

## Perbandingan Metode *Artificial Neural Network* (ANN) dan *Classification And Regression Tree* (CART) Dalam Menentukan Penerima Bantuan Program Keluarga Harapan (PKH)

Dita Septasari<sup>1</sup>, RZ. Abdul Aziz<sup>2\*</sup>

<sup>1,2</sup>Program Studi Magister Teknik Informatika, Institut Informatika dan Bisnis Darmajaya, Indonesia

<sup>1</sup>Universitas Aisyah Pringsewu, Indonesia

<sup>1,2</sup>ZA. Pagar Alam Street No. 93, Bandar Lampung City, Lampung 35141, Indonesia

<sup>1</sup>Jl. A Yani No. 1A Tambahrejo, Kec. Gadingrejo, Kab. Pringsewu, Lampung, 35372, Indonesia

\*Email: [rz\\_aziz@damajaya.ac.id](mailto:rz_aziz@damajaya.ac.id)

[ditaseptasari.2221210069@mail.darmajaya.ac.id](mailto:ditaseptasari.2221210069@mail.darmajaya.ac.id)

Info Artikel	ABSTRAK
Dikirim: - Diterima: - Diterbitkan: -	Kemiskinan yang ada di Indonesia dipengaruhi oleh beberapa faktor. Dalam hal ini pemerintah mengupayakan berbagai program peningkatan kesejahteraan masyarakat. Program tersebut salah satunya adalah Program Bantuan Keluarga Harapan (PKH). Prediksi terhadap penerima bantuan PKH menggunakan metode <i>Artificial Neural Network</i> (ANN) dan <i>Classification And Regression Tree</i> (CART) untuk membantu dalam proses penentuan penerima program keluarga harapan. Metode <i>Artificial Neural Network</i> (ANN) dan <i>Classification And Regression Tree</i> (CART) menghasilkan pengukuran performa (akurasi, presisi, <i>recall</i> dan <i>f1-score</i> ) dari data penerima Program Bantuan Keluarga Harapan (PKH). Hasil dari penelitian ini dapat diperoleh performa jika Metode <i>Artificial Neural Network</i> (ANN) lebih baik dibandingkan dengan <i>Classification And Regression Tree</i> (CART) untuk memprediksi dalam menentukan penerima bantuan PKH di Kabupaten Pringsewu.
<b>Kata kunci:</b> <i>Artificial Neural Network</i> ; <i>Classification And Regression Tree</i> ; Data Mining; Klasifikasi; Program Keluarga Harapan (PKH);	

### 1. PENDAHULUAN

Kemiskinan adalah salah satu tantangan yang dihadapi oleh negara-negara berkembang seperti Indonesia. Permasalahan kemiskinan ini bisa disebabkan oleh beberapa faktor diantaranya seperti kurangnya lapangan pekerjaan sehingga banyak pengangguran, standar atau kesadaran akan pendidikan yang masih rendah, terjadinya inflasi atau beberapa faktor lainnya. Berdasarkan website Badan Pusat Statistik (BPS) Indonesia yang berwenang dalam mendata penduduk Indonesia pada tanggal 16 Januari 2023 menginfokan melalui websitenya presentase penduduk miskin Indonesia pada bulan Maret 2022 mencapai 26.36 juta penduduk dan ini merupakan 9.57% dari seluruh pendudukan Republik Indonesia. Jumlah penduduk miskin mengalami kenaikan mencapai 26.36 juta penduduk yang artinya naik dibanding periode pencatatan sebelumnya dibulan maret, mencapai 9.60 pada bulan September.[1] Pemerintah Indonesia mengupayakan untuk mengatasi kemiskinan dengan berbagai program bantuan untuk kesejahteraan Masyarakat, salah satunya adalah Program Keluarga Harapan (PKH). Program Keluarga Harapan (PKH) adalah inisiatif bantuan sosial yang bertujuan meningkatkan kesejahteraan masyarakat dengan melibatkan partisipasi penerima manfaat untuk menjaga kesehatan dan memastikan anak-anak mereka mendapatkan pendidikan.[2]

