

KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN MENGGUNAKAN METODE *RANDOM FOREST* PADA KELURAHAN 13 ULU SEBERANG ULU DUA PALEMBANG

Agung Nasrullah^{1*}, Muhammad Haviz Irfani², Lastri Widya Astuti³, Zaid Romegar Mair⁴
^{1,2,3,4} Fakultas Ilmu Komputer dan Sains, Universitas Indo Global Mandiri, Palembang
e-mail: ¹2020110055@students.uigm.ac.id, ²m.haviz@uigm.ac.id, ³lastriwidya@uigm.ac.id,
⁴zaidromegar@uigm.ac.id

ABSTRACT

The Family Hope Program (Program Keluarga Harapan, PKH) is a social assistance program aimed at helping poor and vulnerable families. This program is provided to families registered in the Unified Database for Social Welfare (DTKS) and meet certain criteria, such as having school-aged children, toddlers, pregnant women, or elderly members. In its implementation, PKH distribution often faces challenges, such as inaccurate targeting and insufficiently detailed data on poor families. Therefore, this research aims to classify PKH beneficiaries in 13 Ulu Village, Seberang Ulu II District, Palembang, to ensure that social assistance is given to those who truly meet the eligibility criteria. The method used in this research is the Random Forest method, which involves several stages. These stages include Data Preprocessing, Data Splitting, Random Forest Method Implementation, and Evaluation. Several attributes used in this research include House Ownership, Number of Children, School-Aged Children, Toddlers, Pregnant Women, Elderly, Occupation, and Income, with the target attribute being PKH eligibility. In this research, 117 datasets were used, which had gone through the Preprocessing stage. The data was then split with a 70:30 ratio for training and testing, resulting in an accuracy of 97%, precision of 96%, recall of 100%, and an F1-Score of 98%. These results indicate that the Random Forest method is quite effective in accurately classifying PKH beneficiaries.

Keywords: Classify, Family Hope Program, Random Forest

INTISARI

Program Keluarga Harapan (PKH) merupakan salah satu program bantuan sosial yang ditujukan untuk membantu keluarga miskin dan rentan miskin. Program ini diberikan kepada keluarga yang terdaftar dalam Data Terpadu Kesejahteraan Sosial (DTKS) dan memenuhi persyaratan tertentu, seperti memiliki anak usia sekolah, balita, ibu hamil, atau lansia. Dalam pelaksanaannya, penyaluran PKH sering kali menghadapi masalah, seperti ketidaktepatan sasaran dan pendataan keluarga miskin yang kurang akurat. Oleh karena itu, penelitian ini bertujuan untuk melakukan klasifikasi penerima PKH di Kelurahan 13 Ulu, Kecamatan Seberang Ulu II, Palembang, guna memastikan bahwa bantuan sosial ini dapat diberikan kepada mereka yang benar-benar memenuhi kriteria. Metode yang digunakan dalam penelitian ini adalah metode *Random Forest*, yang melibatkan beberapa tahapan. Tahapan tersebut meliputi *Preprocessing Data*, *Split Data*, Implementasi Metode *Random Forest*, dan Evaluasi. Beberapa atribut yang digunakan dalam penelitian ini antara lain Kepemilikan Rumah, Jumlah Anak, Anak Sekolah, Anak Balita, Ibu Hamil, Lansia, Pekerjaan, dan Penghasilan, dengan target atribut berupa Kelayakan penerima PKH. Dalam penelitian ini, digunakan 117 *dataset* yang telah melalui proses *Preprocessing*. Data tersebut kemudian dibagi dengan rasio 70:30 untuk pelatihan dan pengujian, menghasilkan akurasi sebesar 97%, *precision* sebesar 96%, *recall* sebesar 100%, dan *F1-Score* sebesar 98%. Hasil ini menunjukkan bahwa metode *Random Forest* cukup efektif dalam mengklasifikasikan penerima PKH secara tepat.

Kata kunci: Klasifikasi, Program Keluarga Harapan, *Random Forest*

1. PENDAHULUAN

Klasifikasi memiliki peran penting dalam menentukan penerima Program Keluarga Harapan (PKH) di Kelurahan 13 Ulu, Kecamatan Seberang Ulu II, Palembang. Metode yang bisa dipakai untuk melakukan klasifikasi adalah metode *Random Forest*, yang merupakan algoritma *Machine Learning* untuk membangun model keputusan dalam

klasifikasi. Metode *Random Forest* terdiri dari kumpulan pohon keputusan, di mana setiap pohon digunakan untuk mengklasifikasikan data ke dalam kelas tertentu (Suci Amaliah et al., 2022).

Kelurahan 13 Ulu berada di Jln. KH Azhari, Kecamatan Seberang Ulu II, Kota Palembang. Kelurahan ini memiliki luas sekitar 788.380,34 m² yang terdiri dari 34 Rukun Tetangga (RT) dan 6 Rukun Warga (RW). Berdasarkan data, pada tahun 2023, Kelurahan 13 Ulu memiliki jumlah penduduk sekitar 6.448 jiwa. Kelurahan ini merupakan salah satu kelurahan di Kecamatan Seberang Ulu II, Kota Palembang yang menerima bantuan dari pemerintah melalui PKH.

