	Tidak Miskin	Tidak Miskin	BENAR
Pakpak Bharat	Miskin	Miskin	BENAR
Pamekasan	Miskin	Miskin	BENAR
Pandeglang	Miskin	Miskin	BENAR
Pangandaran	Miskin	Miskin	BENAR
Pangkajene Dan Kepulauan	Miskin	Miskin	BENAR
Paniai	Miskin	Tidak Miskin	SALAH
Parigi Moutong	Miskin	Miskin	BENAR
Pasaman	Miskin	Miskin	BENAR
Pasaman Barat	Miskin	Miskin	BENAR
Paser	Miskin	Miskin	BENAR
Pasuruan	Miskin	Miskin	BENAR
Pati	Miskin	Miskin	BENAR
Pegunungan Arfak	Tidak Miskin	Tidak Miskin	BENAR
Pegunungan Bintang	Tidak Miskin	Tidak Miskin	BENAR
Pekalongan	Miskin	Miskin	BENAR
Pelalawan	Tidak Miskin	Tidak Miskin	BENAR
Pemalang	Miskin	Miskin	BENAR
Penajam Paser Utara	Tidak Miskin	Tidak Miskin	BENAR
Penukal Abab Lematang Ilir	Miskin	Miskin	BENAR
Pesawaran	Miskin	Miskin	BENAR

Pesisir Barat	Miskin	Miskin	BENAR
Pesisir Selatan	Miskin	Miskin	BENAR
Pidie	Miskin	Miskin	BENAR
Pidie Jaya	Miskin	Miskin	BENAR
Pinrang	Miskin	Miskin	BENAR
Pohuwato	Miskin	Miskin	BENAR
Polewali Mandar	Miskin	Miskin	BENAR
Ponorogo	Miskin	Miskin	BENAR
Pontianak	Miskin	Miskin	BENAR
Poso	Tidak Miskin	Miskin	SALAH
Pringsewu	Miskin	Miskin	BENAR
Probolinggo	Miskin	Miskin	BENAR
Pulang Pisau	Miskin	Miskin	BENAR
Pulau Morotai	Tidak Miskin	Tidak Miskin	BENAR
Pulau Taliabu	Tidak Miskin	Tidak Miskin	BENAR
Puncak	Tidak Miskin	Tidak Miskin	BENAR
Puncak Jaya	Tidak Miskin	Tidak Miskin	BENAR
Purbalingga	Miskin	Miskin	BENAR
Purwakarta	Miskin	Miskin	BENAR
Purworejo	Miskin	Miskin	BENAR
Raja Ampat	Miskin	Miskin	BENAR
Rejang Lebong	Miskin	Miskin	BENAR
Rembang	Miskin	Miskin	BENAR
Rokan Hilir	Miskin	Miskin	BENAR
Rokan Hulu	Tidak Miskin	Tidak Miskin	BENAR
Rote Ndao	Miskin	Miskin	BENAR
Sabu Raijua	Miskin	Miskin	BENAR
Sambas	Miskin	Miskin	BENAR
Samosir	Miskin	Miskin	BENAR
Sampang	Miskin	Miskin	BENAR
Sanggau	Miskin	Miskin	BENAR
Sarmi	Miskin	Miskin	BENAR
Sarolangun	Miskin	Miskin	BENAR
Sekadau	Miskin	Miskin	BENAR
Seluma	Miskin	Miskin	BENAR
Semarang	Tidak Miskin	Miskin	SALAH
Seram Bagian Barat	Miskin	Miskin	BENAR
Seram Bagian Timur	Miskin	Tidak Miskin	SALAH
Serang	Tidak Miskin	Miskin	SALAH
Serdang Bedagai	Miskin	Miskin	BENAR
Seruyan	Tidak Miskin	Tidak Miskin	BENAR

