

**PENERAPAN ALGORITMA C5.0 DAN METODE SMOTE  
PADA KLASIFIKASI STATUS KESEJAHTERAAN  
RUMAH TANGGA**

**SKRIPSI**

**OLEH:  
AIDA NABILA YUMNA HANIN  
NIM. 17610028**



**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**PENERAPAN ALGORITMA C5.0 DAN METODE SMOTE  
PADA KLASIFIKASI STATUS KESEJAHTERAAN  
RUMAH TANGGA**

**SKRIPSI**

**Diajukan Kepada  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
untuk Memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh:  
AIDA NABILA YUMNA HANIN  
NIM. 17610028**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**PENERAPAN ALGORITMA C5.0 DAN METODE SMOTE  
PADA KLASIFIKASI STATUS KESEJAHTERAAN  
RUMAH TANGGA**

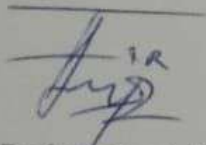
**SKRIPSI**

Oleh:  
**Aida Nabila Yumna Hanin**  
NIM. 17610028

Telah Disetujui Untuk Diuji

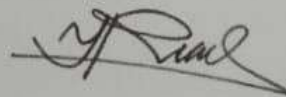
Malang, 27 Desember 2023

Dosen Pembimbing I



Dr. Fachrur Rozi, M.Si  
NIP. 19800527 200801 1 012

Dosen Pembimbing II



Erna Herawati, M.Pd  
NIP. 19760723 202321 2 006

Mengetahui,

Ketua Program Studi Matematika



Dr. Bin Susanti, M.Sc  
NIP. 19741129 200012 2 005

**PENERAPAN ALGORITMA C5.0 DAN METODE SMOTE  
PADA KLASIFIKASI STATUS KESEJAHTERAAN  
RUMAH TANGGA**

**SKRIPSI**

**Oleh:**  
**Aida Nabila Yumna Hanin**  
**NIM. 17610028**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

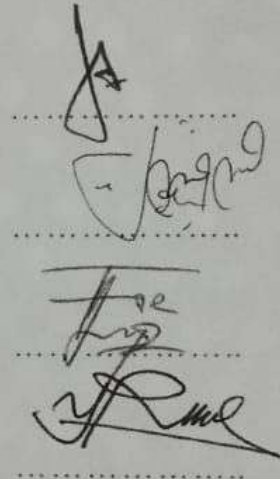
Tanggal 30 Desember 2023

Ketua Penguji : Prof. Dr. Hj. Sri Harini, M.Si.

Anggota Penguji 1 : Ria Dhea Layla Nur Karisma, M.Si.

Anggota Penguji 2 : Dr. Fachrur Rozi, M.Si.

Anggota Penguji 3 : Erna Herawati, M.Pd.



Mengetahui,

Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.

NIP. 19741129 200012 2 005

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Aida Nabila Yumna Hanin

NIM :17610028

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Judul : Penerapan Algoritma C5.0 dan Metode SMOTE pada Klasifikasi Status Kesejahteraan Rumah Tangga.

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-nenar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 30 Desember 2023

Yang Membuat Pernyataan



Aida Nabila Yumna Hanin  
17610028

## MOTO

خَيْرُ النَّاسِ أَنْفَعُهُمْ لِلنَّاسِ

“Sebaik-baiknya manusia adalah yang bermanfaat”

“Jika orang lain bisa, maka aku juga harus bisa melakukannya”

## **PERSEMBAHAN**

Skripsi ini penulis persembahkan untuk:

Kedua orang tua penulis, Bapak M. Muhajir dan Ibu Dewi Mahlikah tercinta, serta kedua adik penulis, Shinta Nafisa dan Vania Zabrina Ramadhani yang senantiasa dengan ikhlas dan istiqomah mendoakan, memberikan kasih sayang dan semangat.

## KATA PENGANTAR

*Assalaamu'alaikum Warahmatullaahi Wabarakaatuh*

Segala puji bagi Allah *Subhanahu wa ta'aala* atas rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul “Penerapan Algoritma C5.0 dan Metode *SMOTE* pada Klasifikasi Kesejahteraan Rumah Tangga”. Shalawat serta salam tercurahkan kepada junjungan Nabi kita Nabi Muhammad *Shallallahu 'alaihi wa sallam* yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang terang benderang yaitu agama Islam. Harapan penulis semoga kita tergolong sebagai orang-orang yang mendapatkan syafaat kelak pada hari kiamat, *Aamiin*.

Dalam penyelesaian proposal skripsi ini tidak dapat lepas dari bantuan berbagai pihak yang telah mendukung dan membantu secara langsung maupun tidak langsung, oleh karena itu penulis mengucapkan terima kasih yang ditujukan:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, S.Pd., M.Sc., selaku ketua Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim.
4. Dr. Fachrur Rozi, M.Si., selaku dosen pembimbing I yang telah meluangkan waktunya serta memberikan bimbingan, nasihat, serta motivasi kepada penulis.
5. Erna Herawati, M.Pd., selaku dosen pembimbing II yang telah meluangkan waktunya serta memberikan bimbingan, nasihat, serta motivasi kepada penulis.
6. Prof. Dr. Hj. Sri Harini, M.Si., selaku Ketua Penguji dalam ujian Skripsi.
7. Ria Dhea Layla Nur Karisma, M.Si., selaku Penguji 1 dalam ujian Skripsi.
8. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim.
9. M. Muhajir dan Dewi Mahlikah, selaku orang tua tercinta yang telah memberikan dukungan terbesar bagi penulis. Serta kedua adik penulis yang



selalu memberikan semangat serta dukungan untuk menyelesaikan skripsi ini.

10. Seluruh teman mahasiswa Program Studi Matematika angkatan 2017 “MAGENTA” yang rela meluangkan waktunya untuk bertukar pikiran dengan penulis.

11. Seluruh pihak yang ikut membantu dalam menyelesaikan skripsi ini, baik yang terlibat secara langsung maupun tidak secara langsung yang tidak bisa penulis sebutkan satu persatu.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi pembaca khususnya mahasiswa Program Studi Matematika.

*Wassalaamu'alaikum Warahmatullaahi Wabarakaatuh*

Malang,30 Desember 2023

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGANTAR .....	ii
HALAMAN PERSETUJUAN .....	iii
HALAMAN PENGESAHAN .....	iv
PERNYATAAN KEASLIAN TULISAN .....	iv
MOTO .....	vi
PERSEMBAHAN .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	x
DAFTAR TABEL .....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN .....	xiv
ABSTRAK .....	xv
ABSTRACT .....	xvi
مستخلص البحث .....	xvii
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	7
1.3 Tujuan Penelitian .....	7
1.4 Manfaat Penelitian .....	8
1.5 Batasan Masalah .....	8
1.6 Definisi Istilah .....	9
<b>BAB II KAJIAN TEORI .....</b>	<b>10</b>
2.1 Teori Pendukung .....	10
2.1.1 Data <i>Training</i> dan Data <i>Testing</i> .....	10
2.1.2 <i>Synthetic Minority Oversampling Technique</i> (SMOTE) .....	11
2.1.3 <i>Decision Tree</i> .....	14
2.1.4 Algoritma C5.0 .....	17
2.1.5 Tingkat Akurasi Ketepatan Klasifikasi .....	19
2.1.7 Kemiskinan dan Kesejahteraan Rumah Tangga .....	23
2.1.8 Faktor-Faktor Kesejahteraan Rumah Tangga .....	26
2.2 Kajian Islam .....	27
2.3 Kajian Kesejahteraan Rumah Tangga dengan Algoritma C5.0 dan Metode <i>Synthetic Minority Oversampling Technique</i> (SMOTE) .....	29
<b>BAB III METODE PENELITIAN .....</b>	<b>32</b>
3.1 Jenis Penelitian .....	32
3.2 Data dan Sumber Data .....	32
3.3 Variabel dan Jenis Data .....	32
3.4 Teknik Analisis Data .....	34
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>37</b>
4.1 Analisis Statistik Deskriptif .....	37
4.1.1 Deskripsi Jenis Kelamin Kepala Rumah Tangga .....	38
4.1.2 Deskripsi Umur Kepala Rumah Tangga .....	39
4.1.3 Deskripsi Banyak Anggota Rumah Tangga .....	41
4.1.4 Deskripsi Tingkat Pendidikan Kepala Rumah Tangga .....	42
4.1.5 Deskripsi Status Pekerjaan Kepala Rumah Tangga .....	43

4.1.6	Lapangan Pekerjaan Kepala Rumah Tangga .....	46
4.2	Pembagian Data Training dan Data Testing .....	47
4.3	<i>Synthetic Minority Oversampling Technique</i> (SMOTE) .....	47
4.4	Algoritma C5.0 .....	49
4.5.1	Pohon Klasifikasi .....	49
4.5.2	Tingkat Akurasi Ketepatan Klasifikasi .....	53
4.5	Tingkat Kepentingan Variabel .....	55
4.6	Kajian Islam pada Hasil Penelitian .....	56
<b>BAB V PENUTUP</b> .....		<b>58</b>
5.1	Kesimpulan .....	58
5.2	Saran .....	58
<b>DAFTAR PUSTAKA</b> .....		<b>60</b>
<b>LAMPIRAN</b> .....		<b>63</b>
<b>RIWAYAT HIDUP</b>		

## DAFTAR TABEL

<b>Tabel 2.1</b>	<i>Confusion Matrix</i> .....	21
<b>Tabel 2.2</b>	Nilai AUC .....	23
<b>Tabel 3.1</b>	Variabel Prediktor .....	33
<b>Tabel 4.1</b>	Deskripsi Pengeluaran Per Kapita Per Bulan .....	38
<b>Tabel 4.2</b>	Jenis Kelamin Berdasarkan Status Kesejahteraan.....	39
<b>Tabel 4.3</b>	Umur berdasarkan Status Kesejahteraan.....	40
<b>Tabel 4.4</b>	Banyaknya ART berdasarkan Status Kesejahteraan .....	41
<b>Tabel 4.5</b>	Pendidikan berdasarkan Status Kesejahteraan .....	43
<b>Tabel 4.6</b>	Status pekerjaan berdasarkan Status Kesejahteraan.....	45
<b>Tabel 4.7</b>	Lapangan Pekerjaan berdasarkan Status Kesejahteraan.....	46
<b>Tabel 4.8</b>	Data Training dan Data Testing .....	48
<b>Tabel 4.9</b>	<i>Confussion Matrix</i> .....	53
<b>Tabel 4.10</b>	Tingkat Kepentingan Variabel .....	56

## DAFTAR GAMBAR

<b>Gambar 2.1</b>	Struktur Pohon Klasifikasi.....	15
<b>Gambar 3.1</b>	Flowchart .....	36
<b>Gambar 4.1</b>	Status Kesejahteraan Rumah Tangga .....	37
<b>Gambar 4.2</b>	Jenis Kelamin .....	38
<b>Gambar 4.3</b>	Umur .....	40
<b>Gambar 4.4</b>	Banyaknya ART .....	41
<b>Gambar 4.5</b>	Tingkat Pendidikan .....	42
<b>Gambar 4.6</b>	Status Pekerjaan .....	44
<b>Gambar 4.7</b>	Lapangan Pekerjaan .....	46
<b>Gambar 4.8</b>	Penerapan SMOTE .....	47
<b>Gambar 4.9</b>	Pohon Klasifikasi.....	52
<b>Gambar 4.10</b>	Curve ROC .....	54

## DAFTAR LAMPIRAN

<b>Lampiran 1</b>	Data Penelitian SUSENAS kota Malang tahun 2020 .....	63
<b>Lampiran 2</b>	Script dan Output .....	67

## ABSTRAK

Hanin, Aida Nabila Yumna. 2023. **Penerapan Algoritma C5.0 dan Metode SMOTE pada Klasifikasi Status Kesejahteraan Rumah Tangga**. Skripsi. Jurusan Matematika. Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Fachrur Rozi, M.Si. (II) Erna Herawati M.Pd.

**Kata kunci:** Klasifikasi, Algoritma C5.0, SMOTE (*Synthetic Minority Oversampling Technique*), Ketepatan Klasifikasi, AUC-ROC, Kesejahteraan Rumah Tangga.

Data tidak seimbang merupakan permasalahan yang sering terjadi pada kasus klasifikasi. Metode yang digunakan untuk menangani data tidak seimbang adalah SMOTE (*Synthetic Minority Oversampling Technique*). Banyak algoritma yang dapat digunakan untuk membangun *decision tree* salah satunya adalah Algoritma C5.0. Dalam algoritma C5.0, pemilihan atribut menggunakan perhitungan ukuran *gain ratio* yang digunakan untuk memilih atribut uji pada tiap *node* di dalam pohon. Salah satu kegunaan Algoritma C5.0, dapat digunakan untuk mengklasifikasikan tingkat kesejahteraan. Kesejahteraan merupakan kondisi kebutuhan di dalam keluarga yang telah terpenuhi dengan baik. Penelitian ini bertujuan untuk mengidentifikasi tingkat akurasi dan tingkat kepentingan variabel pada klasifikasi status kesejahteraan rumah tangga di Kota Malang, yang diklasifikasikan menjadi kategori miskin dan tidak miskin. Variabel yang digunakan adalah jenis kelamin, umur, banyaknya anggota rumah tangga, tingkat pendidikan tertinggi, lapangan pekerjaan dan status pekerjaan kepala rumah tangga. Berdasarkan hasil penelitian dengan metode Algoritma C5.0 dan metode SMOTE menunjukkan bahwa nilai *accuracy* sebesar 76,32%, *sensitivity* sebesar 79,75%, *specificity* sebesar 72,60% dan nilai AUC sebesar 0,7617. Variabel terpenting yang mempengaruhi pohon klasifikasi adalah pendidikan dengan persentase sebesar 32,21%, kemudian variabel status pekerjaan dengan persentase 26,17%. Selanjutnya variabel banyak anggota rumah tangga dengan persentase 25,64%, yang ke-empat ada variabel jenis kelamin dengan persentase 6,71%. Kemudian variabel umur 4,97% dan terakhir variabel lapangan pekerjaan dengan persentase sebesar 4,26%.

## ABSTRACT

Hanin, Aida Nabila Yumna. 2023. **Application of the C5.0 Algorithm and the SMOTE Method in Classification of Household Welfare Status**. Thesis. Mathematics Department. Faculty of Science and Technology. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors : (I) Dr. Fachrur Rozi, M.Si. (II) Erna Herawati M.Pd.

**Keywords:** Classification, C5.0 Algorithm, SMOTE (Synthetic Minority Oversampling Technique), Accuracy of Classification, AUC-ROC, Household Welfare.

Unbalanced data is a problem that often occurs in classification cases. The method used to handle unbalanced data is SMOTE (Synthetic Minority Oversampling Technique). There are many algorithms that can be used to build decision trees, one of which is the C5.0 algorithm. In the C5.0 algorithm, attribute selection uses a gain ratio calculation which is used to select test attributes at each node in the tree. One of the uses of the C5.0 algorithm is that it can be used to classify welfare levels. Welfare is a condition where needs in the family have been fulfilled properly. This research aims to identify the level of accuracy and importance of variables in the classification of the welfare status of households in Malang City, which are classified into poor and non-poor categories. The variables used are gender, age, number of household members, highest level of education, employment and employment status of the head of the household. Based on the results of research using the C5.0 Algorithm method and the SMOTE method, it shows that the accuracy value is 76.32%, the sensitivity is 79.75%, the specificity is 72.60% and the AUC value is 0.7617. The most important variable that influences the classification tree is education with a percentage of 32.21%, then the employment status variable with a percentage of 26.17%. Next, there is the variable number of household members with a percentage of 25.64%, fourthly there is the gender variable with a percentage of 6.71%. Then the age variable is 4.97% and finally the employment variable with a percentage of 4.26%.



## مستخلص البحث

حنين، عايذة نبيلة يمني. ٢٠٢٣. تطبيق خوارزمية *C5.0* وطريقة *SMOTE* في تصنيف حالة الأسرة المزدهرة. البحث العلمي. قسم الرياضيات. كلية العلوم والتكنولوجيا. جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف (1) د. فخر الرازي ، الماجستير. (2) إيرنا هيراواتي، الماجستير.

الكلمات الرئيسية: التصنيف، خوارزمية *C5.0*، *SMOTE* (الاقلية الإصطناعية على اخذ العينات)، دقة التصنيف، *AUC-ROC*.

تعتبر البيانات غير المتوازنة مشكلة متكررة في حالة التصنيف. الطريقة المستخدمة للتعامل مع البيانات غير المتوازنة هي *SMOTE* (الاقلية الإصطناعية على اخذ العينات). هناك العديد من الخوارزميات التي يمكن استخدامها لبناء أشجار القرار، إحداها هي خوارزمية *C5.0* في خوارزمية *C5.0* ، يستخدم تحديد السمة حساب نسبة الكسب الذي يتم استخدامه لتحديد سمة الاختبار في كل عقدة في الشجرة. أحد استخدامات خوارزمية *C5.0* هو أنه يمكن استخدامها لتصنيف مستويات الرفاهية. الرفاهية هي حالة يتم فيها تلبية احتياجات الأسرة بشكل جيد. يهدف هذا البحث إلى تحديد مستوى دقة وأهمية المتغيرات في تصنيف حالة الرفاهية للأسر في مدينة مالانج والتي تم تجميعها إلى فئتين، الفقيرة وغير الفقيرة. والمتغيرات المستخدمة هي الجنس، والعمر، وعدد أفراد الأسرة، وأعلى مستوى تعليمي، والمهنة، والحالة الوظيفية لأرب الأسرة. بناءً على نتائج البحث باستخدام طريقة الخوارزمية *C5.0* وطريقة *SMOTE* ، تم الحصول على قيمة دقة تبلغ 76.32%، وقيمة حساسية 79.75%، ونوعية 72.60%، وقيمة *AUC* تبلغ 0.7617. أما المتغير الأكثر تأثيراً على شجرة التصنيف فهو التعليم بنسبة 32.21%، ثم متغير الحالة الوظيفية بنسبة 26.17%. يليه متغير عدد أفراد الأسرة بنسبة 25.64%، ورابعاً متغير الجنس بنسبة 6.71%. ثم متغير العمر بنسبة 4.97% وأخيراً متغير التوظيف بنسبة 4.26%.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Data tidak seimbang merupakan permasalahan yang sering terjadi pada kasus klasifikasi. Ketidakseimbangan data terjadi ketika jumlah objek dalam satu kelas data lebih besar daripada jumlah objek data lainnya. Sebagian besar algoritma untuk klasifikasi cenderung mengklasifikasikan kelas mayoritas dengan akurasi tinggi dan kelas minoritas dengan akurasi rendah dalam kasus ketidakseimbangan data, yang terjadi ketika salah satu kelas memiliki data yang sangat besar dan kelas lainnya memiliki data yang sangat kecil. Ketidakseimbangan data adalah faktor yang dapat mempengaruhi tingkat akurasi dalam klasifikasi (Qiong Gu, 2016). *Synthetic Minority Oversampling Technique (SMOTE)* merupakan salah satu metode untuk menangani ketidakseimbangan data. Adapun kelebihan dari metode SMOTE adalah tidak menyebabkan adanya informasi yang hilang, menghindari terjadinya *overfitting*, membangun wilayah keputusan yang lebih besar, serta mampu meningkatkan akurasi prediksi kelas minoritas (Wijayanti Trisna dkk, 2021).