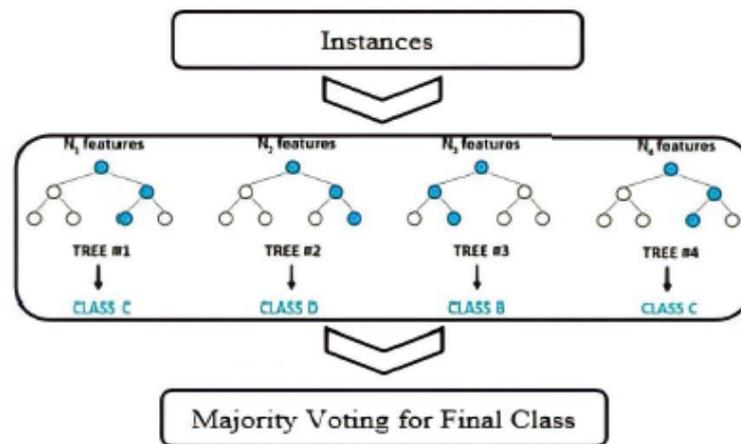
Program Keluarga Harapan (PKH) adalah program bantuan sosial pemerintah yang ditujukan untuk keluarga kurang mampu yang telah terdaftar di Data Terpadu Kesejahteraan Sosial (DTKS) dan memenuhi kriteria yang ditentukan oleh pemerintah. Program Keluarga Harapan (PKH) bertujuan untuk mengurangi kesenjangan sosial dan meningkatkan kesejahteraan di berbagai wilayah di Indonesia (Direktorat Jaminan Sosial & Kementerian Sosial RI, 2021).

Klasifikasi merupakan proses pembelajaran untuk memetakan kumpulan data ke label kelas yang sudah diketahui sebelumnya. Klasifikasi termasuk supervised learning dengan data latih sebelumnya (Wanto et al., 2020). Klasifikasi menemukan model atau fungsi yang menjelaskan dan membedakan konsep atau kelas data. Model ini digunakan untuk memprediksi label kelas objek yang tidak diketahui setelah melakukan analisis set data latih dengan label kelas yang diketahui (Wanto et al., 2020).

Random Forest yaitu algoritma yang awalnya diperkenalkan dalam penelitian dan terbukti unggul dalam menangani masalah regresi dan klasifikasi. Algoritma ini merupakan pengembangan dari *decision tree* yang dirancang untuk mengatasi masalah *overfitting* yang sering terjadi saat menggunakan *decision tree* (Ferdita Nugraha et al., 2022).

Random Forest termasuk metode *ensemble learning* menyatukan sejumlah pohon keputusan untuk akurasi dan stabilitas prediksi (Sholihah & Hermawan, 2023). Dalam kasus klasifikasi, *Random Forest* menggabungkan hasil prediksi dari banyak pohon keputusan dengan menggunakan teknik *voting*. Setiap pohon akan memberikan prediksi kelas untuk suatu sampel data. Kemudian, prediksi kelas terbanyak dari seluruh pohon akan dipilih sebagai keputusan akhir klasifikasi untuk sampel data tersebut (Sholihah & Hermawan, 2023).

Random Forest menggunakan kombinasi pohon keputusan digunakan digunakan model dasar untuk mengklasifikasikan dan memprediksi nilai data, yang membuatnya termasuk dalam kategori pembelajaran kelompok (Ferdita Nugraha et al., 2022). Berikut Gambaran pembentukan *Random Forest*.



Gambar 1. Ilustrasi Proses dari *Random Forest* (Ferdita Nugraha et al., 2022)

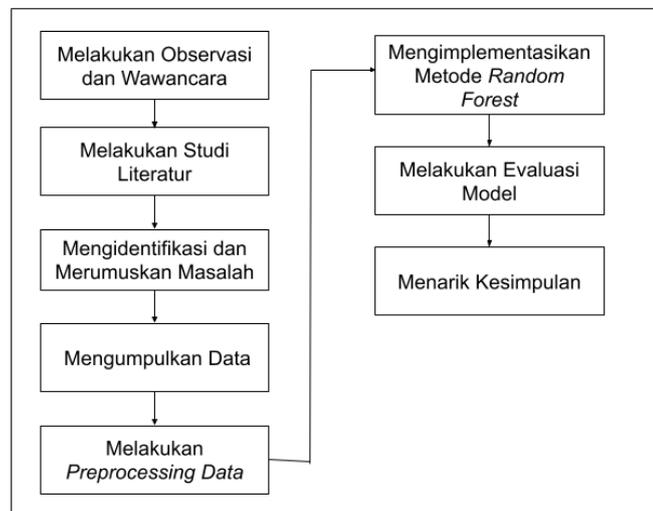
Penelitian tentang klasifikasi penerima bantuan menggunakan metode *Random Forest* bukanlah sesuatu yang baru. Penelitian yang berjudul “Perbandingan Tingkat Akurasi Algoritma *Decision Tree* Dan *Random Forest* Dalam Mengklasifikasi Penerima Bantuan Sosial BPNT Di Desa Slangit”, penulis menyatakan bahwa perbandingan tingkat akurasi pada klasifikasi penerima BPNT menggunakan metode *Random Forest* dan *Decision Tree* menghasilkan akurasi sebesar 97% (Aldiyansyah et al., 2024). Penelitian selanjutnya tentang klasifikasi

penerima bantuan menggunakan metode *Random Forest* yang berjudul “Implementasi Algoritma *Random Forest* untuk Klasifikasi Penerima Bantuan Bedah Rumah”, penulis dalam penelitian ini menjelaskan bahwa klasifikasi bantuan bedah rumah menggunakan metode *Random Forest* menghasilkan akurasi 93% (Maulana, 2023). Penelitian lainnya yang berjudul “Implementasi Algoritma *Random Forest* Untuk Menentukan Penerima Bantuan Raskin”, penulis dalam penelitian ini menjelaskan bahwa hasil menentukan penerima bantuan Raskin menggunakan metode *Random Forest* menghasilkan akurasi sebesar 97% (Kurniawan et al., 2023).