Dalam penelitian yang telah dilakukan oleh peneliti sebelumnya oleh Ir. Adi Sucipto, M.Kom yang berjudul Credit Prediction With Neural Network Algorithm tahun 2015 memamparkan terkait dengan metode *Artificial Neural Network* (ANN) dalam memprediksi koperasi simpan pinjam dalam hal ini ANN berhasil dengan baik

dalam kredit macet yang ada dalam koperasi.[3] Tahun 2019 Nurhadianto, dkk dalam penelitiannya Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode *Neural Network* memberikan informasi jika ANN menunjukkan hasil yang baik dalam klasifikasi nasabah dalam menentukan nasabah yang berhak memperoleh pinjaman Bank.[4] Pada penelitian lainnya, yang berjudul Pengaplikasian *Artificial Neural Network* (ANN) dalam memprediksi curah hujan menggunakan Python yang dilakukan oleh Muhammad Rizaldi M, dkk pada 2021 membuktikan ANN juga memiliki akurasi yang baik dibandingkan dengan menggunakan metode time series dalam memprediksi curah hujan.[5] Penelitian lainnya yang membandingkan 2 metode yakni *Classification And Regression Tree* (CART) dan *Support Vector Machin* (SVM) yang dilakukan oleh M. Anwar Sadat, dkk (2023) dalam penelitian yang berjudul *Comparison Of Algorithm Between Classification & Regression Trees And Support Vector Machine In Determining Student Acceptance In State Universities* yang bagaimana metode CART dan SVM menghasilkan akurasi masing-masing dalam mengklasifikasikan siswa yang masuk ke perguruan tinggi.[6] Tahun 2014 Febtu Eka Pratiwi dan Ismaini Zain menggunakan metode CART dalam penelitiannya yang berjudul *Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree)* di Provinsi Sulawesi Utara dalam penelitian ini menyimpulkan jika ketepatan klasifikasi bisa mencapai lebih tinggi lagi dengan menambahkan variable untuk mewakili karakteristik pengangguran di Sulawesi Utara.[7]

Pada penelitian ini peneliti memilih mengambil penelitian dalam Program Keluarga Harapan (PKH) untuk memberikan keberagaman dalam penelitian dibidang teknologi dan sosial. Metode dalam penerimaan bantuan PKH sebelumnya dilakukan secara manual. Maka dari itu, dilakukan penelitian yang berjudul “Perbandingan Metode *Artificial Neural Network* (ANN) dan *Classification And Regression Tree* (CART) untuk memudahkan pihak terkait dalam mengklasifikasikan keluarga dengan kelayakan penerima bantuan PKH. Perbedaan menggunakan metode ANN dan CART bergantung pada karakteristik data dan tujuan analisis data, Model CART memberikan pemahaman tentang faktor-faktor utama yang mempengaruhi variabel target. CART mampu mengatasi masalah data karena partisi dilakukan berdasarkan pemisahan variabel. Metode ini menawarkan peningkatan yang signifikan dalam pemisahan data dan memiliki kemampuan untuk menangani data non-linear.[7] CART dapat dengan mudah mengelola hubungan non-linear antara variabel independen dan variabel target melalui pemisahan berbasis aturan yang dilakukan pada setiap node.[6] *Artificial Neural Network* terdiri dari sejumlah neuron, yang berfungsi sebagai elemen pemrosesan informasi. Neuron-neuron ini terhubung satu sama lain dan bekerja secara kolaboratif untuk menyelesaikan masalah tertentu, seperti klasifikasi atau prediksi. ANN adalah sistem adaptif yang dapat menyesuaikan strukturnya untuk memecahkan masalah berdasarkan informasi yang diterima dari lingkungan eksternal maupun internal.[8] Metode *Artificial Neural Network* (ANN) dan *Classification And Regression Tree* CART dapat membantu Pembina keluarga PKH dalam memprediksi masyarakat yang berhak menerima bantuan PKH.[5] Perbandingan Metode yang dilakukan untuk mengetahui Tingkat akurasi yang tepat terhadap klasifikasi penerima bantuan PKH di Kabupaten Pringsewu.