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data yang bertujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya belum diketahui. Ada beberapa macam pengklasifikasian dalam *data mining* yaitu *decision tree*, *naïve bayes*, *Support Vector Machine (SVM)* dan lain-lain (Aradea, dkk, 2011). *Decision Tree* atau pohon keputusan merupakan salah satu teknik klasifikasi. *Decision Tree* merupakan suatu metode klasifikasi yang menggunakan struktur pohon, dimana

setiap *node* merepresentasikan atribut dan atribut merepresentasikan nilai dari atribut, sedangkan daunnya digunakan untuk merepresentasikan kelas. Pohon keputusan ini dapat memudahkan manusia dalam melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan menganalisa nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah (Kastawan dkk, 2018). Banyak algoritma yang dapat digunakan untuk membangun *decision tree* seperti ID3, C4.5, C5.0, *Classification And Regression Tree (CART)*, dan GUIDE (Pratiwi, 2020).

Algoritma C5.0 merupakan salah satu Algoritma pohon keputusan dan merupakan penyempurnaan dari Algoritma C4.5. algoritma C5.0 lebih baik dari C4.5 pada kecepatan, memori, dan efisiensi. Dalam algoritma C5.0, pemilihan atribut menggunakan perhitungan ukuran *gain ratio*. Ukuran *gain ratio* digunakan untuk memilih atribut uji pada tiap *node* di dalam pohon. Sehingga *node* pada pohon terbentuk menggunakan perhitungan ukuran *gain ratio*. Atribut dengan nilai *gain ratio* tertinggi dipilih sebagai *parent* untuk *node* . Adapun kelebihan pada algoritma ini adalah dapat mengatasi *missing value* dan dalam jumlah data yang besar serta dapat digunakan untuk data kategorikal dan juga data numerik atau kontinu (Kusrin dan Luthfi, 2009).

Salah satu kegunaan metode di atas, dapat digunakan untuk mengklasifikasikan tingkat kesejahteraan yang masih menjadi hal yang diperjuangkan bagi negara Indonesia. Menurut Badan Pusat Statistik (BPS), tingkat kesejahteraan dapat dinilai melalui pengeluaran rumah tangga yaitu dengan melihat komposisi pengeluaran makanan dan non makanan. Tingkat kesejahteraan akan semakin baik jika pangsa pengeluaran pangan semakin rendah atau rumah tangga

akan dikatakan sejahtera jika pengeluaran makanan lebih kecil dibanding pengeluaran non makanan (BPS, 2021). Menurut Badan Pusat Statistik (BPS), kemiskinan didefinisikan sebagai ketidakmampuan ekonomi seseorang untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Kemiskinan adalah ketika seseorang tidak dapat memenuhi kebutuhan dasar mereka, yang berdampak pada kesejahteraan rumah tangga. Garis kemiskinan, yang dihitung berdasarkan pengeluaran per kapita setiap bulan, sebuah rumah tangga dapat dikategorikan menjadi miskin atau tidak miskin. Garis kemiskinan adalah penjumlahan dari nilai pengeluaran minimum untuk makanan dan non-makanan. Namun, pengeluaran per kapita adalah jumlah uang yang dihabiskan setiap anggota keluarga selama satu bulan, baik itu dari membeli, memberikan, atau membuat sendiri, dibagi dengan jumlah anggota keluarga. Maka definisi penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan (BPS,2020). Di dalam pandangan Islam, kemiskinan merupakan sebuah cobaan bagi manusia untuk menguji kesabaran. Hal tersebut sesuai dengan Q.S Al-Baqarah ayat 155, sebagai berikut:

وَلَنَبْلُوَنَّكُمْ بِشَيْءٍ مِّنَ الْخَوْفِ وَالْجُوعِ وَنَقْصٍ مِّنَ الْأَمْوَالِ وَالْأَنْفُسِ وَالثَّمَرَاتِ وَبَشِّرِ الصَّابِرِينَ ﴿١٥٥﴾  
(سُورَةُ الْبَقَرَةِ)

Artinya: “Dan sungguh akan Kami berikan cobaan kepadamu, dengan sedikit ketakutan, kelaparan, kekurangan harta, jiwa dan buah-buahan. Dan berikanlah berita gembira kepada orang-orang yang sabar.”

Ayat ini menggambarkan bahwa dalam kehidupan, Allah SWT akan menguji manusia dengan berbagai cobaan, termasuk ketakutan, kelaparan, kekurangan harta, jiwa, dan buah-buahan. Salah satu fenomena yang sering terkait dengan kemiskinan adalah ketakutan dan kelaparan, yang merupakan dua elemen

yang disebutkan dalam ayat ini. Kemiskinan bisa menyebabkan ketakutan dan kelaparan karena kurangnya kesejahteraan untuk memenuhi kebutuhan dasar.

Berdasarkan Q.S Al-Baqarah ayat 155, terdapat hadist yang di dalamnya adalah doa agar dijauhkan dari kemiskinan dan kekurangan yang diriwayatkan oleh Anas bin Malik, Rasulullah bersabda:

حَدَّثَنَا أَبُو بَكْرِ حَدَّثَنَا مُحَمَّدُ بْنُ مُصْعَبٍ عَنِ الْأَوْزَاعِيِّ عَنِ إِسْحَاقَ بْنِ عَبْدِ اللَّهِ عَنِ جَعْفَرِ بْنِ عِيَّاضٍ عَنْ أَبِي هُرَيْرَةَ قَالَ قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ تَعَوَّذُوا بِاللَّهِ مِنَ الْفَقْرِ وَالْقِلَّةِ وَالذَّلَّةِ وَأَنْ تَظْلِمَ أَوْ تُظْلَمَ (رواه البخارى)

*Artinya: Telah menceritakan kepada kami [Abu Bakar] telah menceritakan kepada kami [Muhammad bin Mush'ab] dari [Al Auza'i] dari [Ishaq bin Abdullah] dari [Ja'far bin 'Iyadl] dari [Abu Hurairah] dia berkata; Nabi Muhammad Saw berdoa: "Ya Allah, Sesungguhnya aku memohon perlindungan pada-Mu dari kefakiran, kekurangan, dan aku perlindungan kepada-Mu dari kekurangan dan kehinaan, dan aku memohon perlindungan kepada-Mu dari berbuat zhalim atau dizhalimi."*

Hadis riwayat bukhori di dalam kitab *Adabul Mufrod* diatas menjelaskan tentang doa agar terhindar dari kemiskinan, kekurangan dan memohon perlindungan dari perbuatan menganiaya dan dianiaya. Karena itulah pentingnya kesejahteraan bagi tiap manusia dan rumah tangga.

Permasalahan kesejahteraan serta penanganan kemiskinan masih menjadi target utama bagi pemerintah, tak terkecuali bagi pemerintah Kota Malang. Hal yang perlu dilakukan untuk mencapai kesejahteraan adalah dengan mengevaluasi daerah dan mengikuti langkah daerah lain yang telah sejahtera. Kota Malang merupakan salah satu kota dengan persentase penduduk miskin terendah kedua di provinsi Jawa Timur dan merupakan kota terbesar kedua menurut kependudukan di provinsi Jawa Timur setelah Kota Surabaya (BPS, 2020). Hal ini menjadi alasan bahwa Kota Malang dapat menjadi salah satu daerah panutan untuk daerah lain di Provinsi Jawa Timur. Pada tahun 2018, persentase kemiskinan sebesar 4,10%

dengan jumlah penduduk miskin 35490 jiwa dan pada tahun 2019 persentase kemiskinan menurun sebesar 0,03% menjadi 4,07% dengan jumlah penduduk miskin 35390 jiwa. Kemudian, tahun 2020 persentase kemiskinan mengalami kenaikan sebesar 0,37% sehingga mencapai 4,4% dengan jumlah penduduk miskin 38770 jiwa (BPS,2020). Kenaikan angka kemiskinan ini merupakan yang pertama di Kota Malang dalam satu dekade. Berbeda dengan persentase kemiskinan, garis kemiskinan Kota Malang justru menunjukkan trend positif. Pada tahun 2019, garis kemiskinan meningkat dari 507.114 rupiah per kapita per bulan menjadi 543.966 rupiah per kapita per bulan. Pada tahun 2020 garis kemiskinan juga meningkat menjadi 554.791 rupiah per kapita per bulan (BPS, 2020).

Terdapat beberapa indikator yang mampu mencerminkan situasi kesejahteraan rumah tangga, sehingga banyak yang memiliki potensi untuk memengaruhi kesejahteraan rumah tangga di Kota Malang. Metode pengklasifikasian dapat digunakan untuk menentukan faktor-faktor yang berpengaruh. Kesejahteraan rumah tangga tersebut diklasifikasikan dalam 2 kategori, yaitu penduduk miskin dan tidak miskin. Berbagai penelitian telah dilakukan mengenai klasifikasi kesejahteraan khususnya kesejahteraan rumah tangga. Permatasari & Yuliana (2021) menggunakan metode regresi logistik biner untuk menganalisis status kesejahteraan rumah tangga di provinsi bali. Variabel yang memengaruhi secara signifikan adalah jumlah anggota rumah tangga, jenis kelamin kepala rumah tangga, umur kepala rumah tangga, pendidikan terakhir kepala rumah tangga, lapangan usaha kepala rumah tangga dan status pekerjaan kepala rumah tangga. Penelitian ini mendapatkan hasil bahwa analisis regresi logistik biner rumah tangga di provinsi Bali akan cenderung sejahtera apabila

memiliki jumlah anggota rumah tangga yang kecil (paling banyak empat orang), kepala rumah tangga berjenis kelamin perempuan, kepala rumah tangga diatas 47 tahun, kepala rumah tangga berpendidikan tinggi (minimal SMA), kepala rumah tangga yang bekerja di sektor non pertanian dan sebagai pemilik usaha/wirausaha.

Penelitian terdahulu mengenai Algoritma C5.0 dan SMOTE telah dilakukan antara lain oleh Yuni dkk (2017) melakukan penelitian untuk menentukan status berlangganan klien Bank pada kampanye pemasaran langsung. Metode klasifikasi yang digunakan adalah Algoritma C5.0. Data yang dikumpulkan memiliki 4.521 record klien, dengan 521 record kelas yes dan 4000 record kelas no, sehingga dataset tidak seimbang. Masalah data tidak seimbang ditangani dengan metode SMOTE, yang menggunakan persentase duplikasi kelas minoritas sebesar 700% atau perbandingan jumlah record kelas yes:no sebesar 1:1. Nilai akurasi yang didapat dari Algoritma C5.0 dengan penerapan SMOTE adalah 91,3%. Kemudian Sofyan & Prasetyo (2021) menerapkan metode *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani ketidakseimbangan data dalam kasus pendapatan pekerja informal di Provinsi D.I. Yogyakarta pada tahun 2019. Dalam penelitian ini, metode klasifikasi yang diterapkan adalah regresi logistik biner. Model sebelum menerapkan metode SMOTE menghasilkan nilai akurasi sebesar 85,74%, sensitifitas sebesar 98,74%, dan spesifisitas sebesar 11,91%. Setelah diterapkan metode SMOTE, ketepatan klasifikasi yang dihasilkan adalah nilai akurasi sebesar 73,7%, nilai sensitifitas sebesar 73,69%, dan nilai spesifisitas sebesar 74,31%. Sehingga dapat disimpulkan bahwa penerapan SMOTE mampu menangani ketidakseimbangan data dan meningkatkan nilai spesifisitas.

Berdasarkan uraian latar belakang tersebut, maka penulis akan mengkaji metode Algoritma C5.0 untuk klasifikasi tingkat kesejahteraan rumah tangga dan metode SMOTE untuk penanganan data tidak seimbang. Pada penelitian ini penulis menyajikan dengan judul “Penerapan Algoritma C5.0 dan Metode SMOTE pada Klasifikasi Status Kesejahteraan Rumah Tangga”.

## **1.2 Rumusan Masalah**

Berdasarkan uraian pada latar belakang, maka dapat diperoleh susunan rumus masalah sebagai berikut:

1. Bagaimana hasil penerapan dan tingkat akurasi Algoritma C5.0 dengan metode *Synthetic Minority Oversampling Technique* (SMOTE) pada klasifikasi status kesejahteraan rumah tangga di Kota Malang?
2. Bagaimana tingkat kepentingan variabel yang mempengaruhi status kesejahteraan rumah tangga di Kota Malang menggunakan Algoritma C5.0 dengan metode *Synthetic Minority Oversampling Technique* (SMOTE)?

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah dijelaskan, maka tujuan penelitian ini adalah:

1. Mendapatkan hasil penerapan dan tingkat akurasi Algoritma C5.0 dengan metode *Synthetic Minority Oversampling Technique* (SMOTE) pada klasifikasi status kesejahteraan rumah tangga di Kota Malang.
2. Menganalisis tingkat kepentingan variabel yang mempengaruhi status kesejahteraan rumah tangga di Kota Malang menggunakan Algoritma C5.0 dengan metode *Synthetic Minority Oversampling Technique* (SMOTE).



#### 1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Memberikan gambaran dan pengetahuan mengenai penerapan dan tingkat akurasi metode *Synthetic Minority Oversampling Technique* (SMOTE) pada Algoritma C5.0 dalam menentukan tingkat akurasi model klasifikasi status kesejahteraan rumah tangga.
2. Memberikan gambaran dan pengetahuan mengenai tingkat kepentingan variabel yang mempengaruhi status kesejahteraan rumah tangga di Kota Malang menggunakan Algoritma C5.0 dengan metode *Synthetic Minority Oversampling Technique* (SMOTE).
3. Dapat dijadikan referensi bagi peneliti yang sedang meneliti di bidang ini.

#### 1.5 Batasan Masalah

Pada penelitian ini agar permasalahan yang akan diteliti menjadi lebih fokus dan lebih jelas maka permasalahan dibatasi oleh:

1. Data yang digunakan merupakan data mikro Survei Sosial dan Ekonomi Nasional (SUSENAS) Kota Malang pada tahun 2020 oleh Badan Pusat Statistik.
2. Tingkat kesejahteraan pada penelitian ini dinilai melalui pengeluaran rumah tangga yaitu dengan melihat komposisi pengeluaran makanan dan non makanan. Rumah tangga dikatakan sejahtera jika pengeluaran perkapita diatas garis kemiskinan.
3. Menggunakan *confusion matrix* dengan perhitungan *accuracy*, *sensitivity*, *specificity* dan AUC-ROC.

## 1.6 Definisi Istilah

Untuk menghindari kesalahpahaman dalam penafsiran istilah maka dalam penelitian ini perlu adanya batasan-batasan pengertian sebagai berikut:

1. *Accuracy* adalah suatu ukuran yang digunakan untuk mengetahui seberapa tepat model klasifikasi memprediksi kelas data.
2. *Atribut* adalah bagian data yang merepresentasikan karakteristik dari objek data.
3. *Entropy* adalah nilai informasi yang menyatakan ukuran ketidakpastian dari atribut dari kumpulan objek data dalam satuan bit.
4. *Gain* adalah ukuran efisiensi suatu variabel dalam mengklasifikasi data.
5. *Gain ratio* adalah ukuran yang digunakan untuk mengatasi masalah pada atribut yang memiliki nilai sangat bervariasi.
6. *Sensitivity* adalah ukuran seberapa baik tes mampu mengidentifikasi hasil positif yang sebenarnya.
7. *Specificity* adalah ukuran seberapa baik tes mampu mengidentifikasi hasil negatif yang sebenarnya.

## **BAB II**

### **KAJIAN TEORI**

#### **2.1 Teori Pendukung**

##### **2.1.1 Data *Training* dan Data *Testing***

Dalam konteks pengujian klasifikasi, pentingnya pemisahan data menjadi dua kelompok, yaitu data *training* dan data *testing*, tidak dapat diabaikan. Data *training* dan data *testing* memainkan peran kunci dalam membangun dan menguji model klasifikasi. Data *training*, yang merupakan subset pertama, memiliki peran sentral dalam mengenali karakteristik data dan membantu dalam pembentukan model. Dengan demikian, data *training* adalah dasar dalam pembangunan model klasifikasi yang akurat. Namun, sejalan dengan itu, peran data *testing* juga tak kalah penting. Setelah model klasifikasi dibentuk menggunakan data *training*, data *testing* digunakan untuk menguji sejauh mana model tersebut dapat memprediksi dengan benar pada data yang tidak pernah dilihat sebelumnya (Musu, dkk., 2021).

Komposisi atau presentase pembagian data antara *training* dan *testing* sangat memengaruhi evaluasi akurasi model. Jika terlalu sedikit data yang digunakan untuk *training*, model dapat kehilangan informasi penting dalam pembentukan pola, yang dapat mengakibatkan performa buruk saat diuji pada data *testing*. Di sisi lain, jika data *training* terlalu besar, model mungkin mempelajari data *training* dengan sangat baik, tetapi mungkin sulit untuk diterapkan pada data *testing* yang berbeda. Mengenai ketidaktepatan dalam menentukan komposisi jumlah data *training* dan *testing*, ini dapat menjadi penyebab utama penurunan kualitas model. Jika komposisi tidak seimbang,

seperti memiliki data *training* yang sangat banyak dibandingkan data *testing* atau sebaliknya, model dapat menjadi bias terhadap salah satu subset data tersebut. Oleh karena itu, penting untuk merencanakan pembagian data yang seimbang dan mewakili distribusi yang sesungguhnya (Musu, dkk., 2021).

Dalam upaya membagi data menjadi data training dan data testing untuk mengembangkan model klasifikasi yang andal, berbagai kombinasi presentase pembagian umumnya digunakan. Beberapa kombinasi yang lazim adalah 90%:10%, 85%:15%, 80%:20%, dan 75%:25%, di mana angka pertama merujuk pada persentase data training dan angka kedua merujuk pada persentase data testing. Kombinasi terbaik dari kombinasi yang ditentukan diperoleh dari hasil ketepatan klasifikasi dari persentase terbaik pada data *training* juga dat *testing*. Perhitungan ketepatan data *training* dan *testing* menjadi proses terbentuknya pohon klasifikasi yang baik (Nair, dkk., 2001).

### **2.1.2 Synthetic Minority Oversampling Technique (SMOTE)**

Ketika jumlah objek dalam satu kelas data lebih besar daripada jumlah objek dalam kelas data lainnya, itu disebut ketidakseimbangan data. Kelas data dengan jumlah objek yang lebih besar disebut kelas mayor, dan kelas data dengan jumlah objek yang lebih sedikit disebut kelas minor. Penggunaan data yang tidak seimbang dalam pembuatan model memiliki dampak yang besar pada hasil akhir model. Ketika algoritma digunakan tanpa memperhatikan ketidakseimbangan data, mereka cenderung menggunakan algoritma kelas mayor dan mengabaikan algoritma kelas minor (Azmatul dkk, 2013).