Penelitian ini akan dilakukan pemanfaatan metode *Random Forest* untuk mengklasifikasikan penerima bantuan Program Keluarga Harapan (PKH). Hasil penelitian ini akan diperoleh tingkat akurasi metode *Random Forest* dalam memperoleh klasifikasi bagi penerima bantuan Program Keluarga Harapan (PKH). Harapan dengan adanya penelitian ini, dapat membantu meningkatkan efisiensi dan efektivitas dalam penentuan penerima bantuan sosial serta memberikan rekomendasi kebijakan yang lebih tepat sasaran.

2. METODE PENELITIAN

Metode penelitian adalah rencana yang dijadikan pedoman untuk penelitian ini. Berikut tahapan pada metode penelitian.



Gambar 2. Alur Tahapan Metode Penelitian

Gambar 2 memperlihatkan alur dari tahapan penelitian yang digunakan dalam penelitian ini antara lain:

- Melakukan Observasi dan Wawancara: Melakukan Observasi objek penelitian tentang memperoleh gambaran apa saja terkait dengan penerima Program Keluarga Harapan (PKH) Kelurahan 13 Ulu, terutama dalam penentuan penerimaan Program Keluarga Harapan (PKH), dan wawancara dengan pendamping Program Keluarga Harapan (PKH) di Kelurahan 13 Ulu, Kecamatan Seberang Ulu II, Palembang.
- Melakukan Studi Literatur: Pada tahapan studi literatur, memperoleh berbagai jurnal, *e-book*, dan artikel yang dapat mendukung penelitian, secara referensi digunakan juga informasi yang diperlukan penelitian. Pada tahapan studi literatur ini juga didapatkan penelitian terdahulu yang digunakan untuk memperkuat argumen, perbandingan, dan mempertajam penelitian yang sedang dilakukan.
- Mengidentifikasi dan Merumuskan masalah: Identifikasi masalah dalam penelitian ini yaitu ketidakmerataan distribusi bantuan dalam Program Keluarga Harapan (PKH). Program Keluarga Harapan (PKH) tidak selalu sampai kepada keluarga yang benar-benar memerlukan, dan banyak keluarga yang kurang mampu untuk belum mendapatkan manfaat dari bantuan ini. Selain itu, beberapa penerima bantuan sebenarnya sudah tidak lagi memenuhi syarat untuk menerima bantuan. Dari identifikasi masalah tersebut, dapat dirumuskan masalah yaitu bagaimana cara mengklasifikasikan penerima Program Keluarga Harapan (PKH) Kelurahan 13 Ulu, Kecamatan Seberang Ulu II, Palembang untuk memastikan bahwa bantuan diberikan kepada yang membutuhkan. Untuk menyelesaikan masalah ini, digunakan metode *Random Forest*.
- Mengumpulkan Data: Pada tahap ini data penerima Program Keluarga Harapan (PKH) Kelurahan 13 Ulu, Kecamatan Seberang Ulu II, Palembang diperoleh dari pendamping Program Keluarga Harapan (PKH). Data penelitian ini yaitu data sekunder. Data sekunder dalam penelitian ini berupa data penerima bantuan Program Keluarga Harapan (PKH) di Kelurahan 13 Ulu Kecamatan Seberang Ulu II, Palembang tahun 2023 yang berjumlah 140 data. Penelitian ini akan menggunakan 8 atribut (Tabel 1) yang terdiri dari beberapa atribut

meliputi, Kepemilikan Rumah, Jumlah Anak, Anak Sekolah, Anak Balita, Ibu Hamil, Lansia, Pekerjaan, Penghasilan, dan Target atau label Kelayakan.

Tabel 1. Atribut Penelitian

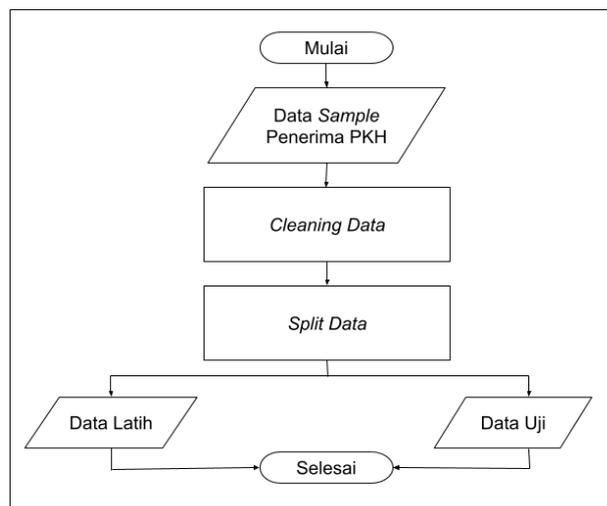
No.	Atribut	Tipe Data
1.	Kepemilikan Rumah	Kategorikal
2.	Jumlah Anak	Numerik
3.	Anak Sekolah	Nominal
4.	Anak Balita	Nominal
5.	Ibu Hamil	Nominal
6.	Lansia	Nominal
7.	Pekerjaan	Kategorikal
8.	Penghasilan	Numerik

Gambar 3 memperlihatkan 10 record data berdasarkan atribut (variabel) yang digunakan dalam *dataset* awal.