## 2. METODE PENELITIAN

### 2.1 Data Mining

Data Mining adalah proses mengumpulkan dan mengolah data untuk mengekstrak informasi penting. [9] Proses ini dapat dilakukan dengan bantuan perangkat lunak yang menggunakan perhitungan statistika, matematika, atau teknologi *Artificial Intelligence* (AI).[10]

### 2.2 *Artificial Neural Network* (ANN) atau Jaringan Saraf Tiruan

Algoritma *Artificial Neural Network* (ANN) merupakan metode dalam pemrosesan sebuah informasi yang sistem kerjanya seperti cara kerja otak manusia.[11] ANN terdiri dari berbagai neuron sederhana yang terhubung satu sama lainnya. Neuron-neuron ini terhubung melalui bobot (*weight*) yang mengirimkan sinyal dari satu neuron ke neuron lainnya. ANN merupakan metode yang meniru bagaimana jaringan saraf biologis bekerja. Melalui proses pembelajaran, jaringan saraf tiruan dapat menyesuaikan diri untuk memberikan respons yang konsisten terhadap berbagai input. Jaringan saraf tiruan dirancang dan dilatih untuk meniru kemampuan manusia. Setiap neuron menerima beberapa input dan menghasilkan satu output. Input pada neuron bisa berupa data mentah atau hasil pemrosesan dari neuron sebelumnya, sementara output neuron dapat menjadi hasil akhir atau input bagi neuron berikutnya.[12]

Jaringan neuron buatan terdiri dari beberapa kelompok neuron yang terbagi dalam tiga lapisan utama:

- Lapisan *Input (Input Layer)*: Bertugas sebagai penghubung antara jaringan dan sumber data eksternal. Neuron-neuron pada lapisan ini tidak memproses data, melainkan hanya menyampaikan data ke lapisan berikutnya.
- Lapisan Tersembunyi (*Hidden Layer*): Jaringan dapat memiliki satu atau lebih lapisan tersembunyi, atau bahkan tidak memilikinya sama sekali. Jika terdapat beberapa lapisan tersembunyi, lapisan yang paling bawah akan menerima input dari lapisan input.
- Lapisan *Output (Output Layer)*: Cara kerja lapisan ini mirip dengan lapisan tersembunyi dan sering menggunakan fungsi sigmoid. Namun, *output* dari lapisan ini merupakan hasil akhir dari proses jaringan.[13]

Dalam melakukan prediksi akurat terhadap pemilihan penerima bantuan program PKH menggunakan metode ANN dilakukan dengan beberapa tahapan:

#### a. *Pre-Processing*

Merupakan tahapan dalam mengkondisikan data. Pada penelitian ini data yang digunakan merupakan data masyarakat penerima bantuan PKH pada tahun 2023.

#### b. *Training dan Testing*

*Training* adalah tahap di mana pola nilai dalam suatu data dipelajari sehingga model dapat memprediksi nilai di masa mendatang. Hasil dari proses training ini akan menghasilkan nilai MSE yang digunakan untuk menentukan jenis uji yang paling mendekati kebenaran. Semakin banyak contoh data yang digunakan, semakin kecil kemungkinan kesalahan prediksi yang akan dihasilkan oleh model ANN, yang dirumuskan sebagai berikut:

$$MSE = \frac{\sum (y_i - \check{y}_i)^2}{n}$$

$y_i$  = Data sebenarnya

$\check{y}_i$  = Nilai Prediksi dari Variabel Y

$n$  = Banyak observasi

*Testing* adalah tahap pengujian yang dilakukan untuk membuktikan kesesuaian atau keakuratan hasil dengan data yang digunakan, di mana data testing ditampilkan dalam bentuk grafik. [5] Pembagian data training dan data testing bisa dilakukan dengan rasio 70%:30%, 80%:20%, atau 90%:10%. Tidak ada aturan pasti mengenai pembagian data ini. Untuk menemukan rasio yang optimal, ketiga rasio tersebut digunakan dalam analisis.