Menurut Chawla, *Synthetic Minority Oversampling Technique* (SMOTE) merupakan salah satu metode oversampling yang berfokus pada peningkatan

jumlah data pada kelas minoritas dalam sebuah dataset. *Oversampling* adalah penanganan data tidak seimbang dengan cara menggandakan sampel-sampel dari kelas minoritas dengan menciptakan contoh sintesis berdasarkan data yang sudah ada. Tujuan utama dari SMOTE adalah untuk mengatasi masalah ketidakseimbangan kelas, dimana kelas minoritas di-oversampling-kan dengan membuat data sintesis. Data sintesis tersebut dibuat berdasarkan k- amatan tetangga. Di sini, k- amatan tetangga merujuk pada sejumlah tetangga terdekat dari suatu contoh data minoritas tertentu. Proses ini dilakukan dengan mencari k- amatan tetangga terdekat untuk setiap contoh data minoritas. Setelah tetangga terdekat ditemukan, data sintesis dibentuk dengan cara mengambil beberapa atribut dari data minoritas asli dan beberapa atribut dari salah satu tetangga terdekat secara acak. Persentase dari atribut yang diambil ini ditentukan oleh faktor oversampling yang diinginkan. Dengan kata lain, berapa kali lipat data sintesis akan dibentuk dibandingkan dengan jumlah data minoritas awal (Azmatul dkk, 2013).

Metode pembentukan data sintesis berbeda antara data berskala numerik dan kategorik. Dalam kasus data numerik, jarak antara contoh-contoh data diukur menggunakan metrik Euclidean. Ini membantu dalam menilai seberapa dekat atau berbedanya suatu contoh data dengan tetangga terdekatnya. Namun, pada data kategorik, jarak diukur dengan menggunakan rumus *Value Difference Metric* (VDM), yang mempertimbangkan perbedaan nilai antar atribut kategorik. Metode SMOTE memiliki keunggulan dalam menghasilkan variasi data yang lebih besar daripada metode oversampling sederhana, seperti duplikasi. Ini membantu mencegah overfitting dan menghasilkan model klasifikasi yang lebih baik dalam

menggeneralisasikan data baru. Namun, seperti halnya dengan metode oversampling lainnya, perlu dilakukan dengan hati-hati dan mungkin memerlukan validasi tambahan untuk memastikan bahwa data sintetis yang dihasilkan benar-benar mewakili variasi dalam data minoritas yang sebenarnya (Azmatul dkk, 2013).

Rumus *Value Difference Metric* (VDM) digunakan untuk menghitung jarak antar contoh kelas data minor dengan peubah berskala kategorik, yaitu (Azmatul dkk, 2013):

$$\Delta(X, Y) = \sum_{i=1}^N \delta(V_1, V_2) \quad (2.4)$$

dengan:

$\Delta(X, Y)$  : Jarak antara amatan X dengan Y

$N$  : Banyaknya variabel prediktor

$\delta(V_1, V_2)$  : Jarak antara amatan X dan Y untuk setiap variabel yang dihitung

Digunakan rumus berikut untuk mengukur jarak antara observasi X dan Y pada setiap variabel.

$$\delta(V_1, V_2) = \sum_{i=1}^n \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right| \quad (2.5)$$

dengan:

$C_{1i}$  : Banyaknya kategori ke-1 yang termasuk pada variabel ke- $i$

$C_{2i}$  : Banyaknya kategori ke-2 yang termasuk pada variabel ke- $i$

$C_1$  : Banyaknya kategori ke-1 terjadi

$C_2$  : Banyaknya kategori ke-2 terjadi

$n$  : Banyaknya kategori pada variabel ke- $i$

Prosedur pembangkitan data buatan untuk (Azmatul dkk, 2013):

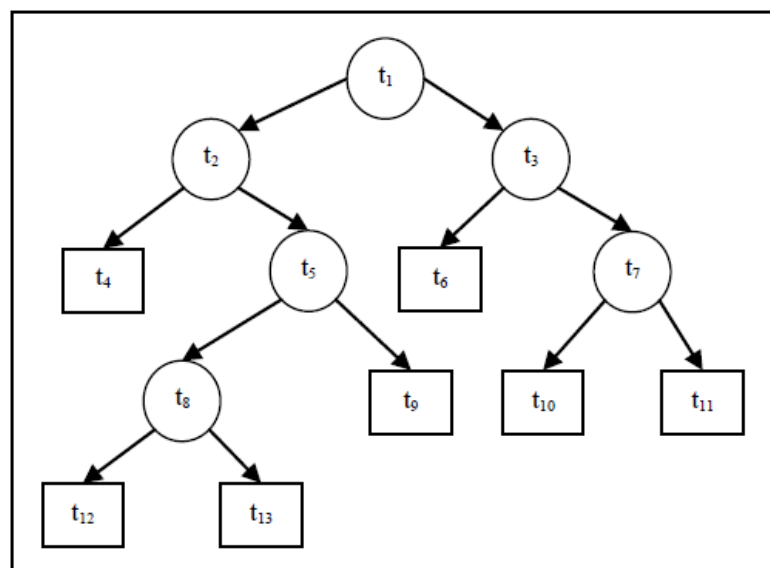
1. Data Numerik
  - a. Menghitung perbedaan antar vektor utama dengan  $k$ -tetangga terdekatnya.
  - b. Mengkalikan perbedaan dengan angka yang diacak diantara 0 dan 1.
  - c. Menambahkan perbedaan tersebut ke dalam nilai utama pada vektor utama asal sehingga diperoleh vektor utama baru.
2. Data Kategorik
  - a. Memilih mayoritas antar vector utama dengan  $k$ -tetangga terdekatnya untuk nilai nominal. Jika terjadi nilai yang sama maka pilih secara acak.
  - b. Menjadikan nilai tersebut sebagai kelas buatan baru.

### 2.1.3 *Decision Tree*

Pohon keputusan (*decision tree*) adalah salah satu metode yang umum digunakan dalam analisis data dan pembuatan keputusan. Metode ini menggambarkan aliran keputusan berdasarkan aturan-aturan yang dideduksi dari data yang ada. Pohon keputusan memiliki struktur mirip pohon, dengan akar sebagai *node* awal yang mewakili fitur atau atribut, cabang-cabang yang mewakili pilihan atau nilai dari fitur tersebut, dan daun-daun yang mewakili hasil keputusan atau kelas. Konsep dasar pohon keputusan adalah membagi populasi data menjadi subset-subset yang semakin kecil dan homogen berdasarkan atribut-atribut yang ada. Proses ini dilakukan dengan memilih atribut terbaik pada setiap langkah berdasarkan kriteria tertentu, seperti keuntungan informasi (*information gain*) atau rasio keuntungan (*gain ratio*). Atribut yang dipilih akan menjadi cabang-cabang pada pohon, dan proses ini berulang hingga mencapai kondisi di mana data dalam setiap cabang cukup homogen atau memenuhi syarat akhir lainnya. Bentuk ilustrasi pohon klasifikasi dapat dipelajari pada Gambar 2.1

Karakteristik dari *decision tree* dibentuk dari sejumlah elemen sebagai berikut (Prasetyo, 2012):



1. *Node*, yang menggambarkan variabel. *Node* bisa berupa variabel akar (*root node*), variabel cabang, dan kelas.
2. *Arm*, setiap cabang menyatakan nilai hasil pengujian di *node* bukan daun.
3. *Node* akar, tidak mempunyai *input arm* yaitu lengan masukan dan mempunyai nol atau lebih *output arm* yaitu lengan keluar.
4. *Node* internal, setiap *node* yang bukan daun (non terminal) yang memiliki tepat satu *input arm* dan dua atau lebih *output arm*, *node* ini menyatakan pengujian yang didasarkan pada nilai fitur.
5. *Node* daun (terminal) adalah *node* yang mempunyai tepat satu input arm dan tidak mempunyai output arm. *Node* ini menyatakan label kelas (keputusan).



Gambar 2.1 Struktur Pohon Klasifikasi

Simpul utama (*root node*) dinotasikan sebagai  $t_1$ , sedangkan simpul  $t_2$ ,  $t_3$ ,  $t_5$ ,  $t_7$  dan  $t_8$  disebut simpul dalam (*internal nodes*). Simpul akhir yang juga disebut sebagai simpul terminal (*terminal nodes*) adalah  $t_4$ ,  $t_6$ ,  $t_9$ ,  $t_{10}$ ,  $t_{11}$ ,  $t_{12}$  dan



$t_{13}$  dimana tidak terjadi lagi pemilahan. Kedalaman pohon (depth) dihitung dimulai dari simpul utama atau  $t_1$  yang berada pada kedalaman 1, sedangkan  $t_2$  dan  $t_3$  berada pada kedalaman 2. Begitu seterusnya sampai pada simpul terminal  $t_{12}$  dan  $t_{13}$  yang berada pada kedalaman 5 yang ditunjukkan pada Gambar 2.1 dengan *node* yang digambarkan  merupakan node utama dan node dalam sedangkan *node* yang digambarkan  merupakan *node terminal* (Pratiwi dan Zain, 2014).

Salah satu keuntungan utama dari pohon keputusan adalah kemampuannya untuk memberikan pemahaman yang visual tentang bagaimana suatu keputusan dibuat. Ini memungkinkan pengguna untuk menganalisis alur logika di balik suatu keputusan dan memahami faktor-faktor yang paling berpengaruh dalam mengklasifikasikan atau memprediksi suatu objek. Selain itu, pohon keputusan juga dapat mengatasi masalah data yang tidak lengkap atau memiliki nilai yang hilang (Nair, dkk., 2001).

Namun, pohon keputusan juga memiliki potensi untuk overfitting, di mana pohon dapat menjadi terlalu kompleks dan “menghafal” data training yang spesifik, sehingga kinerjanya menurun pada data baru. Untuk mengatasi hal ini, teknik pruning atau pengurangan cabang-cabang yang tidak signifikan sering digunakan. Selain itu, dalam kasus data yang tidak seimbang, di mana beberapa kelas memiliki frekuensi yang lebih rendah, pohon keputusan dapat cenderung memihak pada kelas mayoritas. Secara keseluruhan, pohon keputusan adalah alat yang kuat dan fleksibel dalam analisis data dan pembuatan keputusan. Dengan kecanggihan analisis yang tepat, pohon keputusan dapat menghasilkan model

yang mudah diinterpretasi dan dapat memberikan wawasan berharga bagi pengambilan keputusan (Nair, dkk., 2001).

#### **2.1.4 Algoritma C5.0**

Algoritma C5.0, sebuah pengembangan dari Algoritma 1D3 dan C4.5, muncul dari karya Ross Quinlan pada tahun 1987. Sejak kemunculannya, algoritma ini telah menjadi salah satu pendekatan yang sangat berguna dalam dunia analisis data dan pembuatan model klasifikasi. Algoritma C5.0 memperkenalkan pendekatan yang lebih canggih dan efektif dalam memilih atribut yang akan digunakan dalam pembentukan model. Salah satu perubahan utama dalam algoritma C5.0 adalah penggunaan metrik yang berbeda dalam pemilihan atribut dibandingkan dengan pendahulunya. Atribut yang akan dijadikan node pemisah dalam pohon klasifikasi dipilih berdasarkan metode yang disebut gain ratio. Gain ratio adalah rasio antara gain informasi (information gain) dan split information. Pendekatan ini mengatasi kelemahan yang dapat terjadi pada penggunaan hanya informasi gain dalam pemilihan atribut. Proses pembentukan pohon pada algoritma C5.0 memiliki beberapa kesamaan dengan algoritma pendahulunya, seperti 1D3 dan C4.5. Langkah-langkah dalam membangun pohon keputusan melibatkan pembagian data menjadi subset-subset yang lebih kecil dan homogen berdasarkan atribut-atribut yang ada. Hal ini dilakukan dengan melakukan rekursi, di mana pada setiap langkah, atribut terbaik dipilih untuk menjadi node pemisah yang menghasilkan cabang-cabang berdasarkan nilai-nilai yang berbeda dari atribut tersebut. (Yusuf, dkk, 2022).

Meskipun proses ini memiliki kesamaan dengan algoritma 1D3 dan C4.5, algoritma C5.0 memiliki fitur unik dalam hal pembentukan pohon. Pohon yang dihasilkan oleh algoritma C5.0 memiliki jumlah cabang yang fleksibel di setiap

node. Ini berarti bahwa pohon tidak memiliki batasan khusus pada jumlah cabang yang mungkin dihasilkan pada setiap langkah. Ini memungkinkan pohon untuk menggambarkan kompleksitas yang lebih tinggi dalam data yang lebih rumit. Salah satu keunggulan lain dari algoritma C5.0 adalah kemampuannya untuk melanjutkan penghitungan gain ratio setelah perhitungan gain selesai dilakukan. Pembentukan pohon dimulai dengan perhitungan *entropy*, dilanjutkan dengan *gain* dan *gain ratio*. Jika algoritma 1D3 dan C4.5 berhenti pada perhitungan *gain*, algoritma C5.0 akan melangkah lebih jauh dengan menghitung *gain ratio*. Ini memungkinkan algoritma C5.0 untuk memilih atribut yang lebih optimal dalam pembentukan pohon. Adapun rumus untuk mencari nilai *entropy* sebagai berikut (Putri dkk, 2013):

$$Entropy(S) = - \sum_{i=1}^m p_i \log_2 p_i \quad (2.1)$$

dengan:

$S$  : Gabungan data/atribut

$m$  : Jumlah pembagian  $S$

$p_i$  : Jumlah perbandingan keseluruhan atribut ( $S_i$ ) pada keseluruhan total ( $S$ )

Selanjutnya untuk mencari nilai *gain* digunakan rumus berikut (Putri dkk, 2013):

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^m \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2.2)$$

dengan:

$Entropy(S)$  : Entropy keseluruhan

$Entropy(S_i)$  : Entropy atribut

$A$  : Atribut

$m$  : Jumlah pembagian  $S$

$|S_i|$  : Jumlah kasus atribut

$|S|$  : Jumlah kasus pada  $S$

Setelah mendapatkan nilai *entropy* dan nilai *gain*, langkah selanjutnya adalah melakukan perhitungan untuk mendapatkan nilai *gain ratio*. Rumus untuk menghitung *gain ratio* ini adalah sebagai berikut.:

$$Gain\ Ratio = \frac{Gain(S, A)}{\sum_{i=1}^m Entropy(S_i)} \quad (2.3)$$

dengan:

$Gain(S, A)$  : Nilai *gain* dari suatu atribut

$\sum_{i=1}^m Entropy(S_i)$  : Jumlah nilai *entropy* dalam suatu atribut

Proses diulang untuk masing-masing cabang sampai semua kelas pada cabang memiliki kelasnya masing-masing (Yusuf, 2022).

### 2.1.5 Tingkat Akurasi Ketepatan Klasifikasi

Evaluasi model merupakan langkah penting dalam proses pengembangan model klasifikasi untuk mendapatkan wawasan yang lebih mendalam tentang seberapa baik model tersebut dapat melakukan prediksi pada data baru. Salah satu alat yang umum digunakan untuk evaluasi model adalah confusion matrix, yang membantu dalam mengukur kinerja model dengan cara membandingkan kategori kelas yang sebenarnya dengan hasil prediksi yang dihasilkan oleh model. Confusion matrix adalah sebuah matriks tabulasi silang yang menggambarkan empat kemungkinan hasil dari prediksi: *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). *True positive* merujuk pada kasus di mana model benar-benar memprediksi kelas positif dengan benar, sementara *true negative* merujuk pada kasus di mana model memprediksi kelas negatif dengan

benar. *False positive* terjadi ketika model salah memprediksi kelas positif, sedangkan *false negative* terjadi ketika model salah memprediksi kelas negatif.

Dari confusion matrix, beberapa metrik evaluasi model dapat dihitung. Salah satunya adalah akurasi (*accuracy*), yang merupakan rasio antara jumlah prediksi benar (TP dan TN) dengan total jumlah data. Akurasi memberikan gambaran tentang sejauh mana model mampu melakukan prediksi yang tepat secara keseluruhan. Selain akurasi, dua metrik penting lainnya adalah *sensitivity* dan *specificity*. *Sensitivity*, juga dikenal sebagai *true positive rate* atau *recall*, mengukur seberapa baik model dapat mengidentifikasi kasus positif yang sebenarnya. Itu dihitung dengan membagi TP dengan jumlah keseluruhan data positif. Sementara itu, *specificity* mengukur seberapa baik model dapat mengidentifikasi kasus negatif yang sebenarnya. *Specificity* dihitung dengan membagi TN dengan jumlah keseluruhan data negatif. Ketiga metrik ini bersama-sama memberikan gambaran yang lebih lengkap tentang kinerja model. Meskipun akurasi dapat memberikan pandangan umum tentang performa model, *sensitivity* dan *specificity* memberikan wawasan tentang bagaimana model menangani kasus positif dan negatif secara khusus. Kombinasi ketiganya membantu dalam memahami apakah model memiliki kecenderungan untuk salah klasifikasi pada kelas tertentu atau apakah memiliki keseimbangan yang baik antara kemampuan mengidentifikasi kelas positif dan negatif. Berikut ini adalah format dari *confusion matrix* (Bisri, 2015):

**Tabel 2.1** *Confusion Matrix*

Predicted Class	Actual Class	
	C1	-C1
C1	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
-C1	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keterangan:

*Actual class* : Kelas yang sebenarnya

*Predicted class* : Kelas hasil prediksi dari model klasifikasi

*True Positive (TP)* : Jumlah baris kelas C1 yang benar diklasifikasikan sebagai kelas C1

*False Negative (FN)* : Jumlah baris kelas C1 tetapi diklasifikasikan sebagai bukan kelas C1

*False Positive (FP)* : Jumlah baris bukan kelas C1 tetapi diklasifikasikan sebagai kelas C1

*True Negative (TN)* : Jumlah baris bukan kelas C1 yang benar diklasifikasikan sebagai bukan kelas C1

Perhitungan *accuracy*, *specificity*, dan *sensitivity* secara berturut-turut menggunakan rumus sebagai berikut (Bisri, 2015):

$$\begin{aligned}
 accuracy &= \frac{\text{banyaknya prediksi yang benar}}{\text{total banyaknya prediksi}} \\
 &= \frac{TP + TN}{TP + FN + FP + TN}
 \end{aligned}
 \tag{2.6}$$

$$sensitivity = \frac{\text{banyaknya true positive}}{\text{banyaknya actual positive}} = \frac{TP}{TP + FN}
 \tag{2.7}$$

$$specificity = \frac{\text{banyaknya true negatif}}{\text{banyaknya actual negatif}} = \frac{TN}{TN + FP}
 \tag{2.8}$$

Evaluasi dengan *Receiver Operating Character Curve* (ROC Curve), secara teknis menggambarkan graifk dua dimensi, dimana tingkat *True Positive* (TP) terletak pada garis sumbu Y, sedangkan untuk *False Positive* (FP) terletak pada garis sumbu X. dengan demikian ROC menggambarkan *tradeoff* antara TP dan FP. Pencatatan dalam ROC dinyatakan dalam sebuah klausa yaitu semakin rendah titik kekiri (0.0), maka dinyatakan sebagai klasifikasi prediksi mendekati/menjadi negatif, sedangkan semakin keatas titik kekanan (1.1), maka dinyatakan sebagai klasifikasi prediksi mendekati/menjadi positif. Titik dengan nilai 1 dinyatakan sebagai tingkat *True Positif* (TP), sedangkan titik dengan nilai 0 dinyatakan sebagai tingkat *False Positive* (FP). Pada titik (0.1) merupakan klasifikasi prediksi adalah sempurna karena semua kasus baik positif maupun negatif dinyatakan dengan benar (*True*). Sedangkan untuk (1.0) klasifikasi prediksi semuanya dinyatakan sebagai tidak benar (*False*) (Bisri, 2015).