NO	NAMA	KEPEMILIKAN RUMAH	JUMLAH ANAK	ANAK SEKOLAH	BALITA	IBU HAMIL	LANSIA	PEKERJAAN	PENGHASILAN	KELAYAKAN
1	HERYATI	Milik sendiri	2	Tidak	Tidak	Tidak	Ya	Wiraswasta	2000000	Kurang layak
2	ZAINUR	Kontrak	4	Ya	Ya	Tidak	Tidak	Tidak bekerja	1500000	Layak
3	AYUDA	Rumah orang tua	1	Ya	Tidak	Tidak	Tidak	Tidak bekerja	700000	Layak
4	FITRIYANI	Milik sendiri	3	Ya	Ya	Tidak	Tidak	ART	600000	Layak
5	ENNI	Kontrak	3	Ya	Ya	Tidak	Tidak	ART	500000	Layak
6	MAIMUNAH	Kontrak	2	Tidak	Tidak	Tidak	Ya	Tidak bekerja	1200000	Layak
7	RISKA FITRIANA	Kontrak	0	Tidak	Tidak	Ya	Tidak	Karyawan	2000000	Layak
8	YASMI	Rumah orang tua	3	Tidak	Tidak	Tidak	Ya	Tidak bekerja	1000000	Layak
9	NOVITA SURYANI	Milik sendiri	2	Tidak	Tidak	Tidak	Tidak	Karyawan	3000000	Kurang layak
10	ONA ABDUL RONI	Milik sendiri	1	Ya	Tidak	Tidak	Tidak	Buruh	1400000	Kurang layak

Gambar 3. 10 Baris Pertama *Dataset* Awal

- e. Melakukan *Preprocessing Data*: *Preprocessing data* pada tahap penelitian yang menyiapkan data mentah agar dapat diolah. Data sekunder yang didapatkan dari pendamping Program Keluarga Harapan (PKH) sebelumnya masih berupa data mentah, sehingga harus dilakukan *Cleaning Data* terlebih dahulu mencakup penghapusan data duplikat, normalisasi data dalam atribut, dan pemeriksaan pada *dataset* yang bertujuan untuk mengetahui apakah terdapat data yang kosong (*null*) atau tidak, setelah itu *split data* dengan rasio 70:30 untuk pelatihan dan pengujian. Tahap *preprocessing data* sebagai berikut (Gambar 4).



Gambar 4. Tahap *Preprocessing Data*

- f. Mengimplementasikan Metode *Random Forest*: Menerapkan Metode *Random Forest* pada data yang telah di *preprocessing* dengan melakukan pengujian untuk membangun model klasifikasi.

Perhitungan nilai *Information Gain* dan nilai *Entropy* pada rumus berikut (Nalatissifa et al., 2021).

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

Keterangan :

S = Himpunan *Case* n = jumlah pembagian S_i terhadap S

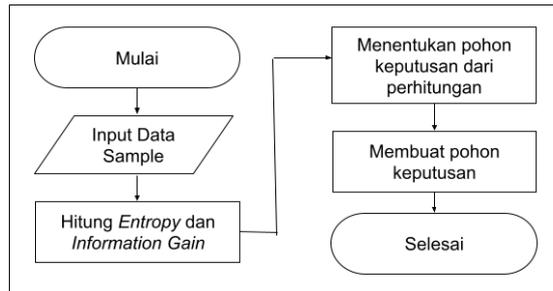
p_i = Bagian S_i terhadap S

$$Information\ Gain(S,A) = Entropy(S) - \sum_{i=1}^n -p_i \frac{|S_i|}{S} * Entropy(S_i) \dots\dots\dots (2)$$

Keterangan:

S = Himpunan *Case* A = Atribut n = jumlah pembagian atribut A

$|S_i|$ = Bagian S_i terhadap S $|S|$ = jumlah *case* dalam S



Gambar 5. Tahapan Implementasi *Random Forest*

- g. Mengevaluasi Model: Mengevaluasi model menggunakan *Confusion Matrix* untuk menghitung akurasi, *precision*, *recall*, dan *f1-score*.
- h. Menarik Kesimpulan: Menyimpulkan hasil penelitian berdasarkan analisis dan evaluasi model yang telah dilakukan.

3. HASIL DAN PEMBAHASAN

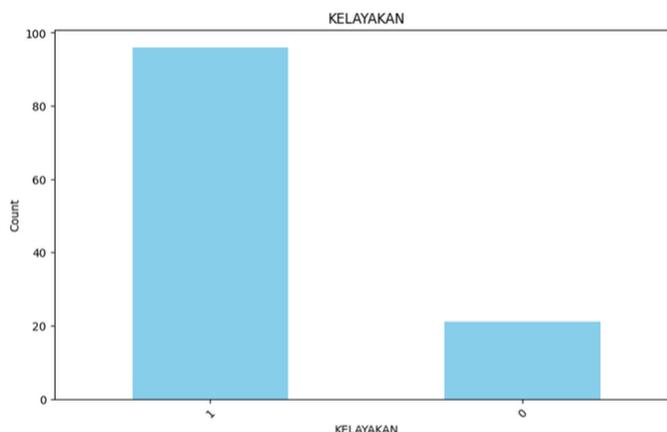
3.1. Data Cleaning

Data Cleaning merupakan *dataset* yang telah di *preprocessing* sebelumnya, yang mencakup penghapusan data duplikat, normalisasi data dalam atribut, dan pemeriksaan pada *dataset* yang bertujuan untuk mengetahui apakah terdapat data yang kosong (*null*). Berikut merupakan hasil dari jumlah data duplikat.