Siak	Tidak Miskin	Tidak Miskin	BENAR
Siau Tagulandang Biaro	Miskin	Miskin	BENAR
Sidenreng Rappang	Miskin	Miskin	BENAR
Sidoarjo	Tidak Miskin	Tidak Miskin	BENAR
Sigi	Miskin	Miskin	BENAR
Sijunjung	Miskin	Miskin	BENAR
Sikka	Miskin	Miskin	BENAR
Simalungun	Miskin	Miskin	BENAR
Simeulue	Miskin	Miskin	BENAR
Sinjai	Miskin	Miskin	BENAR
Sintang	Miskin	Miskin	BENAR
Situbondo	Miskin	Miskin	BENAR
Sleman	Miskin	Miskin	BENAR
Solok	Miskin	Miskin	BENAR
Solok Selatan	Miskin	Miskin	BENAR
Soppeng	Miskin	Miskin	BENAR
Sorong	Miskin	Miskin	BENAR
Sorong Selatan	Miskin	Miskin	BENAR
Sragen	Miskin	Miskin	BENAR
Subang	Miskin	Miskin	BENAR
Sukabumi	Miskin	Miskin	BENAR
Sukamara	Miskin	Miskin	BENAR
Sukoharjo	Miskin	Miskin	BENAR
Sumba Barat	Miskin	Miskin	BENAR
Sumba Barat Daya	Miskin	Miskin	BENAR
Sumba Tengah	Miskin	Miskin	BENAR
Sumba Timur	Miskin	Miskin	BENAR
Sumbawa	Miskin	Miskin	BENAR
Sumbawa Barat	Tidak Miskin	Miskin	SALAH
Sumedang	Miskin	Miskin	BENAR
Sumenep	Miskin	Tidak Miskin	SALAH
Supiori	Miskin	Miskin	BENAR
Tabalong	Miskin	Miskin	BENAR
Tabanan	Miskin	Miskin	BENAR
Takalar	Miskin	Miskin	BENAR
Tambrau	Tidak Miskin	Tidak Miskin	BENAR
Tana Tidung	Miskin	Tidak Miskin	SALAH
Tana Toraja	Miskin	Miskin	BENAR
Tanah Bumbu	Miskin	Miskin	BENAR
Tanah Datar	Miskin	Miskin	BENAR
Tanah Laut	Miskin	Miskin	BENAR

Tangerang	Tidak Miskin	Tidak Miskin	BENAR
Tanggamus	Miskin	Miskin	BENAR
Tanjung Jabung Barat	Tidak Miskin	Tidak Miskin	BENAR
Tanjung Jabung Timur	Miskin	Miskin	BENAR
Tapanuli Selatan	Miskin	Miskin	BENAR
Tapanuli Tengah	Miskin	Miskin	BENAR
Tapanuli Utara	Miskin	Miskin	BENAR
Tapin	Miskin	Miskin	BENAR
Tasikmalaya	Miskin	Miskin	BENAR
Tebo	Miskin	Miskin	BENAR
Tegal	Miskin	Miskin	BENAR
Teluk Bintuni	Miskin	Miskin	BENAR
Teluk Wondama	Miskin	Miskin	BENAR
Temanggung	Miskin	Miskin	BENAR
Timor Tengah Selatan	Miskin	Miskin	BENAR
Timor Tengah Utara	Miskin	Miskin	BENAR
Toba Samosir	Miskin	Miskin	BENAR
Tojo Una-Una	Tidak Miskin	Tidak Miskin	BENAR
Toli-Toli	Miskin	Miskin	BENAR
Tolikara	Tidak Miskin	Tidak Miskin	BENAR
Toraja Utara	Miskin	Miskin	BENAR
Trenggalek	Miskin	Miskin	BENAR
Tuban	Miskin	Miskin	BENAR
Tulang Bawang Barat	Miskin	Miskin	BENAR
Tulangbawang	Tidak Miskin	Miskin	SALAH
Tulungagung	Miskin	Miskin	BENAR
Wajo	Miskin	Miskin	BENAR
Wakatobi	Miskin	Miskin	BENAR
Waropen	Miskin	Miskin	BENAR
Way Kanan	Miskin	Miskin	BENAR
Wonogiri	Miskin	Miskin	BENAR
Wonosobo	Miskin	Miskin	BENAR
Yahukimo	Tidak Miskin	Tidak Miskin	BENAR
Yalimo	Miskin	Tidak Miskin	SALAH