Sebuah metode umum untuk menghitung daerah dibawah kurva ROC adalah *Area Under Curve* (AUC) dimana bidang yang berada dibawah kurva mempunyai nilai yang selalu berada pada nilai 0 sampai dengan 1. Namun yang menarik untuk dihitung adalah yang mempunyai luas diatas 0,5, semakin tinggi luasnya maka akan semakin baik (Hardianto dkk, 2019). Adapun perhitungan AUC didefinisikan sebagai berikut (Sari, dkk, 2020):

$$AUC = \frac{1}{2} \left( \frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (2.9)$$

Dalam pengklasifikasian keakuratan dari tes diagnostik menggunakan *Area Under Curve* (AUC) (Gorunescu, 2011, hal. 325-326) sebuah sistem nilai yang disajikan pada Tabel 2.2.

**Tabel 2.2** Nilai AUC

<b>AUC</b>	<b>Keterangan</b>
0.90 – 1.00	<i>Excellent Classification</i>
0.80 – 0.90	<i>Good Classification</i>
0.70 – 0.80	<i>Fair Classification</i>
0.60 – 0.70	<i>Poor Classification</i>
0.50 – 0.70	<i>Failure</i>

### 2.1.6 Variabel Penting

Variabel terpenting adalah variabel yang paling berpengaruh. Pada klasifikasi Algoritma C5.0, setelah terbentuk pohon klasifikasi akan terlihat variabel mana yang memiliki pengaruh terbesar dalam proses klasifikasi. Variabel penting dapat diketahui dengan urutan variabel pada ketentuan dibawah simpul induk. Selain itu, variabel penting bisa diketahui dari data *overall* tiap variabel yang sudah dihasilkan pohon klasifikasi yang dibagi data keseluruhan (Bobbit, 2020). Perhitungan *variable importance* dapat dihitung dengan formula:

$$VarImp = \frac{\text{nilai overall tiap variabel}}{\text{total overall}} \times 100\% \quad (2.10)$$

### 2.1.7 Kemiskinan dan Kesejahteraan Rumah Tangga

Rumah tangga atau keluarga adalah bagian terkecil dari masyarakat yang terdiri dari ayah, ibu, dan anak, dan kesejahteraan adalah hal yang didambakan setiap orang. Dimana kesejahteraan merupakan kondisi kebutuhan di dalam keluarga yang telah terpenuhi dengan baik, tingkat kesejahteraan keluarga sangat berpengaruh pada kualitas keluarga, dan peran anggota keluarga tentunya penting dalam mencapai tujuan keluarga sejahtera itu sendiri (Fitriyah dkk, 2022).



Berdasarkan Bappenas (Badan Perencanaan Pembangunan Nasional), tingkat kesejahteraan penduduk dapat diukur dengan mengamati sejauh mana pengeluaran rumah tangga dialokasikan untuk berbagai kebutuhan. Secara garis besar, kebutuhan tersebut dapat dikelompokkan menjadi dua kategori utama: kebutuhan dasar, yang terutama mencakup kebutuhan akan makanan, dan kebutuhan non-dasar, yang mencakup aspek-aspek lain selain makanan. Dalam konteks ini, penting untuk memahami bahwa distribusi pengeluaran rumah tangga dapat memberikan gambaran tentang sejauh mana individu atau keluarga dapat memenuhi kebutuhan-kebutuhan esensial dalam kehidupan sehari-hari. Konsep ini mengakui bahwa sektor makanan memiliki urgensi yang lebih tinggi dibandingkan kebutuhan bukan makanan. Suatu rumah tangga dianggap dalam keadaan sejahtera jika pengeluaran mereka untuk kebutuhan makanan melebihi atau setidaknya setara dengan pengeluaran untuk kebutuhan selain makanan. Hal ini menandakan bahwa mereka dapat memprioritaskan pemenuhan kebutuhan makanan yang mendasar, yang penting bagi kelangsungan hidup dan kesejahteraan fisik mereka. Ini juga mengindikasikan bahwa rumah tangga tersebut memiliki kapasitas untuk memenuhi kebutuhan non makanan yang juga penting dalam peningkatan kualitas hidup.

Sebaliknya, rumah tangga dianggap sebagai kelompok miskin jika pengeluaran mereka untuk kebutuhan makanan lebih besar daripada pengeluaran untuk kebutuhan selain makanan. Sesuai dengan definisi dari BPS (Badan Pusat Statistik), penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan (BPS,2020). Dalam situasi ini, ketidakseimbangan ini dapat mengisyaratkan bahwa rumah tangga

tersebut mungkin mengalami kesulitan dalam memenuhi kebutuhan dasar yang terkait dengan makanan, yang secara langsung dapat memengaruhi tingkat kesejahteraan mereka. Kesimpulannya, pendekatan yang digunakan oleh Bappenas dalam mengukur status kesejahteraan penduduk melalui proporsi pengeluaran untuk kebutuhan rumah tangga mengakui pentingnya prioritas dalam pemenuhan kebutuhan makanan dan bukan makanan. Distribusi pengeluaran ini dapat memberikan pandangan yang lebih mendalam tentang kemampuan individu atau rumah tangga untuk mengakses kebutuhan dasar, serta memberikan arah dalam upaya pemerintah untuk meningkatkan kesejahteraan penduduk secara keseluruhan.

Secara umum, kesejahteraan sosial dapat didefinisikan sebagai keadaan dari kebutuhan dasar seperti makanan, pakaian, perumahan, pendidikan dan perawatan kesehatan yang dapat terpenuhi. Badan Pusat Statistik mengukur kemiskinan menggunakan konsep kebutuhan dasar tersebut. The World Bank dalam Permatasari dan Yuliana mengungkapkan bahwa kemiskinan merupakan kekurangan dalam kesejahteraan. Misalnya, jika seseorang sangat kekurangan dalam kesejahteraan maka dia sedang berada dalam kemiskinan. Istilah keduanya saling terkait, hanya saja memiliki cara pandang yang berbeda. Dengan begitu, untuk mengetahui tingkat kesejahteraan rumah tangga dapat digunakan kriteria pengkategorian rumah tangga miskin dan tidak miskin dengan asumsi yang dijelaskan sebelumnya yaitu kesejahteraan merupakan negasi dari kemiskinan (Permatasari & Yuliana, 2021).

### 2.1.8 Faktor-Faktor Kesejahteraan Rumah Tangga

Aspek rumah tangga sejahtera dapat dilihat dari kemampuan pemenuhan dasar, seperti kebutuhan pokok pangan, sandang, papan, kesehatan, dan pendidikan. Selain pengeluaran rumah tangga sebagai indikator kesejahteraan rumah tangga, faktor-faktor yang mempengaruhi status kesejahteraan rumah tangga dapat diukur dari indikator lainnya. Indikator kesejahteraan rumah tangga dapat dilihat dari variabel yang berhubungan dengan kepala rumah tangga (KRT), yaitu daerah tempat tinggal, jenis kelamin, banyaknya anggota rumah tangga, ijazah yang dimiliki, dan lapangan usaha dari kepala rumah tangga (BPS, 2021).

Haughton & Khandker mengungkapkan bahwa penting untuk mengetahui penjelasan dari penyebab kemiskinan. Penyebab utama yang berkaitan dengan kemiskinan sebagai berikut (Permatasari & Yuliana, 2021):

- Karakteristik tingkat wilayah, mencakup kerentanan terhadap bencana alam (banjir atau topan), keterpencilan, kualitas tata kelola dan hak milik serta penegakannya.
- Karakteristik tingkat komunitas, mencakup ketersediaan dari infrastruktur (listrik, air, batu) dan layanan (kesehatan dan pendidikan), kedekatan pada pasar, dan hubungan sosial.
- Karakteristik rumah tangga dan individual, mencakup
  1. Demografi, seperti jumlah anggota rumah tangga, struktur umur, angka ketergantungan, dan jenis kelamin KRT.
  2. Ekonomi, seperti status pekerjaan, jam kerja, dan kepemilikan aset.
  3. Sosial, seperti kesehatan dan status nutrisi, pendidikan dan perlindungan.

## 2.2 Kajian Islam

Kesejahteraan dalam islam mencakup pembebasan individu dari kemiskinan, kebodohan dan rasa takut, sehingga mereka dapat mencapai kehidupan yang aman dan tenteram secara lahir dan batin. Kesejahteraan bukan hanya terpenuhi harta tetapi juga iman. Kenikmatan dunia harus terpenuhi begitu juga dengan ibadah untuk akhirat, keduanya harus seimbang. Firman Allah SWT tentang keseimbangan ada dalam surat Al-Qashash ayat 77 sebagai berikut:

وَأَبْتَعْ فِيْمَا ءَاتَكَ اللهُ الدَّارَ الْآخِرَةَ وَلَا تَنْسَ نَصِيْبَكَ مِنَ الدُّنْيَا وَأَحْسِنَ كَمَا أَحْسَنَ اللهُ إِلَيْكَ وَلَا تَتَّبِعِ الْفُسَادَ فِي الْأَرْضِ إِنَّ اللَّهَ لَا يُحِبُّ الْمُفْسِدِينَ ۝ (سُورَةُ الْقَصَصِ)

Artinya:

*“Dan carilah pada apa yang telah dianugerahkan Allah kepadamu (kebahagiaan) negeri akhirat, dan janganlah kamu melupakan bahagiamu dari (kenikmatan) duniawi dan berbuat baiklah (kepada orang lain) sebagaimana Allah telah berbuat baik kepadamu, dan janganlah kamu berbuat kerusakan di (muka) bumi. Sesungguhnya Allah tidak menyukai orang-orang yang berbuat kerusakan.”*

Ibnu Katsir menjelaskan bahwa pada ayat diatas, kita dianjurkan untuk menggunakan harta dan nikmat sebagai bentuk ibadah, dengan tujuan untuk mendekatkan diri kepada Allah SWT. Melakukan berbagai macam kebaikan agar mendapatkan pahala di dunia dan kebaikan di akhirat. Allah SWT memperbolehkan hamba-Nya untuk makan, minum, pakaian, rumah dan menikah karena memiliki kewaiban terhadap terhadap tuhan, diri sendiri dan keluarga yang harus dipenuhi. Serta agar selalu berbuat baik sesama makhluk sebagaimana Allah SWT berbuat baik kepada kita. Dalam konteks ini, kita harus menghindari perbuatan merusak lingkungan dan berperilaku buruk terhadap ciptaan Allah SWT (Tafsir Al-qur'an,2019).

Menurut Zubdatut Tafsir min Fathil Qadir ayat diatas ditafsirkan agar membelanjakan harta pada apa yang diridhai Allah SWT, bukan untuk menyombongkan diri. Pada ayat ini juga dilarang melalaikan bagian di dunia dan menikmati harta yang halal. Dengan kenikmatan yang telah Allah SWT berikan di dunia maka kita dilarang bermaksiat kepada Allah SWT di bumi dan tidak membuat kerusakan (Tafsir web, 2022).

Berdasarkan ayat dan penjelasan diatas, dapat diketahui bahwa Allah SWT memerintahkan kita agar seimbang antara dunia dan akhirat. Allah SWT tidak melarang kita meraih kenikmatan dunia tetapi juga dengan memperhatikan akhiratnya. Di dunia ini yang kita miliki adalah pemberian Allah SWT yang harus kita jaga dan syukuri.

Kesejahteraan juga disinggung di al-qur'an dalam surat An-Nahl ayat 97 sebagai berikut:

مَنْ عَمِلَ صَالِحًا مِنْ ذَكَرٍ أَوْ أُنْثَىٰ وَهُوَ مُؤْمِنٌ فَلَنُحْيِيَنَّهٗ حَيَاةً طَيِّبَةً وَلَنَجْزِيَنَّهُمْ أَجْرَهُمْ بِأَحْسَنِ مَا كَانُوا يَعْمَلُونَ (سُورَةُ النَّحْلِ)

Artinya: “Barangsiapa yang mengerjakan amal saleh, baik laki-laki maupun perempuan dalam keadaan beriman, maka sesungguhnya akan Kami berikan kepadanya kehidupan yang baik dan sesungguhnya akan Kami beri balasan kepada mereka dengan pahala yang lebih baik dari apa yang telah mereka kerjakan.”

Menurut tafsir al-muyasar, ayat tersebut menjelaskan bahwa siapapun yang melakukan hal baik atau beramal shalih yang berjenis kelamin laki-laki maupun perempuan dan beriman kepada Allah SWT maka akan diberikan kehidupan yang bahagia dan tenteram di dunia meskipun tidak memiliki harta yang banyak serta balasan akhirat yang mulia dan lebih baik dari amalan mereka di dunia.

Berdasarkan ayat dan penjelasan diatas, dapat diketahui bahwa kesejahteraan bisa didapatkan oleh siapa saja yang mau beramal shalih tanpa

memandang jenis kelamin ataupun bentuk fisik seseorang. Allah SWT menjanjikan kehidupan yang baik di dunia dan balasan yang baik di akhirat bagi siapa saja yang beriman dan beramal shalih. Kehidupan yang baik dapat didefinisikan sebagai kehidupan yang aman, tenteram, rezeki yang lapang serta terhindar dari rasa takut dan kesulitan yang dihadapinya. Al-Qur'an telah menyinggung indikator kesejahteraan dalam surat Quraisy ayat 3-4:

فَلْيَعْبُدُوا رَبَّ هَذَا الْبَيْتِ ۖ الَّذِي أَطْعَمَهُمْ مِنْ جُوعٍ وَآمَنَهُمْ مِنْ خَوْفٍ ۖ (سُورَةُ قُرَيْشٍ)

Artinya: “Maka hendaklah mereka menyembah Tuhan (pemilik) rumah ini (Ka'bah). Yang telah memberikan makanan kepada mereka untuk menghilangkan lapar dan mengamankan mereka dari rasa takut”.

Ada tiga indikator kesejahteraan dalam Al-Qur'an, menurut ayat di atas: menyembah Tuhan (pemilik) Ka'bah, menghilangkan lapar, dan menghilangkan rasa takut. Indikator pertama yaitu beribadah, hal ini menunjukkan bahwa jika memiliki harta yang banyak belum menjamin mereka bahagia. Ada seseorang yang memiliki harta yang melimpah tetapi ada rasa tidak tenang dalam dirinya dan juga memiliki perasaan gelisah, karena inilah manusia memiliki ketergantungan kepada Allah SWT yaitu dengan beribadah untuk memiliki sama aman dan tenang dalam dirinya. Indikator kedua yaitu menghilangkan rasa lapar yaitu terpenuhinya kebutuhan konsumsi. Indikator ketiga yaitu mengamankan dari rasa takut, ini merupakan representasi dari terciptanya rasa aman dan nyaman.

### **2.3 Kajian Kesejahteraan Rumah Tangga dengan Algoritma C5.0 dan Metode *Synthetic Minority Oversampling Technique* (SMOTE)**

Rumah tangga dapat dikategorikan menjadi rumah tangga miskin dan rumah tangga tidak miskin. Ada beberapa indikator yang digunakan dalam penelitian ini untuk mengklasifikasikan rumah tangga miskin dan tidak miskin. Klasifikasi yang

digunakan dalam penelitian ini adalah Algoritma C5.0 dan metode *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani data yang tidak seimbang.

Setelah memperoleh data, data tersebut diolah dengan mengklasifikasi status kesejahteraan rumah tangga sesuai dengan definisi miskin menurut BPS dengan melihat pengeluaran per kapita per bulan berdasarkan garis kemiskinan yang berlaku di Kota Malang pada tahun 2020. Setelah mendapatkan klasifikasi status kemiskinan rumah tangga yang berupa miskin dan tidak miskin, status kemiskinan rumah tangga ditetapkan sebagai variabel respon dan 6 variabel lainnya sebagai variabel prediktor. Kemudian melakukan analisis statistika deskriptif pada data untuk mengetahui karakteristik dari variabel-variabel yang akan dianalisis.

Kemudian dilakukan pembagian data menjadi dua bagian, yaitu data *training* dan data *testing* dengan beberapa perbandingan. Setelah mendapatkan kombinasi terbaik dilakukan penanganan data tidak seimbang dengan metode SMOTE pada data *training* untuk mendapatkan tingkat akurasi yang lebih baik. Metode SMOTE menghasilkan data sintesis baru dengan data yang seimbang.

Proses pengklasifikasian dilakukan dengan Algoritma C5.0. Pemilihan *node* akar diawali dengan perhitungan *entropy*. Kemudian dilanjutkan dengan perhitungan *gain*. Selanjutnya dilakukan perhitungan *gain ratio*. Setelah itu, menentukan cabang untuk masing-masing *node* dengan menghitung nilai *gain ratio* tertinggi dari variabel prediktor, lalu kelas dibagi dalam cabang yang telah ditentukan. Langkah menentukan cabang untuk masing-masing *node* dilakukan berulang-ulang sampai semua kelas memiliki cabangnya masing-masing dan menghasilkan pohon keputusan.

Setelah prediksi dari Algoritma C5.0 muncul, dilakukan perhitungan tingkat akurasi model menggunakan *confusion matrix* hingga diperoleh nilai *accuracy*, *sensitivity*, dan *specificity*. Kemudian dilakukan juga perhitungan AUC-ROC untuk mengetahui seberapa baik model dapat memisahkan kelas positif dan negatif. Terakhir, dari proses Algoritma C5.0 akan diketahui variabel apa yang berpengaruh dan juga dilakukan perhitungan untuk mendapatkan variabel terpenting pada klasifikasi kesejahteraan rumah tangga di Kota Malang.



## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Jenis Penelitian**

Pendekatan penelitian yang digunakan adalah pendekatan deskriptif kuantitatif. Penelitian deskriptif kuantitatif meneliti populasi atau sampel tertentu dengan menggunakan analisis data kuantitatif atau ilmu statistika. Pendekatan ini menekankan analisis data numerik, menghasilkan kesimpulan yang akan memperjelas gambaran objek dan menunjukkan hubungan penting antara variabel yang diteliti.

#### **3.2 Data dan Sumber Data**

Data yang dianalisis dalam penelitian ini merupakan data sekunder yang diperoleh dari Survei Sosial dan Ekonomi Nasional (SUSENAS) tahun 2020 yang dilakukan oleh Badan Pusat Statistik (BPS). Sumber data ini dapat diakses melalui situs web <https://silastik.bps.go.id/>. Kegiatan SUSENAS dilakukan pada bulan Maret 2020 dengan penyajian pada tingkat kabupaten atau kota. Jumlah sampel yang digunakan dalam penelitian ini adalah sebanyak 775.