Jumlah data duplikat dalam dataset: 23

Gambar 6. Jumlah Data Duplikat

Dataset yang akan disiapkan berupa 117 dari 140 data penerima bantuan Program Keluarga Harapan (PKH) di Kelurahan 13 Ulu, Kecamatan Seberang Ulu II, Palembang karena terdapat 23 data duplikat dihapus, yang terdiri dari 96 ‘Layak’ dan 21 ‘Kurang Layak’ pada label ‘Kelayakan’.



Gambar 7. Jumlah Data Label Kelayakan

Berikut merupakan *dataset* setelah dinormalisasi.

KEPEMILIKAN RUMAH	JUMLAH ANAK	ANAK SEKOLAH	BALITA	IBU HAMIL	LANSIA	PEKERJAAN	PENGHASILAN	KELAYAKAN
0	2	0	0	0	1	0	2000000	0
1	4	1	1	0	0	1	1500000	1
1	1	1	0	0	0	1	700000	1
0	3	1	1	0	0	1	600000	1
1	3	1	1	0	0	1	500000	1
1	2	0	0	0	1	1	1200000	1
1	0	0	0	1	0	0	2000000	1
1	3	0	0	0	1	1	1000000	1
0	2	0	0	0	0	0	3000000	0
0	1	1	0	0	0	1	1400000	0

Gambar 8. *Dataset* Setelah Dinormalisasi

Berikut merupakan hasil dari pemeriksaan pada setiap atribut dan label, apakah terdapat data yang kosong (*null*).

Jumlah data yang kosong untuk setiap atribut dan label:	
KEPEMILIKAN RUMAH	0
JUMLAH ANAK	0
ANAK SEKOLAH	0
BALITA	0
IBU HAMIL	0
LANSIA	0
PEKERJAAN	0
PENGHASILAN	0
KELAYAKAN	0
dtype: int64	

Gambar 9. Jumlah Data Kosong (*null*)

3.2. Split Data

Sebelum melakukan *Split Data*, pada tahap *Split Data* dimulai dengan menentukan fitur atau atribut (*X*) dan target atau label (*y*) yang akan digunakan dalam *Random Forest*. Fitur atau atribut (*X*) terdiri dari Kepemilikan Rumah, Jumlah Anak, Anak Sekolah, Anak Balita, Ibu Hamil, Lansia, Pekerjaan, Penghasilan, dan target atau label (*y*) yaitu Kelayakan, *Split Data* yang dilakukan yaitu rasio 70:30. Berikut merupakan jumlah data pelatihan dan pengujian yang terdiri dari 81 data pelatihan dan 36 data pengujian dari 8 atribut dan 1 label.

$$((81, 8), (36, 8))$$

Gambar 10. Jumlah Data Pelatihan dan Pengujian

3.3 Pohon Keputusan

Mengawali pembuatan model *Random Forest* dilakukan *Sampling* data dan menentukan variabel *m* yang digunakan untuk membandingkan nilai *Information Gain* dari atribut yang digunakan. Variabel *m* dapat ditentukan secara acak tetapi jika telah digunakan pada node sebelumnya maka atribut tersebut tidak dapat digunakan kembali. Berikut ini perhitungan manual *Entropy* dan *Information Gain* untuk menentukan root dari pohon keputusan (Tabel 2).

Tabel 2. Perhitungan Root Pohon Keputusan

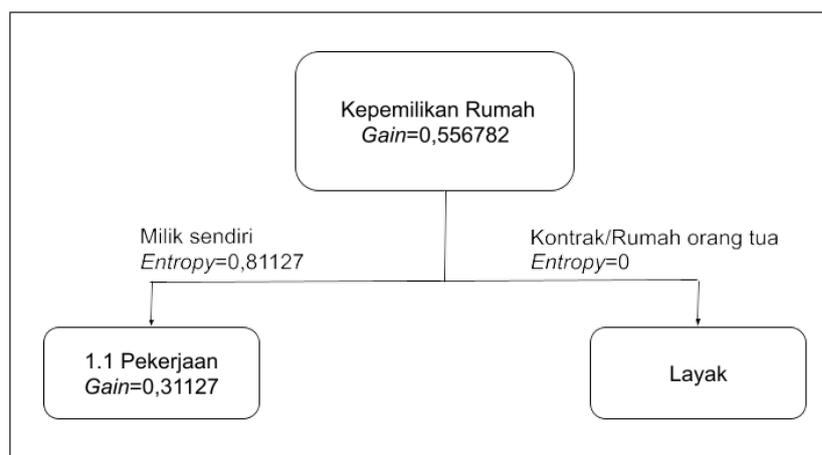
<i>Entropy (Total)</i> =	$= - \left[\frac{7}{10} \log_2 \frac{7}{10} + \frac{3}{10} \log_2 \frac{3}{10} \right] = 0,88129$
<i>Entropy Kepemilikan Rumah (Milik sendiri)</i> =	$= - \left[\frac{1}{4} \log_2 \frac{1}{4} + \frac{3}{4} \log_2 \frac{3}{4} \right] = 0,81127$