#### **3.3 Variabel dan Jenis Data**

Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon ( $Y$ ) dan variabel prediktor ( $X$ ) sebagai berikut:

1. Variabel Respon

Variabel respon yang digunakan adalah status kesejahteraan rumah tangga yang diperoleh dari pengeluaran per kapita per bulan berdasarkan garis kemiskinan Kota Malang tahun 2020 yaitu 554.791 rupiah per kapita per bulan. Rumah tangga dikategorikan miskin ketika pengeluaran per kapita

per bulan di bawah garis kemiskinan dan rumah tangga dikategorikan tidak miskin jika pengeluaran perkapita per bulan diatas garis kemiskinan, sehingga dapat dikategorikan menjadi:

$Y = 0$ , untuk kategori miskin

$Y = 1$ , untuk kategori tidak miskin

Skala data pada variabel status kesejahteraan rumah tangga adalah nominal.

## 2. Variabel Prediktor

**Tabel 3.1** Variabel Prediktor

No	Variabel		Skala	Keterangan
1.	Jenis Kelamin kepala rumah tangga	X1	Nominal	1 : Laki-laki 2 : Perempuan
2.	Usia kepala rumah tangga	X2	Nominal	1 : produktif 2 : nonproduktif
3.	Banyak anggota rumah tangga	X3	Nominal	1 : $\leq 4$ orang 2 : $> 4$ orang
4.	Tingkat Pendidikan tertinggi kepala rumah tangga	X4	Ordinal	1 : SD sederajat 2 : SMP sederajat 3 : SMA sederajat 4 : Perguruan tinggi
5.	Status Pekerjaan kepala rumah tangga	X5	Nominal	1 : Berusaha sendiri 2 : Berusaha dibantu buruh tidak tetap 3 : Berusaha dibantu buruh tetap 4 : Buruh / karyawan / karyawan 5 : Pekerja bebas 6 : Pekerja keluarga atau tidak dibayar

No	Variabel		Skala	Keterangan
6.	Lapangan Usaha kepala rumah tangga	X6	Nominal	1 : Pertanian 2 : Industri 3 : Jasa

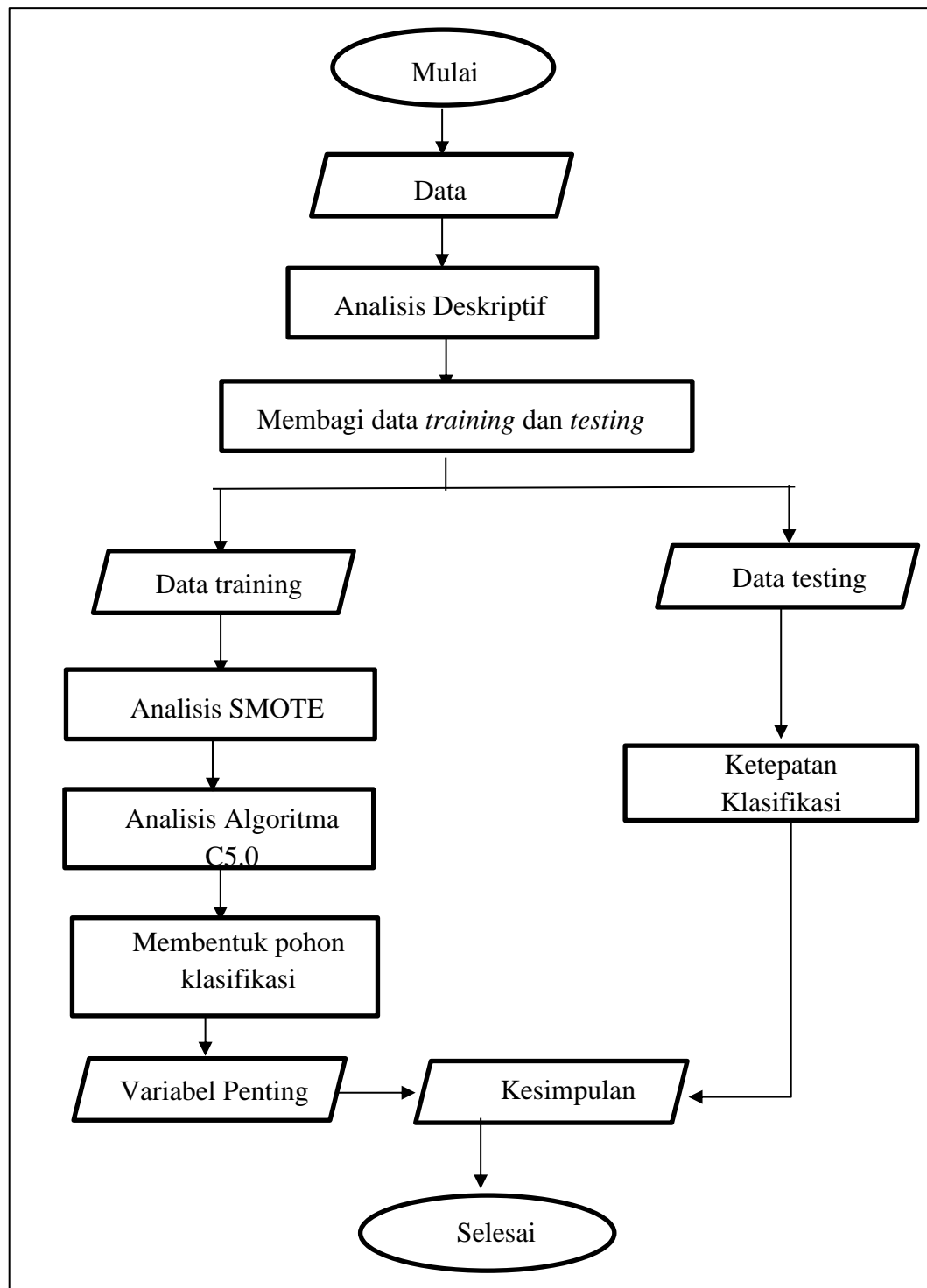
### 3.4 Teknik Analisis Data

Penelitian ini menggunakan Algoritma C5.0 untuk klasifikasi dan metode SMOTE untuk penanganan data tidak seimbang dengan bantuan *software* Rstudio dengan tahapan sebagai berikut :

1. Mengklasifikasi status kesejahteraan rumah tangga sesuai dengan definisi miskin menurut BPS dengan melihat pengeluaran per kapita per bulan berdasarkan garis kemiskinan yang berlaku di Kota Malang pada tahun 2020.
2. Melakukan analisis statistika deskriptif pada variabel-variabel penelitian yang digunakan.
3. Penerapan metode SMOTE pada data *training* untuk menyeimbangkan data.
  - a. Menghitung jarak antar amatan pada kelas minor menggunakan persamaan (2.4).
  - b. Menentukan nilai  $k$  dan persentase *oversampling*.
  - c. Dipilih satu contoh dari kelas minor secara acak.
  - d. Menentukan amatan  $k$  tetangga terdekat dengan mengurut jarak contoh terpilih dengan semua amatan pada kelas minor.

- e. Data sintesis dibuat dengan menentukan nilai per peubah penjelasnya. Nilai tersebut diperoleh dari mayoritas nilai pada  $k$  tetangga terdekat. Jika semua peubah telah dibuat maka diperoleh satu amatan baru.
  - f. Ulangi langkah c-e hingga banyaknya oversampling yang diinginkan tercapai.
4. Membagi data menjadi dua bagian, yaitu data *training* dan data *testing* dengan variasi perbandingan yang berbeda, yakni 90%:10%, 85%:15%, 80%:20% dan 75%:25%. Pembagian ini dilakukan secara acak.
  5. Penerapan Algoritma C5.0 untuk klasifikasi.
    - a. Pemilihan *node* akar dimulai dengan perhitungan nilai *entropy* menggunakan persamaan (2.1).
    - b. Mencari nilai *gain* menggunakan persamaan (2.2).
    - c. Mencari nilai *gain ratio* menggunakan persamaan (2.3).
    - d. Menentukan cabang untuk masing-masing *node* dengan menghitung nilai *gain ratio* tertinggi dari variabel prediktor.
    - e. Kelas dibagi dalam cabang yang telah ditentukan.
    - f. Ulangi langkah a-c hingga semua kelas memiliki cabangnya masing-masing.
  6. Menentukan tingkat akurasi model menggunakan *confusion matrix*.
  7. Menghitung ketepatan klasifikasi dengan metode AUC-ROC pada persamaan (2.9).
  8. Menghitung variabel penting dengan persamaan (2.10).

Tahapan-tahapan berikut akan digambarkan dalam flowchart pada Gambar 3.1



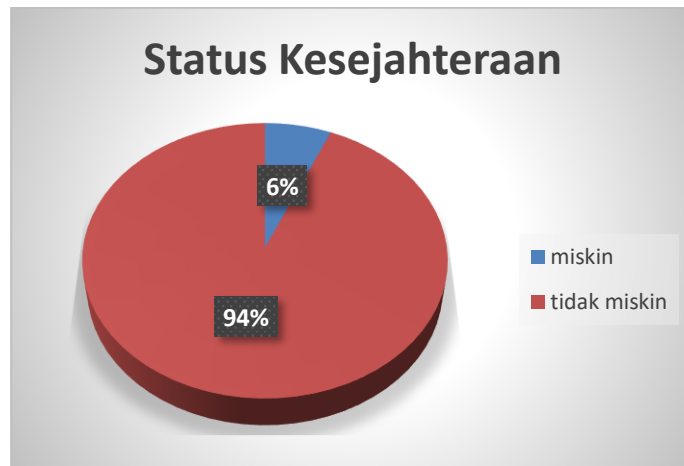
Gambar 3.1 Flowchart

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Analisis Statistik Deskriptif

Analisis statistik deskriptif merupakan analisis statistik yang bertujuan untuk menggambarkan karakteristik dari masing-masing variabel yang akan diteliti. Status kesejahteraan rumah tangga dikategorikan menjadi miskin dan tidak miskin. Gambaran mengenai status kesejahteraan di kota Malang disajikan dalam Gambar 4.1



**Gambar 4.1** Status Kesejahteraan Rumah Tangga

Gambar 4.1 menjelaskan bahwa status kesejahteraan rumah tangga di Kota Malang lebih banyak rumah tangga tidak miskin atau sejahtera dibandingkan dengan keluarga miskin. Tercatat dari 775 sampel data terdapat 94 persen atau sebanyak 725 rumah tangga tidak miskin. Sedangkan, 6 persen atau 50 rumah tangga di kategorikan rumah tangga miskin. Status kesejahteraan rumah tangga dapat dikategorikan miskin dan tidak miskin berdasarkan pengeluaran per kapita rumah tangga per bulan dan garis kemiskinan daerah tersebut. Pada tahun 2020, angka garis kemiskinan di Kota Malang adalah Rp.554.791. Sedangkan, deskripsi

pengeluaran per kapita per bulan rumah tangga di Kota Malang tahun 2020 ditunjukkan pada Tabel 4.1.

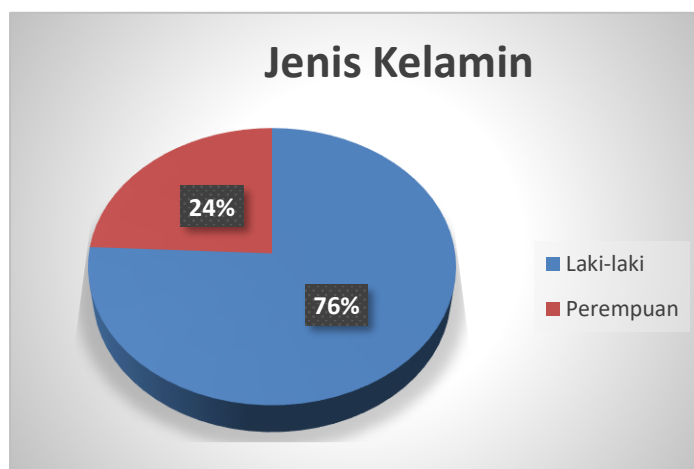
**Tabel 4.1** Deskripsi Pengeluaran Per Kapita Per Bulan

	<i>Mean</i>	<i>Median</i>	<i>Minimum</i>	<i>Maximum</i>
<b>Pengeluaran Per Kapita Per Bulan</b>	2.067.336,227	1.651.213,381	308.629,167	20.896.811,26

Tabel 4.1 menunjukkan bahwa rata-rata pengeluaran per kapita rumah tangga Kota Malang tahun 2020 bernilai 2.067.336 rupiah dengan median atau nilai tengah sebesar 1.651.213 rupiah. Nilai minimum atau nilai terendah pengeluaran per kapita per bulan yaitu 308.629,167 rupiah dan nilai maximum atau nilai tertinggi pada angka 20.896.811,26 rupiah.

#### 4.1.1 Deskripsi Jenis Kelamin Kepala Rumah Tangga

Kepala rumah tangga adalah anggota rumah tangga yang bertanggung jawab atas kebutuhan sehari-hari rumah tangga. Jenis kelamin kepala rumah tangga termasuk salah satu faktor yang memengaruhi status kesejahteraan rumah tangga. Gambaran mengenai jenis kelamin kepala rumah tangga disajikan dalam gambar 4.2.



**Gambar 4.2** Jenis Kelamin

Gambar 4.2 menunjukkan bahwa kepala rumah tangga berjenis kelamin laki-laki lebih banyak dibandingkan dengan kepala rumah tangga yang berjenis kelamin perempuan. Tercatat dari 775 sampel, sebesar 76 persen atau sebanyak 587 kepala rumah tangga berjenis kelamin laki-laki. Sedangkan, sebesar 24 persen atau sebanyak 188 kepala rumah tangga berjenis kelamin perempuan. Deskripsi jenis kelamin kepala rumah tangga berdasarkan status kesejahteraan rumah tangga ditunjukkan pada tabel 4.2

**Tabel 4.2** Jenis Kelamin Berdasarkan Status Kesejahteraan

		Status Kesejahteraan		Total
		Miskin	Tidak Miskin	
Jenis Kelamin	Laki-Laki	42	545	587
	Perempuan	8	180	188
Total		50	725	775

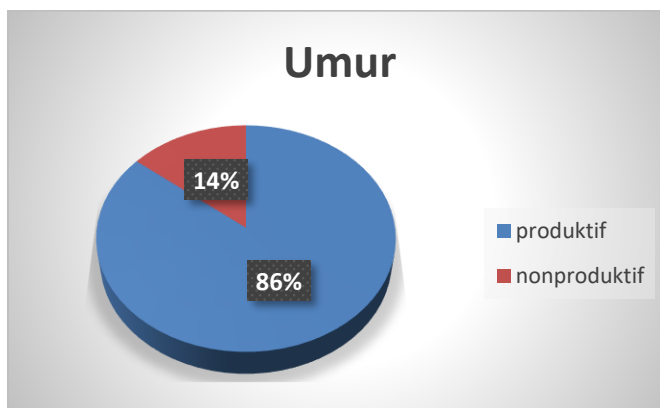
Tabel 4.2 menjelaskan tentang jenis kelamin kepala rumah tangga berdasarkan status kesejahteraan rumah tangga. Tercatat dari 50 rumah tangga miskin terdapat sebanyak 42 kepala keluarga berjenis kelamin laki-laki dan sebanyak 8 kepala keluarga berjenis kelamin perempuan. Sedangkan, pada rumah tangga tidak miskin atau sejahtera terdapat 545 kepala keluarga berjenis kelamin laki-laki dan sisanya 180 kepala rumah tangga adalah perempuan. Sehingga dapat disimpulkan bahwa mayoritas kepala keluarga rumah tangga di Kota Malang berjenis kelamin laki-laki baik pada rumah tangga miskin ataupun tidak miskin.

#### **4.1.2 Deskripsi Usia Kepala Rumah Tangga**

Usia kepala rumah tangga termasuk salah satu faktor yang mempengaruhi kesejahteraan rumah tangga. Variabel usia dikategorikan menjadi 2 yaitu usia



produktif dan nonproduktif. Berdasarkan Badan Pusat Statistik (BPS), usia produktif adalah 15 sampai 64 tahun dan usia nonproduktif adalah lebih dari 64 tahun. Gambaran umur kepala rumah tangga ditunjukkan pada gambar 4.3.



**Gambar 4.3** Umur

Gambar 4.3 menjelaskan bahwa kepala rumah tangga yang berusia produktif berjumlah 665 atau 86 persen. Kemudian kepala rumah tangga yang berusia nonproduktif berjumlah 110 atau 14 persen. Deskripsi umur kepala rumah tangga berdasarkan status kesejahteraan rumah tangga ditunjukkan pada tabel 4.3

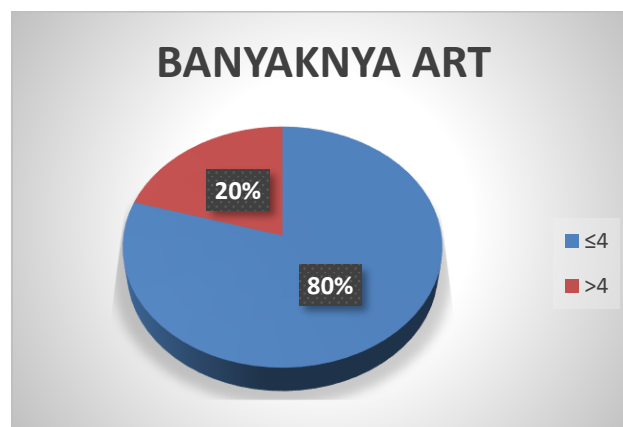
**Tabel 4.3** Usia berdasarkan Status Kesejahteraan

		Status Kesejahteraan		Total
		Miskin	Tidak Miskin	
Umur	Produktif	42	623	665
	Non Produktif	8	102	110
	Total	50	725	775

Berdasarkan tabel 4.3 diketahui bahwa pada kategori miskin terdapat 42 KRT berusia produktif dan 8 KRT berusia nonproduktif. Sedangkan pada 725 data kategori tidak miskin terdapat kepala rumah tangga berusia produktif sebanyak 623 KRT dan kepala rumah tangga berusia nonproduktif ada sebanyak 102 KRT.

### 4.1.3 Deskripsi Banyak Anggota Rumah Tangga

Banyak anggota rumah tangga merupakan salah satu faktor yang mempengaruhi status kesejahteraan rumah tangga. Banyaknya anggota rumah dikategorikan menjadi 2 yaitu  $\leq 4$  dan  $>4$ , kategori tersebut berdasarkan anjuran pemerintah untuk keluarga berencana dengan mempunyai 2 anak. Gambaran banyaknya anggota rumah tangga ditunjukkan pada gambar 4.4.