Entropy Kepemilikan Rumah (Kontrak /Rumah orang tua) =	$= - \left[\frac{6}{6} \log_2 \frac{6}{6} + \frac{0}{6} \log_2 \frac{0}{6} \right] = 0$
Entropy Penghasilan (Penghasilan > Rp. 1.000.000) =	$= - \left[\frac{4}{7} \log_2 \frac{4}{7} + \frac{3}{7} \log_2 \frac{3}{7} \right] = 0,98522$
Entropy Penghasilan (Penghasilan < Rp. 1.000.000) =	$= - \left[\frac{3}{3} \log_2 \frac{3}{3} + \frac{0}{3} \log_2 \frac{0}{3} \right] = 0$
Information Gain (Kepemilikan Rumah) =	$= 0,88129 - \left[\frac{4}{10} * 0,81127 + \frac{6}{10} * 0 \right] = 0,556782$
Information Gain (Penghasilan) =	$= 0,88129 - \left[\frac{7}{10} * 0,98522 + \frac{3}{10} * 0 \right] = 0,191636$

Pada *Root* Pohon keputusan dengan menggunakan atribut "Kepemilikan Rumah" sebagai *Root* berdasarkan nilai *information gain* sebesar 0,556782, yang membagi data menjadi dua cabang yaitu "Milik Sendiri" dengan *entropy* 0,81127 dan "Kontrak/Rumah orang tua" dengan *entropy* 0. Langkah selanjutnya menentukan *Node* 1.1 pada pohon keputusan. Berikut ini perhitungan *Entropy* dan *Information Gain* *Node* 1.1 pada pohon keputusan (Tabel 3).

Tabel 3. Perhitungan *Node* 1.1 Pohon Keputusan

Entropy (Total) =	$= - \left[\frac{1}{4} \log_2 \frac{1}{4} + \frac{3}{4} \log_2 \frac{3}{4} \right] = 0,81127$
Entropy (Jumlah anak >= 3) =	$= - \left[\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3} \right] = 0,91829$
Entropy (Jumlah anak < 3) =	$= - \left[\frac{0}{1} \log_2 \frac{0}{1} + \frac{1}{1} \log_2 \frac{1}{1} \right] = 0$
Entropy (Wiraswasta/Karyawan) =	$= - \left[\frac{0}{2} \log_2 \frac{0}{2} + \frac{2}{2} \log_2 \frac{2}{2} \right] = 0$
Entropy (Tidak Bekerja/Buruh/ ART) =	$= - \left[\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right] = 1$
Information Gain (Jumlah Anak) =	$= 0,81127 - \left[\frac{3}{4} * 0,91829 + \frac{1}{4} * 0 \right] = 0,1225525$
Information Gain (Pekerjaan) =	$= 0,81127 - \left[\frac{2}{4} * 0 + \frac{2}{4} * 1 \right] = 0,31127$



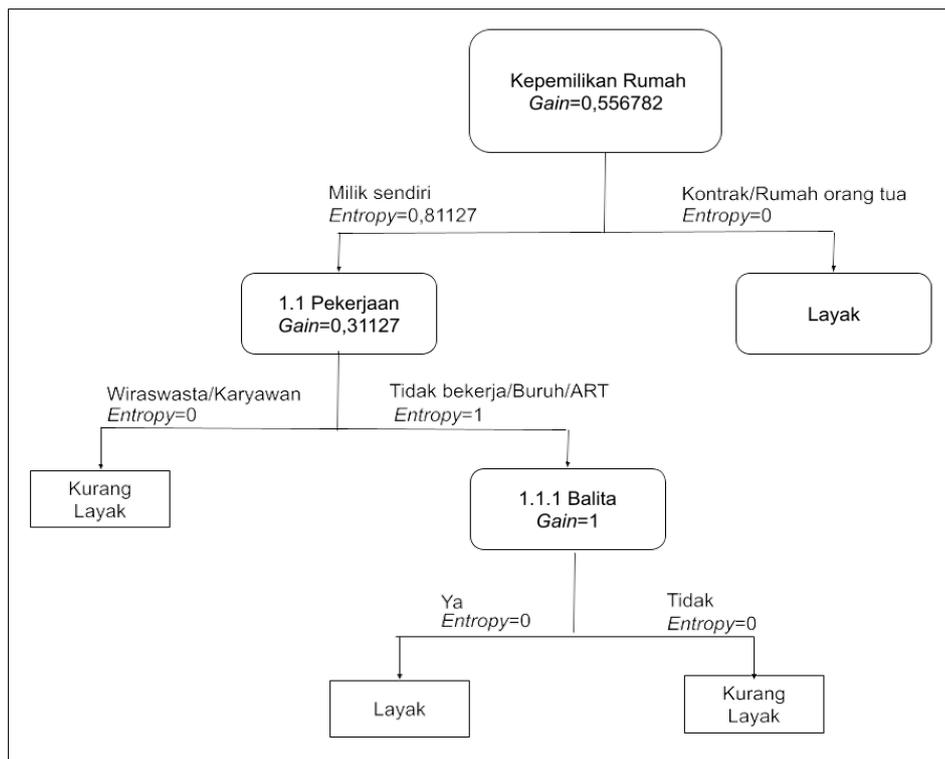
Gambar 11. *Node* 1.1 Pohon Keputusan

Pohon keputusan ini mengoptimalkan klasifikasi berdasarkan pengurangan *entropy* dan *informasi gain* dari atribut yang dipilih. Setelah mendapatkan *Node* 1.1, selanjutnya menentukan *Node* terakhir yaitu *Node* 1.1.1. Berikut perhitungan untuk nilai *Entropy* dan *Information Gain* *Node* 1.1.1 pada Pohon Keputusan (Tabel 4).