**Gambar 4.4** Banyaknya ART

Gambar 4.4 menjelaskan bahwa anggota rumah tangga yang banyaknya  $\leq 4$  memiliki persentase yang tinggi yaitu sebanyak 80 persen atau berjumlah 617 dan banyaknya anggota rumah tangga  $>4$  memiliki persentase 20 persen atau berjumlah 158. Deskripsi banyaknya anggota rumah tangga menurut status kesejahteraan rumah tangga ditunjukkan pada tabel 4.4

**Tabel 4.4** Banyaknya ART berdasarkan Status Kesejahteraan

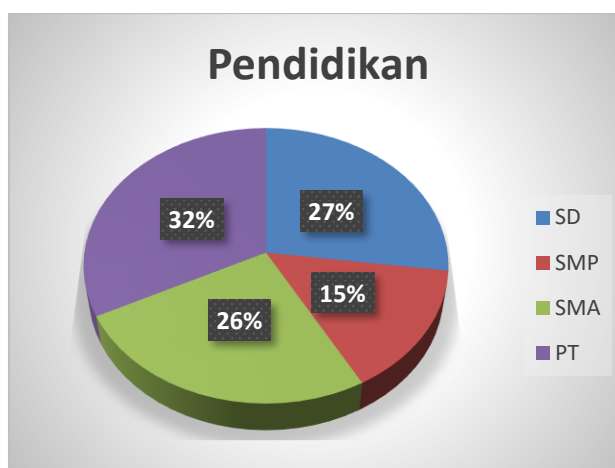
		Status Kesejahteraan		Total
		Miskin	Tidak Miskin	
Banyak ART	$\leq 4$	24	593	617
	$>4$	26	132	158
Total		50	725	775

Tabel 4.1.4 menunjukkan bahwa, dari 50 kategori miskin terdapat 24 yang banyak anggotanya  $\leq 4$  dan sebanyak 26 yang memiliki banyak anggota rumah

tangga >4. Selanjutnya dari kategori tidak miskin yang berjumlah 725 terdapat 593 rumah tangga yang beranggota  $\leq 4$ . Kemudian banyaknya anggota rumah tangga >4 pada kategori tidak miskin berjumlah 132.

#### 4.1.4 Deskripsi Tingkat Pendidikan Kepala Rumah Tangga

Tingkat pendidikan kepala rumah tangga termasuk salah satu faktor yang mempengaruhi status kesejahteraan rumah tangga. Gambaran mengenai tingkat pendidikan kepala rumah tangga disajikan pada gambar berikut.



Gambar 4.5 Tingkat Pendidikan

Gambar 4.5 menjelaskan mengenai tingkat pendidikan kepala rumah tangga. Tingkat pendidikan terbanyak adalah perguruan tinggi sebesar 32 persen atau sebanyak 251 kepala rumah tangga, selanjutnya yaitu pendidikan SD sebesar 27 persen atau sebanyak 210. Kemudian, pada tingkat pendidikan SMA sebanyak 26 persen atau sebanyak 199 dan yang terakhir pada tingkat pendidikan SMP dengan angka terendah yaitu 15 persen atau sebanyak 110 kepala rumah tangga. Deskripsi mengenai tingkat pendidikan berdasarkan status kesejahteraan rumah tangga ditunjukkan pada tabel 4.5

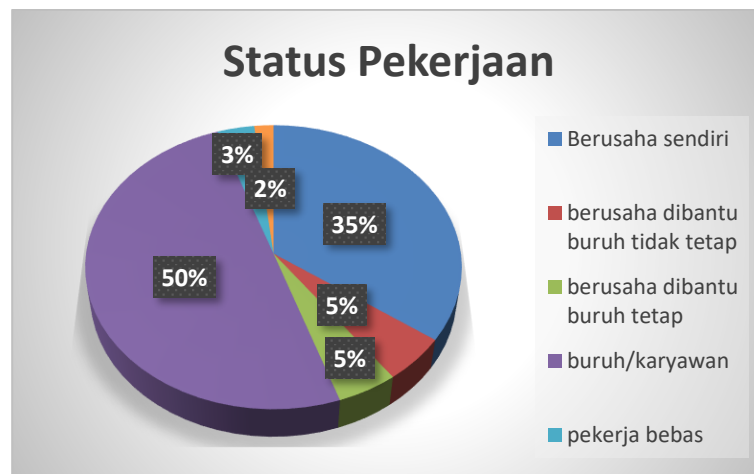
**Tabel 4.5** Pendidikan berdasarkan Status Kesejahteraan

		Status Kesejahteraan		Total
		Miskin	Tidak Miskin	
Pendidikan	SD	23	187	210
	SMP	12	103	115
	SMA	15	184	199
	Perguruan Tinggi	0	251	251
Total		50	725	775

Tabel 4.5 menjelaskan tentang tingkat pendidikan terakhir kepala rumah tangga berdasarkan status kesejahteraan rumah tangga. Tercatat dari 50 kategori miskin, terdapat 23 KRT yang berpendidikan SD, kemudian sebanyak 12 KRT berpendidikan SMP dan 15 KRT berpendidikan SMA. Pada kategori miskin tidak ada KRT yang pendidikan terakhirnya perguruan tinggi. Kemudian, di kategori tidak miskin yang berjumlah 725, terdapat 187 KRT berpendidikan SD, terdapat 103 KRT berpendidikan SMP, selanjutnya terdapat 184 KRT berpendidikan SMA dan yang terbanyak adalah KRT yang berpendidikan perguruan tinggi yaitu sebanyak 251.

#### **4.1.5 Deskripsi Status Pekerjaan Kepala Rumah Tangga**

Status pekerjaan adalah jenis kedudukan kepala rumah tangga dalam pekerjaan utama. Status pekerjaan kepala rumah tangga termasuk faktor yang berpengaruh pada status kesejahteraan rumah tangga. Gambaran mengenai status pekerjaan kepala rumah tangga disajikan pada Gambar 4.6.



**Gambar 4.6** Status Pekerjaan

Gambar 4.6 menjelaskan mengenai status pekerjaan kepala rumah tangga. Status pekerjaan paling banyak yaitu sebagai buruh/karyawan dengan persentase 50 persen atau berjumlah 387 KRT. Selanjutnya terdapat status pekerjaan berusaha sendiri sebanyak 270 atau 35 persen. KRT yang status pekerjaannya berusaha dibantu buruh tidak tetap sebanyak 40 KRT atau 5 persen. Pada status pekerjaan berusaha dibantu buruh tetap ada sebanyak 37 atau 5 persen. Kemudian KRT yang status pekerjaannya pekerja bebas terdapat 27 KRT atau 3 persen. Terakhir, KRT yang status pekerjaannya pekerja keluarga terdapat 14 KRT atau 2 persen. Deskripsi mengenai status pekerjaan berdasarkan status kesejahteraan rumah tangga ditunjukkan pada tabel 4.6.

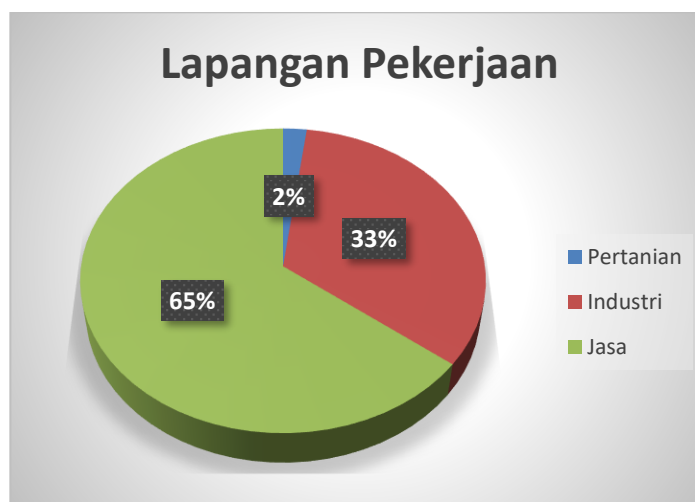
**Tabel 4.6** Status pekerjaan berdasarkan Status Kesejahteraan

		Status Kesejahteraan		Total
		Miskin	Tidak Miskin	
Status Pekerjaan	Berusaha sendiri	15	255	270
	berusaha dibantu buruh tidak tetap	1	39	40
	berusaha dibantu buruh tetap	0	37	37
Status Pekerjaan	buruh/karyawan	30	357	387
	pekerja bebas	3	24	27
	pekerja keluarga	1	13	14
Total		50	725	775

Tabel 4.6 menjelaskan mengenai status pekerjaan kepala rumah tangga berdasarkan status pekerjaan. Pada kategori miskin yang berjumlah 50, terdapat 15 KRT yang status pekerjaannya berusaha sendiri, 1 KRT yang berusaha dibantu buruh tidak tetap, 30 karyawan/buruh, 3 pekerja bebas dan 1 pekerja keluarga. Selanjutnya pada 725 KRT yang kategori tidak miskin, terdapat 255 KRT yang status pekerjaannya berusaha sendiri, 39 KRT yang berusaha dibantu buruh tidak tetap, 37 KRT yang berusaha dibantu buruh tetap, yang terbanyak yaitu KRT yang status pekerjaannya sebagai karyawan yaitu 357 KRT. Selanjutnya ada 24 KRT yang status pekerjaannya sebagai pekerja bebas dan terakhir sebanyak 13 KRT yang status pekerjaannya sebagai pekerja keluarga.

#### 4.1.6 Lapangan Pekerjaan Kepala Rumah Tangga

Lapangan pekerjaan kepala rumah tangga menjadi faktor yang berpengaruh terhadap kesejahteraan rumah tangga. Gambaran mengenai lapangan pekerjaan kepala rumah tangga terdapat pada Gambar 4.7.



**Gambar 4.7** Lapangan Pekerjaan

Gambar 4.7 menjelaskan mengenai lapangan pekerjaan kepala rumah tangga. Lapangan pekerjaan terbanyak terdapat pada bidang jasa yaitu memiliki 65 persen atau berjumlah 500 KRT. Selanjutnya terdapat 33 persen atau sebanyak 259 KRT yang bekerja di bidang industri. Terakhir, dengan persentase paling sedikit terdapat 2 persen atau 16 KRT yang bekerja di bidang pertanian. Deskripsi mengenai lapangan pekerjaan berdasarkan status kesejahteraan rumah tangga ditunjukkan pada Tabel 4.7

**Tabel 4.7** Lapangan Pekerjaan berdasarkan Status Kesejahteraan

		Status Kesejahteraan		Total
		Miskin	Tidak Miskin	
Lapangan Pekerjaan	Pertanian	1	15	16
	Industri	14	245	259
	Jasa	35	465	500
Total		50	725	775

Tabel 4.7 menjelaskan mengenai lapangan pekerjaan kepala rumah tangga berdasarkan status kesejahteraan rumah tangga. Tercatat dari 50 kategori miskin, terdapat 35 KRT yang bekerja di bidang jasa, 14 KRT yang bekerja di bidang industri, dan 1 KRT yang bekerja di bidang pertanian. Kemudian, pada 725 KRT kategori tidak miskin, terdapat 465 KRT yang bekerja di bidang jasa, 245 KRT yang bekerja di bidang industri dan yang paling sedikit di bidang pertanian yaitu berjumlah 16 KRT.

#### 4.2 *Synthetic Minority Oversampling Technique (SMOTE)*

Metode SMOTE digunakan untuk mengatasi data yang tidak seimbang. Metode SMOTE hanya digunakan pada data training untuk meningkatkan kinerja klasifikasi. Data berjumlah 775 data, dari data tersebut dibagi menjadi dua kategori yaitu miskin dan tidak miskin. Tercatat dari 775 data, terdapat 50 data kategori miskin dan 725 data kategori tidak miskin. Proses pembuatan data sintesis dilakukan dengan meningkatkan data minoritas dengan percentage over sebesar 1400%, percentage under sebesar 110% dengan tetangga terdekat 5. Perbedaan data minoritas dan mayoritas sebelum dan sesudah dilakukan metode SMOTE terdapat pada Gambar 4.8



**Gambar 4.8** Penerapan SMOTE



Berdasarkan Gambar 4.8 diketahui bahwa sebelum dilakukan metode SMOTE, data pada kategori miskin dan tidak miskin sangat tidak seimbang. Sedangkan, setelah diterapkan metode SMOTE data menjadi seimbang. Pada bagan diatas dapat diketahui juga bahwa data miskin sebelum diterapkan metode SMOTE berjumlah 50, kemudian setelah diterapkan metode SMOTE data kategori miskin bertambah menjadi 750. Pada data kategori tidak miskin, sebelum diterapkan SMOTE berjumlah 720 dan setelah diterapkan metode SMOTE berjumlah 770 data.

### 4.3 Pembagian Data Training dan Data Testing

Pada penelitian ini, pembagian presentase data training dan data testing menggunakan 4 macam kombinasi presentase, yaitu: (90% : 10%), (85% : 15%), (80% : 20%), dan (75% : 25%). Selanjutnya membentuk pohon klasifikasi di setiap pembagian persentase data yang ada, kemudian dihitung ketepatan klasifikasi pada tiap-tiap klasifikasi pembagian persentase data. Setelah diperoleh hasil dari 4 ketepatan klasifikasi, dilanjutkan dengan perbandingan, kombinasi dari persentase akan dipilih jika ketepatan klasifikasinya mempunyai nilai data training yang paling baik. Data training digunakan pada proses membentuk pohon klasifikasi sedangkan data testing digunakan untuk melakukan pengukuran kelayakan atau kebaikan model. Pemilihan kombinasi terbaik ditunjukkan berdasarkan Tabel 4.8.

**Tabel 4.8** Data Training dan Data Testing

No.	Persentase Data		Ketepatan Klasifikasi	
	Training	Testing	Training	Testing
1.	90%	10%	<b>0,7654</b>	<b>0,7632</b>
2.	85%	15%	0,7709	0,7325
3.	80%	20%	0,7664	0,7401
4.	75%	25%	0,7693	0,7526

Berdasarkan Tabel 4.8 dapat diketahui bahwa dari empat kombinasi didapatkan satu kombinasi terbaik. Hasil dari kombinasi data tersebut yang dipertimbangkan adalah nilai dari ketepatan klasifikasi data *training* dan data *testing* karena dari data *training* dapat dilihat kelayakan dan ketepatan pohon klasifikasi dan dari data *testing* dapat diketahui evaluasi model klasifikasi. Kombinasi terbaik dari empat kombinasi yaitu kombinasi data *training* 90% dan data *testing* 10%, pada data *training*-nya yaitu 0,7654 dan data *testing* sebesar 0,7632. Kombinasi 90%:10% menjadi kombinasi terbaik, sehingga pada penelitian ini data dibagi menjadi dua, yaitu 90% menjadi data *training* dan 10% menjadi data *testing*. Pada penelitian ini data berjumlah 1520, sehingga 90% nya yaitu 1368 menjadi data *training* dan 10% data yaitu 152 data menjadi data *testing*.

#### **4.4 Algoritma C5.0**

Algoritma C5.0 digunakan untuk mengetahui klasifikasi status kesejahteraan rumah tangga yang dikategorikan menjadi dua, yaitu miskin dan tidak miskin. Pembagian data *training* dan data *testing* penting untuk klasifikasi. Data *training* digunakan untuk menguji model, sedangkan data *testing* digunakan untuk mengukur nilai akurasi klasifikasi. Pembagian data *training* dan data *testing* telah dilakukan sebelumnya. Pada penelitian ini, algoritma C5.0 menggunakan data *training* berjumlah 1368 data dan data *testing* berjumlah 152 data.

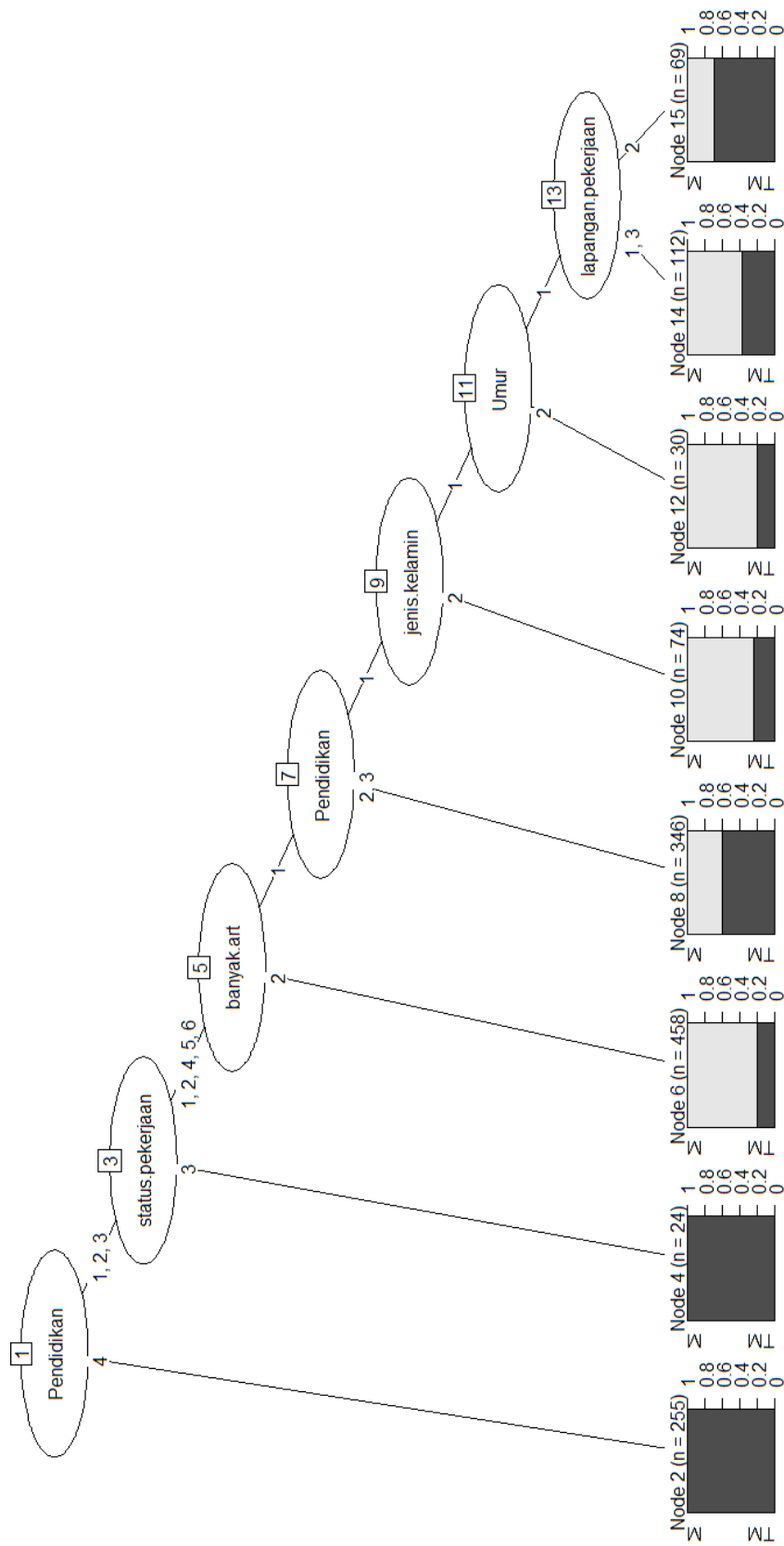
##### **4.4.1 Pohon Klasifikasi**

Pohon klasifikasi dapat dilihat pada Gambar 4.9. Pada gambar 4.9 diperoleh pohon klasifikasi dengan simpul utama (*root node*) yang dinotasikan  $t_1$  yaitu variabel pendidikan. Pada pohon klasifikasi, diketahui ada 6 simpul dalam (*internal nodes*) yaitu  $t_3$ ,  $t_5$ ,  $t_7$ ,  $t_9$ ,  $t_{11}$ ,  $t_{13}$  dan 8 simpul terminal (*terminal*

*nodes*) yaitu  $t_2, t_4, t_6, t_8, t_{10}, t_{12}, t_{14}, t_{15}$ . Berdasarkan pohon klasifikasi pada Gambar 4.9, dapat diketahui bahwa :

1. Apabila KRT yang pendidikan tertingginya Perguruan Tinggi cenderung tidak miskin.
2. Apabila KRT yang memiliki pendidikan terakhir SD, SMP, SMA dan status pekerjaannya berusaha dibantu buruh tetap cenderung tidak miskin.
3. Apabila KRT yang memiliki pendidikan terakhir SD, SMP, SMA dan status pekerjaannya berusaha sendiri, berusaha dibantu buruh tidak tetap, buruh/karyawan, pekerja bebas, pekerja keluarga yang banyaknya anggota rumah tangga lebih dari 4 cenderung miskin.
4. Apabila KRT yang status pekerjaannya berusaha sendiri, berusaha dibantu buruh tidak tetap, buruh/karyawan, pekerja bebas, pekerja keluarga yang banyaknya anggota rumah tangga kurang dari 4 dengan pendidikan terakhirnya SMP dan SMA cenderung tidak miskin.
5. Apabila KRT yang status pekerjaannya berusaha sendiri, berusaha dibantu buruh tidak tetap, buruh/karyawan, pekerja bebas, pekerja keluarga yang banyaknya anggota rumah tangga kurang dari 4 dengan pendidikan terakhirnya SD yang berjenis kelamin perempuan cenderung miskin.
6. Apabila KRT yang status pekerjaannya berusaha sendiri, berusaha dibantu buruh tidak tetap, buruh/karyawan, pekerja bebas, pekerja keluarga yang banyaknya anggota rumah tangga kurang dari 4 dengan pendidikan terakhirnya SD yang berjenis kelamin laki-laki dan memiliki umur nonproduktif yaitu lebih dari 56 tahun maka cenderung miskin.

7. Apabila KRT yang status pekerjaannya berusaha sendiri, berusaha dibantu buruh tidak tetap, buruh/karyawan, pekerja bebas, pekerja keluarga yang banyaknya anggota rumah tangga kurang dari 4 dengan pendidikan terakhirnya SD yang berjenis kelamin laki-laki dan memiliki umur produktif yaitu 15-56 tahun dan bekerja di bidang pertanian dan jasa cenderung miskin.
8. Apabila KRT yang status pekerjaannya berusaha sendiri, berusaha dibantu buruh tidak tetap, buruh/karyawan, pekerja bebas, pekerja keluarga yang banyaknya anggota rumah tangga kurang dari 4 dengan pendidikan terakhirnya SD yang berjenis kelamin laki-laki dan memiliki umur produktif yaitu 15-56 tahun dan bekerja di bidang industri cenderung tidak miskin.



Gambar 4.9 Pohon Klasifikasi

#### 4.4.2 Tingkat Akurasi Ketepatan Klasifikasi

Ketepatan klasifikasi status kesejahteraan rumah tangga dapat dilihat dari perhitungan nilai *accuracy*. Juga dapat dilihat dari nilai *sensitivity* untuk mengukur seberapa baik model dapat mengidentifikasi kategori miskin dan nilai *specificity* untuk mengukur seberapa baik model dapat mengidentifikasi kategori tidak miskin. Pengujian ketepatan klasifikasi menggunakan data *testing* sebanyak 152 sampel. Tabel *confussion matrix* bisa dilihat di Tabel 4.9

**Tabel 4.9** *Confussion Matrix*

<i>Actual Class</i>	<i>Predicted Class</i>		<b>Total</b>
	<b>Miskin</b>	<b>Tidak Miskin</b>	
<b>Miskin</b>	63	16	79
<b>Tidak Miskin</b>	20	53	73
<b>Total</b>	83	69	152

Berdasarkan Tabel 4.9 dapat diketahui bahwa jumlah data *testing* sebanyak 152 sampel., data yang diprediksi miskin ada 63 sampel yang miskin dan 20 sampel yang tidak miskin dan pada data yang diprediksi tidak miskin ada 53 sampel dikategorikan tidak miskin dan 16 sampel dikategorikan miskin.

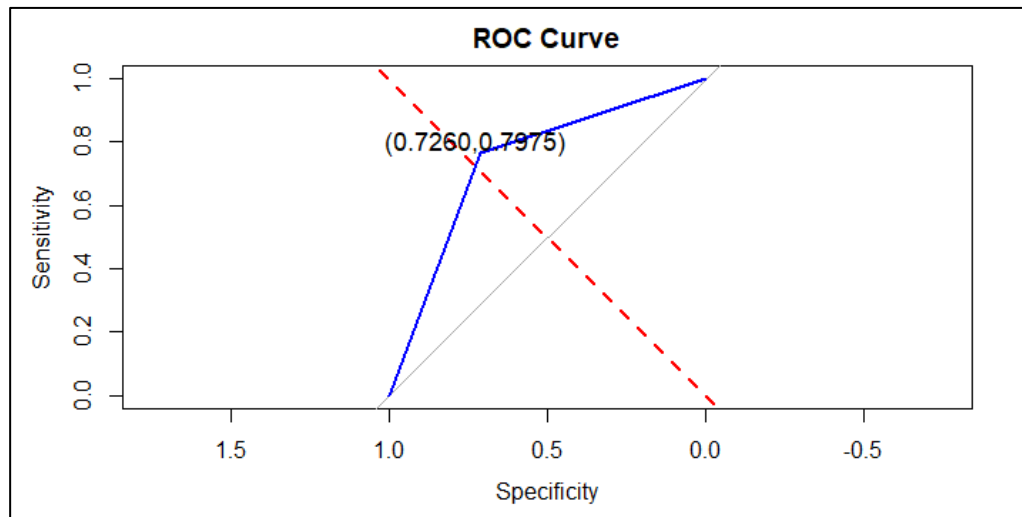
Perhitungan nilai *accuracy*, *sensitivity*, dan *specificity* yaitu:

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} = \frac{63 + 53}{63 + 16 + 20 + 53} = \frac{116}{152} = 0,7632$$

$$sensitivity = \frac{TP}{TP + FN} = \frac{63}{63 + 16} = \frac{63}{79} = 0,7975$$

$$specificity = \frac{TN}{TN + FP} = \frac{53}{53 + 20} = \frac{53}{73} = 0,7260$$

Berdasarkan perhitungan diatas, diperoleh nilai *accuracy* sebesar 0,7632 atau 76,32% dengan nilai *sensitivity* sebesar 0,7975 atau 79,75% dan nilai *specificity* sebesar 0,7260 atau 72,6%. Selanjutnya dilakukan pengujian dengan AUC-ROC. AUC-ROC memberikan gambaran yang baik tentang kemampuan model membedakan antara kelas-kelas. ROC Curve disajikan dalam Gambar 4.10



**Gambar 4.10** Curve ROC

Gambar 4.10 dapat dilihat bahwa garis berwarna biru menunjukkan titik dari nilai *specificity* dan *sensitivity*. Berdasarkan kurva ROC dapat diketahui nilai *specificity* yaitu 0,7260 dan nilai *sensitivity* yaitu 0,7975. Berikut ini adalah perhitungan AUC:

$$\begin{aligned} AUC &= \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) = \frac{1}{2} \left( \frac{63}{63 + 16} + \frac{53}{53 + 20} \right) \\ &= \frac{1}{2} (0,7260 + 0,7975) \\ &= \frac{1}{2} (1,5235) = 0,7617 \end{aligned}$$

Berdasarkan perhitungan diatas dapat dilihat bahwa nilai AUC adalah 0,7617. Nilai pengklasifikasian AUC dibagi menjadi beberapa kelompok, pada penelitian ini nilai AUC yaitu 0,7617 sehingga masuk dalam *fair classification*.

#### 4.5 Tingkat Kepentingan Variabel

Klasifikasi status kesejahteraan rumah tangga pada penelitian ini memiliki enam variabel prediktor. Pada pohon klasifikasi, dapat diketahui variabel yang paling penting pada klasifikasi status kesejahteraan rumah tangga. Perhitungan tingkat kepentingan variabel pada klasifikasi sebagai berikut:

$$VarImp = \frac{\text{nilai overall tiap variabel}}{\text{total overall}} \times 100\%$$

1. Jenis Kelamin

$$VarImp = \frac{20,83}{310,45} \times 100\% = 6,71\%$$

2. Umur

$$VarImp = \frac{15,43}{310,45} \times 100\% = 4,97\%$$

3. Banyak anggota rumah tangga

$$VarImp = \frac{79,61}{310,45} \times 100\% = 25,64\%$$

4. Pendidikan

$$VarImp = \frac{100}{310,45} \times 100\% = 32,21\%$$

5. Lapangan pekerjaan

$$VarImp = \frac{13,23}{310,45} \times 100\% = 4,26\%$$

6. Status Pekerjaan

$$VarImp = \frac{81,26}{310,45} \times 100\% = 26,17\%$$



**Tabel 4.10** Tingkat Kepentingan Variabel

No.	Variabel	Overall	Tingkat Kepentingan
1.	Pendidikan	100	32,21%
2.	Status Pekerjaan	81,36	26,17%
3.	Banyak anggota rumah tangga	79,61	25,64%
4.	Jenis Kelamin	20,83	6,71%
5.	Umur	15,42	4,97%
6.	Lapangan Pekerjaan	13,23	4,26%

Berdasarkan Tabel 4.10 dapat diketahui bahwa variabel terpenting yaitu pendidikan dengan tingkat kepentingan 32,21 persen, kemudian variabel status pekerjaan dengan persentase 26,17 persen. Selanjutnya variabel banyak anggota rumah tangga dengan persentase 25,64 persen, yang ke-empat ada variabel jenis kelamin dengan persentase 6,71 persen. Kemudian variabel umur 4,97 persen dan terakhir variabel lapangan pekerjaan dengan persentase sebesar 4,26 persen.

#### **4.6 Kajian Islam pada Hasil Penelitian**

Rumah tangga merupakan tempat bernaungnya sebuah keluarga. Bagi keluarga, rumah tangga berfungsi untuk mendapatkan ikatan saudara dalam satu kesatuan ekonomi sosial yang saling mendukung. Salah satunya saling bergantung dalam pemenuhan kebutuhan ekonomi. Kesejahteraan keluarga mencakup kesejahteraan secara materiil untuk kebutuhan jasmani dan secara moral untuk kebutuhan rohani. Harus ada keseimbangan antara keduanya untuk mencapai kesejahteraan bagi keluarga. Kesejahteraan secara materiil yaitu pemenuhan kebutuhan sehari-hari seperti sandang, pangan, dan papan. Sedangkan kesejahteraan secara moral adalah pemenuhan kebutuhan religi dan rasa kasih

sayang dalam keluarga. Al-Qur'an telah menyinggung tentang kehidupan dalam rumah tangga yaitu pada surat Ar-Ruum Ayat 21:

وَمِنْ آيَاتِهِ أَنْ خَلَقَ لَكُمْ مِنْ أَنْفُسِكُمْ أَزْوَاجًا لِتَسْكُنُوا إِلَيْهَا وَجَعَلَ بَيْنَكُمْ مَوَدَّةً وَرَحْمَةً ۗ إِنَّ فِي ذَلِكَ لَآيَاتٍ لِقَوْمٍ يَتَفَكَّرُونَ

*Artinya :” Dan di antara tanda-tanda kekuasaan-Nya ialah Dia menciptakan untukmu isteri-isteri dari jenismu sendiri, supaya kamu cenderung dan merasa tenteram kepadanya, dan dijadikan-Nya diantaramu rasa kasih dan sayang. Sesungguhnya pada yang demikian itu benar-benar terdapat tanda-tanda bagi kaum yang berfikir.”*

Surat Ar-Ruum ayat 21 menjelaskan bahwa ada tiga indikator terkait rumah tangga sejahtera menurut Islam, yaitu *Sakinnah*, *Mawaddah*, *Rahmah*. Imam Fachruddin Ar Razi dalam tafsir *Mafatihul Ghaib* menjelaskan bahwa *sakinah* adalah rasa tenang dan tentramnya hati yang dirasakan dan didapatkan dari suami, istri serta anggota keluarga yang lain dan berlaku. Imam Fachruddin Ar Razi menjelaskan bahwa *mawaddah* sebagai rasa cinta kasih yang tumpah untuk pasangan. *Rahmah* adalah rasa kasih sayang yang mengalir dari suami, istri dan anggota keluarga lainnya.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan klasifikasi status kesejahteraan rumah tangga, didapatkan kesimpulan sebagai berikut:

1. Tingkat akurasi ketepatan klasifikasi status kesejahteraan rumah tangga di Malang menggunakan Algoritma C5.0 dan metode *Synthetic Minority Oversampling Technique* (SMOTE) yaitu nilai *accuracy* sebesar 0,7632 atau 76,32% dengan nilai *sensitivity* sebesar 0,7975 atau 79,75% dan nilai *specificity* sebesar 0,7260 atau 72,6%. Tingkat ketepatan klasifikasi juga diukur dengan metode AUC-ROC, nilai AUC yaitu sebesar 0,7617. Nilai AUC masuk dalam kategori *fair classification*.
2. Tingkat kepentingan variabel tertinggi yaitu pada variabel pendidikan dengan tingkat kepentingan 32,21 persen, kemudian variabel status pekerjaan dengan persentase 26,17 persen. Selanjutnya variabel banyak anggota rumah tangga dengan persentase 25,64 persen, yang ke-empat ada variabel jenis kelamin dengan persentase 6,71 persen. Kemudian variabel umur 4,97 persen dan terakhir variabel lapangan pekerjaan dengan persentase sebesar 4,26 persen.

#### 5.2 Saran

Saran bagi peneliti selanjutnya yang menggunakan metode yang sama dapat memberikan perbedaan hasil akurasi sebelum SMOTE dan sesudah SMOTE. Selain itu, peneliti selanjutnya dapat mencoba menggunakan metode oversampling lain untuk menangani data tidak seimbang yang

mungkin dapat menghasilkan tingkat akurasi yang lebih baik. Peneliti selanjutnya juga dapat membandingkan algoritma C5.0 dengan metode *decission tree* yang lain sehingga akan didapatkan metode terbaik untuk klasifikasi.

## DAFTAR PUSTAKA

- Al-Quran Yasmina. 2009. Departemen Agama Republik Indonesia. Bandung: Sygma Exagrafika.
- Aradea, Satriyo, dkk. (2011). Penerapan Decision Tree untuk Penentuan Pola Data Penerimaan Mahasiswa Baru. *Jurnal Penelitian Sitrotika*, 7 (1), ISSN: 1693-9670.
- Azmatul B., R., Dina S., I., M. Afendi, Farit. (2013). Penerapan Synthetic Minority Oversampling Technique (SMOTE) terhadap data tidak seimbang pada Pembuatan Model Komposisi Jamu. *Xplore*. 1 (1), 1-6.
- Badan Pusat Statistik. (2021). *Pengeluaran Untuk Konsumsi Penduduk Indonesia (Berdasarkan Hasil Susenas September 2020)*. Katalog BPS: 3201004
- Badan Pusat Statistik. *Pengertian Kesejahteraan dan Kemiskinan*. Retrieved from Badan Pusat Statistik: <https://www.bps.go.id>.
- Badan Pusat Statistik. *Survei Sosial dan Ekonomi Nasional*. Retrieved from Badan Pusat Statistik: <http://www.sirusa.bps.go.id>
- Bisri, Achmad, and Romi S. Wahono. (2015). Penerapan Adaboost untuk Penyelesaian Ketidakeimbangan Kelas pada Penentuan Kelulusan Mahasiswa dengan Metode Decision Tree. *Journal of Intelligent Systems*. 1 (1), 27-32.
- Bobbit, Z. (2020, November 23). *How to perform Bagging in R (step-by-step)*. Retrieved Desember 23, 2023, from Statistics Simplified Statology: <https://www.statology.org/bagging-in-r/>
- Dwiyanti, Y., dkk. (2017). Memprediksi Status Berlangganan Klien Bank pada Kampanye Pemasaran Langsung dengan Menggunakan Metode Klasifikasi dengan Algoritma C5.0. *e-Proceeding of Engineering*, 4 (1), 3138-3147.
- Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Verlag Berlin Heidelberg: Springer.
- Hadianto, N., Novitasari, H., & Rahmawati, A. (2019). KLASIFIKASI PEMINJAMAN NASABAH BANK MENGGUNAKAN METODE NEURAL NETWORK. *Jurnal Pilar Nusa Mandiri*, 15(2), 163-170.
- Kastawan, Putu Wirya, Dewa Made, dan I Made Sudarma. (2018). Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil. *Majalah Ilmiah Teknologi Elektro*, Vol. 17, No. 3.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta : Penerbit Andi.

- Musu, W., Ibrahim, A., & Heriadi. (2021). Prngaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5. *Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi. X* (1), 186-195.
- Nair, A., Kuban, B. D., Obuchowsk, N., & Vince, D. G., (2001). Assesing Sprectal Algorithm To Predict Atherosclerotic Plaque Composition with Normalized and Raw Intravascular Ultrasound Data. *Ultrasound in Medicine and Biology*, 27 (10), 1319-1331.
- Permatasari, V., & Yuliana, L. (2021). Penerapan Regresi Logistik Biner pada Status Kesejahteraan Rumah Tangga di Provinsi Bali Tahun 2020. *Seminar Nasional Official Statistics, 2021*(1), 497-506.
- Pratiwi, F. E. dan Zain, I. (2014). Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara. *Jurnal Sains dan Seni Pomits*, 3(1), 2337-3520.
- Pratiwi, Dinda A., dkk. (2020). Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Tenik Informatika Menggunakan Metode Naïve Bayes. Bandung: Kreatif Industri Nusantara.
- Putri, Y. R., Mukhlash, I. dan Hidayat, N. (2013). Prediksi Pola Kecelakaan Kerja pada Perusahaan Non Ekstraktif Menggunakan Algoritma Decision Tree: C4.5 dan C5.0. *Jurnal Sains dan Seni Pomits*, 2(1), 2337-3520.
- Qiong, Gu, dkk. (2016). An Improved SMOTE Algorithm Based On Genetic Algorithm For Imbalanced Data Classification. *Journal of Digital Information Management*, 14 (2), 92-103.
- Sari, Veronica Retno, Feranandah, Firdausi & Yufis Azhar. (2020). Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naïve Bayes. *Jurnal Pendidikan Informatika*, 4 (2), 1-9.
- Siringoringo, R. (2018). Klasifikasi Data tidak seimbang menggunakan Algoritma SMOTE dan K-Nearest Neighbor. *Jurnal ISD*. 3 (1).
- Sofyan, S. & Prasetyo, A. (2021). Penerapan Synthetic Minority Oversampling Technique (SMOTE) terhadap data tidak seimbang pada Tingkat Pendapatan Pekerja Informal di Provinsi D.I. Yogyakarta Tahun 2019. *Seminar Nasional Official Statistics 2021*.
- Tafsir Alquran. (2021, September). *Tafsir Surah Al-Qasas Ayat 77*. Retrieved from Tafsir Alquran: <https://tafsiralquran.id/>
- Tafsir Web. (2018, September). *Tafsir Web*. Retrieved from Tafsir Web: <https://tafsirweb.com/>

Wahyu U., T. & Arianti, I. (2020). Principal Component Analysis Support Vektor Machine (PCA-SVM) untuk Klasifikasi Kesejahteraan Rumah Tangga di Kabupaten Brebes. *Proceeding SENDIU 2020*. ISBN : 978-979-3649-72-6.