Tabel 4. Perhitungan *Node 1.1.1* Pohon Keputusan

$Entropy (Total) =$	$= - \left[\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right] = 1$
$Entropy (Anak Sekolah Ya) =$	$= - \left[\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right] = 1$
$Entropy (Anak Sekolah Tidak) =$	$= - \left[\frac{0}{0} \log_2 \frac{0}{0} + \frac{0}{0} \log_2 \frac{0}{0} \right] = 0$
$Entropy (Balita Ya) =$	$= - \left[\frac{1}{1} \log_2 \frac{1}{1} + \frac{0}{1} \log_2 \frac{0}{1} \right] = 0$
$Entropy (Balita Tidak) =$	$= - \left[\frac{0}{1} \log_2 \frac{0}{1} + \frac{1}{1} \log_2 \frac{1}{1} \right] = 0$
$Information Gain (Anak Sekolah) =$	$= 1 - \left[\frac{2}{2} * 1 + \frac{0}{2} * 0 \right] = 0$
$Information Gain (Balita) =$	$= 1 - \left[\frac{1}{10} * 0 + \frac{1}{10} * 0 \right] = 1$

Pada *Node 1.1.1* pohon Keputusan (Gambar 12), kelayakan penerima bantuan ditentukan melalui beberapa atribut jika status kepemilikan rumah adalah kontrak atau tinggal di rumah orang tua, maka penerima bantuan dinilai "Layak", karena pada atribut "Kepemilikan Rumah" pada data sampel penerima bantuan menunjukkan bahwa kebanyakan penerima bantuan dengan status kepemilikan rumah berupa kontrak atau tinggal di rumah orang tua dinilai sebagai "Layak".



Gambar 12. *Node 1.1.1* Pohon Keputusan

3.4 Pengujian dan Evaluasi

Tahapan pengujian dan evaluasi dilakukan dengan menggunakan *Confusion Matrix* terlihat pada Tabel 4 untuk membuktikan tingkat keberhasilan dalam penelitian pada saat membangun model klasifikasi menggunakan *Random Forest*. *Confusion Matrix* adalah kombinasi antara prediksi dan nilai sebenarnya. Terdapat istilah merepresentasikan proses klasifikasi pada *Confusion Matrix*, yaitu *true positif* (TP), *true negatif* (TN), *false positif* (FP), dan *false negatif* (FN). Pengujian klasifikasi pada metode yang digunakan memerlukan sebuah metode yang digunakan untuk mengevaluasi kinerjanya (Fajriyan et al., 2022).

Tabel 5. *Confusion Matrix*

<i>Confusion Matrix</i>	Prediksi Negatif	Prediksi Positif
Actual Negatif	TN	FP
Actual Positif	FN	TP

Keterangan:

TP : jumlah kelas positif yang diklasifikasi sebagai benar

FP : jumlah kelas positif yang diklasifikasi sebagai salah

TN : jumlah kelas negatif yang diklasifikasi sebagai salah

FN : jumlah kelas negatif yang diklasifikasi sebagai salah

Berdasarkan nilai *true* negatif (TN), *false* positif (FP), *false* negatif (FN), dan *true* positif (TP) dapat diperoleh nilai akurasi, *precision*, *recall*, dan *f1-score*.

Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan (1). Nilai *precision* menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar lalu dibagi dengan total data yang diklasifikasi positif. Nilai *Precision* dapat diperoleh dengan Persamaan (2). Sementara itu, nilai *recall* menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar. Nilai *recall* diperoleh dengan Persamaan (3). Nilai *f1-score* merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif. Nilai *f1-score* dapat dilihat pada Persamaan (4) (Kanaka et al., 2024).

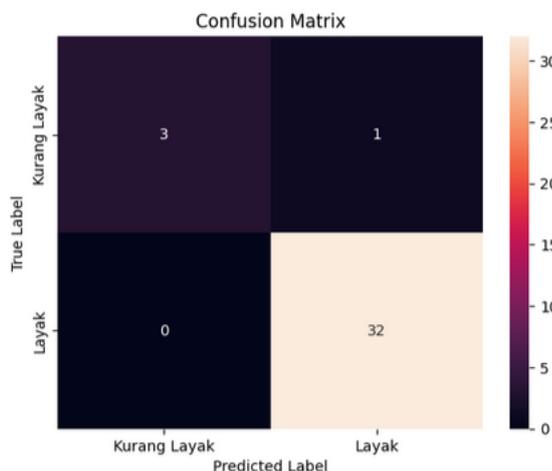
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Pengujian dilakukan menggunakan rasio 70:30. Pengujian menggunakan *dataset* pengujian 36 data pengujian dari 117 *dataset* yang terdiri 96 data 'Layak' dan 21 data 'Kurang Layak' pada label Kelayakan. Berikut merupakan hasil pengujian dan evaluasi menggunakan *Confusion Matrix* (Gambar 13).