Yusuf A., M., Charis F., A. & Maarif M., H. (2022). Perbandingan Kinerja Sistem Klasifikasi berbasis K-Fold Cross Validation pada Algoritma Decision Tree ID3 dan C5.0. *Jurnal Information Communication & Technology*. 21 (1), 44-52.

## LAMPIRAN

**Lampiran 1 : Data Penelitian SUSENAS Kota Malang tahun 2020**

No.	Keterangan Miskin	Jenis Kelamin	Umur	Banyak Art	Lapangan Pekerjaan	Pendidikan	Status Pekerjaan
1	Tidak Miskin	Lk	>49	>4	jasa	SMP	Karyawan
2	Tidak Miskin	Lk	<49	<=4	industri	SD	Karyawan
3	Tidak Miskin	Lk	>49	<=4	industri	SMA	Karyawan
4	Tidak Miskin	Lk	<49	<=4	jasa	SMP	Wirausaha
5	Tidak Miskin	Lk	<49	<=4	jasa	PT	Karyawan
6	Tidak Miskin	Lk	>49	>4	industri	SMA	Wirausaha
7	Tidak Miskin	Lk	>49	<=4	industri	SD	Karyawan
8	Tidak Miskin	Lk	>49	>4	jasa	PT	Karyawan
9	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
10	Tidak Miskin	Pr	<49	<=4	pertanian	SD	Wirausaha
11	Tidak Miskin	Lk	>49	<=4	industri	SMA	Karyawan
12	Miskin	Lk	>49	>4	industri	SMA	Karyawan
13	Tidak Miskin	Lk	49	>4	industri	PT	Karyawan
14	Tidak Miskin	Lk	>49	>4	jasa	PT	Wirausaha
15	Tidak Miskin	Lk	<49	<=4	jasa	SD	Karyawan
16	Tidak Miskin	Pr	<49	<=4	jasa	SD	Wirausaha
17	Tidak Miskin	Lk	<49	>4	jasa	SMA	Karyawan
18	Tidak Miskin	Lk	>49	<=4	industri	SD	Karyawan
19	Tidak Miskin	Lk	>49	>4	industri	SD	Karyawan
20	Tidak Miskin	Lk	<49	<=4	industri	SD	Wirausaha
21	Tidak Miskin	Lk	>49	<=4	industri	SD	Wirausaha
22	Tidak Miskin	Pr	>49	<=4	industri	SD	Karyawan
23	Tidak Miskin	Lk	>49	<=4	jasa	SD	Wirausaha
24	Tidak Miskin	Lk	<49	<=4	jasa	SD	Karyawan
25	Tidak Miskin	Lk	<49	>4	industri	SMP	Karyawan



<b>No.</b>	<b>Keterangan Miskin</b>	<b>Jenis Kelamin</b>	<b>Umur</b>	<b>Banyak Art</b>	<b>Lapangan Pekerjaan</b>	<b>Pendidikan</b>	<b>Status Pekerjaan</b>
26	Tidak Miskin	Lk	49	<=4	pertanian	SMA	Karyawan
:							
:							
40	Tidak Miskin	Lk	<49	<=4	pertanian	SMP	Wirausaha
41	Miskin	Lk	>49	>4	jasa	SD	Karyawan
42	Tidak Miskin	Lk	<49	<=4	jasa	SMA	Karyawan
43	Miskin	Lk	<49	<=4	jasa	SMA	Karyawan
44	Miskin	Lk	<49	<=4	jasa	SD	Karyawan
45	Tidak Miskin	Lk	>49	<=4	industri	SD	Wirausaha
46	Tidak Miskin	Lk	<49	<=4	jasa	SD	Wirausaha
47	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
48	Tidak Miskin	Lk	<49	<=4	industri	SD	Wirausaha
49	Tidak Miskin	Lk	>49	<=4	industri	SD	Karyawan
50	Miskin	Lk	<49	>4	jasa	SD	Karyawan
51	Tidak Miskin	Lk	<49	<=4	industri	SMP	Karyawan
52	Tidak Miskin	Lk	>49	<=4	industri	SD	Karyawan
53	Tidak Miskin	Lk	<49	<=4	jasa	SD	Karyawan
54	Tidak Miskin	Lk	<49	<=4	industri	SMA	Karyawan
55	Tidak Miskin	Lk	<49	<=4	industri	SMP	Karyawan
56	Tidak Miskin	Lk	<49	<=4	jasa	SD	Karyawan
57	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
58	Tidak Miskin	Lk	>49	<=4	industri	SD	Wirausaha
59	Tidak Miskin	Lk	>49	>4	jasa	SMA	Wirausaha
60	Tidak Miskin	Lk	<49	<=4	industri	SD	Wirausaha
61	Miskin	Lk	<49	<=4	jasa	SD	Wirausaha
62	Miskin	Lk	<49	>4	jasa	SD	Karyawan
63	Tidak Miskin	Lk	>49	<=4	jasa	SD	Wirausaha

No.	Keterangan Miskin	Jenis Kelamin	Umur	Banyak Art	Lapangan Pekerjaan	Pendidikan	Status Pekerjaan
64	Tidak Miskin	Lk	<49	<=4	industri	SD	Karyawan
65	Miskin	Pr	<49	<=4	jasa	SD	Karyawan
66	Tidak Miskin	Lk	<49	>4	industri	SD	Wirausaha
67	Tidak Miskin	Pr	49	<=4	industri	SMA	Karyawan
68	Tidak Miskin	Lk	>49	>4	jasa	PT	Wirausaha
69	Tidak Miskin	Lk	<49	<=4	industri	PT	Wirausaha
70	Tidak Miskin	Lk	>49	<=4	industri	PT	Wirausaha
71	Tidak Miskin	Lk	>49	<=4	industri	SMP	Karyawan
72	Tidak Miskin	Lk	<49	<=4	industri	SMP	Karyawan
73	Tidak Miskin	Pr	<49	>4	jasa	SD	Karyawan
74	Tidak Miskin	Lk	>49	<=4	industri	SMA	Wirausaha
75	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
76	Tidak Miskin	Lk	49	<=4	industri	SMA	Wirausaha
:							
:							
425	Tidak Miskin	Lk	<49	<=4	industri	SMP	Wirausaha
426	Tidak Miskin	Lk	<49	<=4	industri	SMP	Wirausaha
427	Tidak Miskin	Lk	<49	<=4	jasa	SD	Karyawan
428	Tidak Miskin	Lk	49	<=4	jasa	SMP	Karyawan
429	Tidak Miskin	Pr	<49	<=4	industri	SD	Wirausaha
430	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
431	Tidak Miskin	Pr	<49	>4	jasa	SD	Karyawan
432	Miskin	Lk	>49	<=4	industri	SMP	Karyawan
433	Tidak Miskin	Lk	>49	<=4	industri	SMA	Wirausaha
434	Tidak Miskin	Lk	>49	>4	industri	SD	Wirausaha
435	Tidak Miskin	Lk	>49	<=4	jasa	PT	Wirausaha
436	Miskin	Lk	>49	<=4	industri	SMA	Karyawan

<b>No.</b>	<b>Keterangan Miskin</b>	<b>Jenis Kelamin</b>	<b>Umur</b>	<b>Banyak Art</b>	<b>Lapangan Pekerjaan</b>	<b>Pendidikan</b>	<b>Status Pekerjaan</b>
437	Tidak Miskin	Lk	<49	<=4	jasa	SD	Karyawan
438	Miskin	Lk	>49	>4	jasa	SMP	Wirausaha
439	Tidak Miskin	Pr	<49	<=4	Industri	SMP	Wirausaha
440	Miskin	Pr	<49	>4	jasa	SD	Karyawan
:							
:							
755	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
756	Tidak Miskin	Lk	>49	<=4	industri	PT	Karyawan
757	Tidak Miskin	Lk	>49	<=4	industri	SMA	Karyawan
758	Tidak Miskin	Lk	<49	>4	industri	SD	Wirausaha
759	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
760	Tidak Miskin	Lk	<49	<=4	pertanian	SD	Wirausaha
761	Tidak Miskin	Lk	>49	<=4	industri	SMP	Wirausaha
762	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
763	Tidak Miskin	Lk	<49	<=4	jasa	SMA	Wirausaha
764	Tidak Miskin	Lk	>49	<=4	jasa	SMP	Karyawan
765	Tidak Miskin	Lk	>49	<=4	pertanian	SMA	Karyawan
766	Tidak Miskin	Pr	>49	<=4	jasa	PT	Karyawan
767	Tidak Miskin	Pr	<49	<=4	jasa	SD	Karyawan
768	Tidak Miskin	Pr	<49	<=4	jasa	SMA	Karyawan
769	Tidak Miskin	Pr	>49	<=4	jasa	PT	Karyawan
770	Tidak Miskin	Pr	>49	<=4	jasa	PT	Karyawan
771	Tidak Miskin	Lk	<49	>4	jasa	SMP	Karyawan
772	Tidak Miskin	Lk	>49	<=4	jasa	SMA	Wirausaha
773	Tidak Miskin	Pr	>49	<=4	jasa	PT	Karyawan
774	Tidak Miskin	Lk	>49	<=4	industri	PT	Wirausaha
775	Tidak Miskin	Lk	>49	<=4	jasa	PT	Karyawan

## Lampiran 2 : Script dan Output

```
dataku <- read.csv("dummyData1.csv", sep=";")

#SMOTE
library("DMwR")
dataku$keterangan.Miskin<-as.factor(dataku$keterangan.Miskin)
dataku$jenis.kelamin<-as.factor(dataku$jenis.kelamin)
dataku$status.pekerjaan<-as.factor(dataku$status.pekerjaan)
dataku$umur<-as.factor(dataku$umur)
dataku$banyak.art<-as.factor(dataku$banyak.art)
dataku$lapangan.pekerjaan<-as.factor(dataku$lapangan.pekerjaan)
dataku$pendidikan<-as.factor(dataku$pendidikan)
str(dataku)
table(dataku$keterangan.Miskin)
prop.table(table(dataku$keterangan.Miskin))

target_variable <- "keterangan.Miskin"
predictor_variables <- setdiff(names(dataku), target_variable)
table(dataku$keterangan.Miskin)

data_balanced <- SMOTE(as.formula(paste(target_variable,"~", paste(predictor_variables, collapse = "+"))),
                      dataku, perc.over = 1450, perc.under = 110, k=5)
table(data_balanced$keterangan.Miskin)
prop.table(table(data_balanced$keterangan.Miskin))

> table(dataku$keterangan.Miskin)

  M  TM
50 725
> prop.table(table(dataku$keterangan.Miskin))

          M          TM
0.06451613 0.93548387

> table(data_balanced$keterangan.Miskin)

  M  TM
750 770
> prop.table(table(data_balanced$keterangan.Miskin))

          M          TM
0.4934211 0.5065789
~
~

#Membuat data training dan testing
n <- round(nrow(data_balanced)*0.90);n
set.seed(99191)
samp=sample(1:nrow(data_balanced),n)
data.train = data_balanced[samp,]
dim(data.train)
data.test = data_balanced[-samp,]
dim(data.test)
str(data.test$keterangan.Miskin)

library("c50")
modelku <- C5.0(keterangan.Miskin ~., data=data.train)
modelku
summary(modelku)
plot(modelku)
```

```

> modelku <- C5.0(Keterangan.Miskin ~ ., data=data.train)
> modelku

Call:
C5.0.formula(formula = Keterangan.Miskin ~ ., data = data.train)

Classification Tree
Number of samples: 1368
Number of predictors: 6

Tree size: 8

Non-standard options: attempt to group attributes

> summary(modelku)

Call:
C5.0.formula(formula = Keterangan.Miskin ~ ., data = data.train)

C5.0 [Release 2.07 GPL Edition]           Thu Dec 28 04:52:40 2023
-----

Class specified by attribute `outcome'

Read 1368 cases (7 attributes) from undefined.data

Decision tree:

Pendidikan = 4: TM (255)
Pendidikan in {1,2,3}:
...status.pekerjaan = 3: TM (24)
  status.pekerjaan in {1,2,4,5,6}:
  ...banyak.art = 2: M (458/96)
    banyak.art = 1:
    ...Pendidikan in {2,3}: TM (346/138)
      Pendidikan = 1:
      ...jenis.kelamin = 2: M (74/18)
        jenis.kelamin = 1:
        ...Umur = 2: M (30/6)
          Umur = 1:
          ...lapangan.pekerjaan in {1,3}: M (112/42)
            lapangan.pekerjaan = 2: TM (69/21)

```

Evaluation on training data (1368 cases):

```
      Decision Tree
-----
Size      Errors
      8  321(23.5%)  <<

      (a)  (b)  <-classified as
-----  -----
      512  159  (a): class M
      162  535  (b): class TM
```

Attribute usage:

```
100.00% Pendidikan
 81.36% status.pekerjaan
 79.61% banyak.art
 20.83% jenis.kelamin
 15.42% Umur
 13.23% lapangan.pekerjaan
```

```
# Melakukan prediksi pada data pengujian
predictions <- predict(modelku, data.test)

# Menghitung matriks kebingungan
confusion_matrix <- table(Actual=data.test$keterangan.Miskin, Predicted=predictions)
print(confusion_matrix)

class_predict <- predict(modelku, newdata = data.test, type = "class")
class_predict

#konversi variabel target ke format biner
binary_truth <- ifelse(data.test$keterangan.Miskin == "M", 1, 0)
binary_prediction <- ifelse(class_predict == "M", 1, 0)
length(binary_prediction)
length(binary_truth)

library("pROC")
#Hitung ROC
roc_curve <- roc(binary_truth, binary_prediction)
roc_curve

#Hitung AUC
auc_value <- auc(roc_curve)
auc_value

#Gambar kurva ROC
plot(roc_curve, main = "ROC Curve", col = "blue", lwd = 2)
abline(a=0, b=1, col = "red", lwd = 2, lty = 2)
text(0.7260, 0.7975, "(0.7260,0.7975)", col="black", cex=1.1)
|
```

```

> roc_curve

Call:
roc.default(response = binary_truth, predictor = binary_prediction)

Data: binary_prediction in 73 controls (binary_truth 0) < 79 cases (binary_truth 1).
Area under the curve: 0.7617

> auc_value
Area under the curve: 0.7617

#menghitung variabel penting menggunakan caret
vp <- varImp(modelku, scale = FALSE)
print(vp)
str(vp)

> print(vp)

```

	Overall
Pendidikan	100.00
status.pekerjaan	81.36
banyak.art	79.61
jenis.kelamin	20.83
Umur	15.42
lapangan.pekerjaan	13.23

## RIWAYAT HIDUP



Aida Nabila Yumna Hanin, lahir di Malang pada 19 Februari 1999. Akrab dengan sapaan Nabila. Anak sulung dari tiga bersaudara dari pasangan bapak M. Muhajir dan Ibu Dewi Mahlikah. Adik pertama bernama Shinta Nafisa dan adik kedua bernama Vania Zabrina Ramadhani.

Adapun riwayat pendidikannya diawali di TK Islam Sunan Giri Malang pada tahun 2004 sampai 2005. Pendidikan dasarnya ditempuh di MI KH.

Hasyim Asy'ari Blimbing Malang lulus pada tahun 2011. Setelah itu melanjutkan sekolah menengah di SMP Al-Rifa'ie Gondanglegi Malang lulus pada tahun 2014 dan SMA Al-Rifa'ie Gondanglegi Malang lulus pada tahun 2017. Pada tahun yang sama melanjutkan pendidikannya di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada Jurusan Matematika melalui jalur SNMPTN.

Selama menjadi mahasiswa, berperan aktif di organisasi kampus baik intra maupun ekstra dalam rangka mengembangkan potensi atau *softskill* dan membangun jaringan. Nabila pernah menjadi Wakil Bendahara Umum tahun 2018 dan Bendahara Umum tahun 2019 di HMJ "Integral" Matematika. Selain itu menjadi Wakil Gubernur DEMA Fakultas Sains dan Teknologi tahun 2020. Sekaligus menjadi kader Pergerakan Mahasiswa Islam Indonesia (PMII) Rayon "Pencerahan" Galileo pada tahun 2018 dan Sekretaris KOPRI PMII Rayon "Pencerahan" Galileo pada tahun 2019. Penulis dapat dihubungi melalui email: [aidanabilahanin@gmail.com](mailto:aidanabilahanin@gmail.com)





BUKTI KONSULTASI SKRIPSI

Nama : Aida Nabila Yumna Hanin  
NIM : 17610028  
Fakultas/ Program Studi : Sains dan Teknologi/ Matematika  
Judul Skripsi : Penerapan Algoritma C5.0 dan Metode SMOTE pada Klasifikasi Status Kesejahteraan Rumah Tangga  
Pembimbing I : Dr. Fachrur Rozi, M.Si.  
Pembimbing II : Erna Herawati, M.Pd.

No	Tanggal	Hal	Tanda Tangan	
1	30 Januari 2023	Konsultasi Bab I, II, dan III	1.	
2	31 Januari 2023	Konsultasi Kajian Islam		2.
3	13 Februari 2023	Revisi Bab I dan II	3.	
4	14 Februari 2023	Revisi Kajian Islam		4.
5	02 Mei 2023	Revisi Bab III	5.	
6	15 Mei 2023	ACC Seminar Proposal		6.
7	01 September 2023	Konsultasi Revisi Proposal	7.	
8	08 September 2023	Konsultasi Kajian Islam Bab IV		8.
9	10 Oktober 2023	Konsultasi Bab IV	9.	
10	15 November 2023	Revisi Bab IV dan Konsultasi Bab V		10.
11	20 November 2023	Revisi Kajian Islam Bab IV	11.	

No	Tanggal	Hal	Tanda Tangan
12	29 November 2023	Revisi Bab V dan Format Penulisan	12. <i>[Signature]</i>
13	04 Desember 2023	Revisi Kajian Islam Bab IV	13. <i>[Signature]</i>
14	12 Desember 2023	ACC Seminar Hasil	14. <i>[Signature]</i>
15	12 Desember 2023	ACC Seminar Hasil Kajian Islam	15. <i>[Signature]</i>
16	22 Desember 2023	Revisi Seminar Hasil dan ACC Sidang Skripsi	16. <i>[Signature]</i>
17	30 Desember 2023	ACC Naskah Skripsi Keseluruhan	17. <i>[Signature]</i>



Mataram, 30 Desember 2023

Mengetahui,  
Ketua Program Studi Matematika

*[Signature]*  
Dr. Elly Susanti, M.Sc

NIP. 19741129 200012 2 005