Gambar 13. Hasil *Confusion Matrix*

Keterangan pada *Confusion Matrix*:

Total ada 36 data pengujian

32 data pengujian diklasifikasikan sebagai layak/positif

3 data pengujian diklasifikasikan sebagai kurang layak/negatif

1 data sampel layak/positif yang diprediksi sebagai kurang layak/negatif

0 data sampel kurang layak/negatif yang diprediksi sebagai layak/positif

Dari hasil *Confusion Matrix* berikut dapat dihitung:

True Positif (TP) = 32

True Negatif (TN) = 3

False Positif (FP) = 1

False Negatif (FN) = 0

Berdasarkan perhitungan *Confusion Matrix* sebelumnya, berikut perhitungan akurasi, *precision*, *recall*, dan *f1-score* dari data pengujian menggunakan Persamaan (1), (2), (3), dan (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{32 + 3}{32 + 3 + 1 + 0} = \frac{35}{36} = 0.9722 = 97.22\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{32}{32 + 1} = \frac{32}{33} = 0.9697 = 96.97\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{32}{32 + 0} = \frac{32}{32} = 1 = 100\%$$

$$F_1Score = 2 \times \frac{0.9697 \times 1}{0.9697 + 1} = 2 \times \frac{0.9697}{1.9697} = 2 \times 0.4922 = 0.9844 = 98.44\%$$

Hasil pengujian dan evaluasi *Confusion Matrix* menunjukkan performa yang sangat baik. Menghasilkan tingkat akurasi 97%, *precision* 96%, *recall* 100%, dan *f1-score* 98%, model pada *Random Forest* mendemonstrasikan kemampuan prediksi yang sangat akurat dan seimbang. Tingginya nilai semua metrik (di atas 90%) mengindikasikan bahwa model mampu mengklasifikasi dengan sangat baik dan membuat klasifikasi yang akurat., secara keseluruhan hasil ini menggambarkan model klasifikasi yang sangat andal dan efektif pada *dataset* yang digunakan.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan model klasifikasi pada metode *Random Forest* menggunakan 117 *dataset* penerima bantuan Program Keluarga Harapan (PKH) di Kelurahan 13 Ulu, Kecamatan Seberang Ulu II, Palembang yang telah di *preprocessing*, dengan rasio pembagian data 70:30 dengan jumlah data *training* 81 data dan data *testing* 36 data yang menghasilkan akurasi sebesar 97%, *precision* sebesar 96%, *recall* sebesar 100% dan *f1-score* sebesar 98%. Hal ini menunjukkan bahwa penggunaan metode *Random Forest* dapat melakukan klasifikasi dengan baik.

DAFTAR PUSTAKA

- Aldiyansyah, A., Irma Purnamasari, A., & Ali, I. (2024). Perbandingan Tingkat Akurasi Algoritma Decision Tree Dan Random Forest Dalam Mengklasifikasi Penerima Bantuan Sosial Bpnt Di Desa Slangit. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 127–132. <https://doi.org/10.36040/jati.v8i1.8290>
- Direktorat Jaminan Sosial, D. J. P. dan J. S., & Kementerian Sosial RI. (2021). *Pedoman Umum PKH Edisi 2021*.
- Fajriyan, F. N., Ahsan, M., & Harianto, W. (2022). Komparasi Tingkat Akurasi Information Gain Dan Gain Ratio Pada Metode K-Nearest Neighbor. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(1), 386–391. <https://doi.org/10.36040/jati.v6i1.4694>
- Ferdita Nugraha, A., Aziza, R. F. A., & Pristyanto, Y. (2022). Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing. *Jurnal Infomedia*, 7(1), 39. <https://doi.org/10.30811/jim.v7i1.2959>
- Kanaka, N. A. S., Heriansyah, R., & Puspasari, S. (2024). Perbandingan Algoritma Decision Tree dan Support Vector Machine Dalam Pemilihan Calon Mahasiswa Penerima KIP-K. *TIN: Terapan Informatika Nusantara*, 4(9), 613–619. <https://doi.org/10.47065/tin.v4i9.4902>
- Kurniawan, I., Buani, D. C. P., Abdussomad, A., Apriliah, W., & Saputra, R. A. (2023). Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(2), 421–428. <https://doi.org/10.25126/jtiik.20231026225>

- Maulana, Y. I. (2023). Implementasi Algoritma Random Forest untuk Klasifikasi Penerima Bantuan Bedah Rumah. *JSI: Jurnal Sistem Informasi (E-Journal)*, 15(1), 2999–3005. <https://doi.org/10.18495/jsi.v15i1.20449>
- Sholihah, N. N., & Hermawan, A. (2023). Implementation of Random Forest and Smote Methods for Economic Status Classification in Cirebon City. *Jurnal Teknik Informatika (Jutif)*, 4(6), 1387–1397. <https://doi.org/10.52436/1.jutif.2023.4.6.1135>
- Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121–127. <https://doi.org/10.35580/variasiunm31>
- Wanto, M. N. H., Windarto, Agus Perdana, Hartama, D., Ginantra, Ni Luh Wiwik Sri Rahayu, Napitupulu, D., Negara, Edi Surya, Lubis, M. R., & Dewi, Sarini Vita, Prianto, C. (2020). *Data Mining: Algoritma dan Implementasi*. Yayasan Kita Menulis.