**Hasil klasifikasi ANN** dengan parameter kemudian diuji validitasnya dengan menggunakan *confusion matrix* berupa *Precision, Recall* dan *Accuracy*. [14]

### 2.3 *Classification And Regression Tree (CART)*

CART Merupakan salah satu metode atau algoritma yang digunakan untuk mengeksplorasi data dengan penggambaran pohon Keputusan. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penjelasan lebih lanjut dari suatu pengklasifikasian. Keluaran dari algoritma ini berupa klasifikasi jika variable targetnya adalah data numerik atau kontinu maka keluaran berupa pohon regres algoritma. Rumus 1:

$$i(t) = \sum_{j \neq k} p(j|t)p(k|t)$$

$p(j|t)$  = proporsi kelas j pada simpul t

$p(k|t)$  = proporsi kelas k pada simpul t

#### a. **Model Klasifikasi *Classification And Regression Tree (CART)***

Model klasifikasi *Classification And Regression Tree (CART)* adalah algoritma pembelajaran mesin yang digunakan untuk membangun pohon keputusan dalam rangka melakukan klasifikasi pada data.[15] CART dapat diterapkan baik untuk masalah klasifikasi (memprediksi kelas diskrit) maupun regresi (memprediksi nilai kontinu). Kelebihan utama dari CART dalam tugas klasifikasi adalah proses perhitungannya yang sederhana, relatif cepat, serta mudah diterapkan dalam berbagai kasus. Pemilah terbaik dipilih berdasarkan nilai penurunan rata-rata persemester yang tertinggi dari semua kemungkinan pemilah yang diuji.[6] Metode CART adalah

metode yang mudah untuk diinterpretasikan. Selain itu, metode ini juga menggunakan teknik pencarian data terstruktur dalam bentuk pohon. CART menghasilkan dua jenis pohon: pohon regresi dan pohon klasifikasi. Pohon regresi digunakan untuk mengubah respons yang bersifat kontinu, sementara pohon klasifikasi digunakan untuk respons kategorik. Proses pembentukan pohon klasifikasi dilakukan secara rekursif.[16]

## b. Tahapan *Classification And Regression Tree (CART)*

### 1) Melakukan data *Preprocessing*

Data preprocessing dilakukan untuk menyiapkan data agar sesuai dengan kebutuhan analisis, sehingga meminimalkan masalah selama proses pengolahan. Proses ini mengubah data mentah menjadi data berkualitas yang siap diolah dalam tahap analisis berikutnya. Ada lima langkah dalam data preprocessing: pembersihan data, optimasi data, transformasi data, integrasi data, dan konversi data. Dalam penelitian ini, hanya dua langkah yang diterapkan, yaitu pembersihan data dan transformasi data. Pembersihan data melibatkan penanganan data yang hilang dan *undersampling* pada data yang tidak seimbang.

### 2) Membagi data *Training* dan *Data Testing*

Data *training* digunakan untuk melatih algoritma dalam pengembangan model, sementara data *testing* digunakan untuk menguji model yang telah dibuat.

### 3) Melakukan Klasifikasi Menggunakan Algoritma CART

CART adalah metode klasifikasi yang menghasilkan model berbentuk pohon keputusan. Metode klasifikasi dalam CART terdiri dari dua jenis: pohon klasifikasi dan pohon regresi. Pohon klasifikasi digunakan untuk variabel respons yang bersifat kategorikal, sementara pohon regresi diterapkan pada variabel respons yang bersifat numerik atau kontinu.

## 4) Melakukan Klasifikasi dengan Algoritma CART

### a) Pemilihan pemisah

Pemisahan data dilakukan berdasarkan aturan pemisah dan kriteria goodness of split. Setelah data dipisahkan, himpunan bagian yang dihasilkan harus lebih homogen dibandingkan dengan pemisahan sebelumnya. Indeks gini merupakan fungsi keheterogenan yang sangat mudah digunakan dan cocok untuk berbagai kasus. Pemisahan terbaik adalah pemisahan yang memiliki penurunan nilai keheterogenan tertinggi. Fungsi indeks gini didefinisikan sebagai berikut[6]:

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j$$

dimana:

$p(j|t)$ : Proporsi kelas  $j$  pada simpul  $t$

$p(i|t)$ : Proporsi kelas  $i$  pada simpul  $t$

Setelah menghitung indeks gini, langkah selanjutnya adalah menghitung goodness of split ( $\phi(s, t)$ ) untuk menilai pemisah. Goodness of split mengukur penurunan keheterogenan dengan menggunakan persamaan berikut.

$$(\phi(s, t)) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R)$$

Pemisah terbaik adalah pemisah yang memiliki nilai  $\Delta i(s, t)$  tertinggi karena mampu mengurangi heterogenitas secara signifikan. Nilai  $\Delta i(s, t)$  menunjukkan perubahan keheterogenan pada simpul  $t$  akibat pemisah  $s$ . Jika kelas pada simpul masih belum homogen, proses ini diulangi hingga pohon klasifikasi mencapai konfigurasi yang memenuhi kriteria berikut:

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1)$$

### b) Penentuan Simpul Terminal

Simpul terminal terbentuk ketika data pada simpul termasuk dalam kelas yang sama. Pengembangan pohon keputusan dihentikan jika terdapat minimal 5 pengamatan pada simpul tersebut.

### c) Penanda Label Kelas

Label kelas diberikan berdasarkan kelas yang memiliki jumlah pengamatan terbanyak pada simpul terminal, yaitu jika:

$$p(j|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)}$$

dimana:

$p(j|t)$  : proposi kelas  $j$  pada simpul  $t$

$N_j(t)$  : jumlah amatan kelas  $j$  pada terminal node  $t$

$N(t)$  : jumlah total amatan pada terminal node  $t$

### 5) Pemangkasan Pohon Keputusan

Ukuran pohon yang terlalu besar bisa menyebabkan overfitting, sementara pohon yang terlalu kecil bisa menyebabkan underfitting. Oleh karena itu, untuk mendapatkan pohon keputusan yang optimal, dilakukan pemangkasan pohon keputusan. Pemangkasan ini bertujuan untuk menghasilkan pohon yang lebih sederhana. [13] Pemangkasan dapat dilakukan menggunakan Cost Complexity Pruning (CPP). Berikut adalah persamaan cost complexity:

$$R_\alpha(T) = R(T) + \alpha|\bar{T}|$$

### 6) Menghitung Ketepatan Klasifikasi

Metode *confusion matrix* digunakan untuk mengevaluasi kinerja atau akurasi proses klasifikasi, mencakup nilai akurasi, presisi, dan sensitivitas/*recall* dengan persentase antara 0% hingga 100%. Semakin tinggi persentase nilai tersebut, semakin baik model dianggap.

Tabel 2. 1 Metode *confusion matrix*

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	True Negatif (TN)	False Positif (FP)
Aktual Positif	False Negatif (FN)	True Positif (TP)

Berikut adalah cara menghitung nilai akurasi, presisi, dan sensitivitas/*recall* berdasarkan tabel *confusion matrix*[17].

#### a) Akurasi

Akurasi mengukur seberapa tepat model dalam mengklasifikasikan data dengan benar. Akurasi dihitung dengan membandingkan jumlah observasi yang diklasifikasikan dengan benar dengan total jumlah observasi. Semakin tinggi persentase nilai akurasi, semakin besar tingkat kedekatan antara nilai prediksi dan nilai aktual.

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

#### b) Presisi

Presisi mengukur sejauh mana akurasi model dalam memprediksi kelas positif. Nilai presisi dihitung dengan membandingkan jumlah prediksi benar positif dengan total jumlah prediksi positif.

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

#### c) Recall

*Recall* mengukur seberapa baik model dalam menemukan informasi yang relevan. Nilai *recall* dihitung dengan membandingkan jumlah prediksi benar positif dengan total jumlah observasi yang sebenarnya positif.[18]

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

Keterangan:

- TP : jumlah observasi positif yang benar-benar terdeteksi sebagai positif.
- TN : jumlah observasi negatif yang benar-benar terdeteksi sebagai negatif.
- FP : jumlah observasi negatif yang terdeteksi sebagai positif.
- FN : jumlah observasi positif yang terdeteksi sebagai negatif.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Data Penelitian

Dataset keluarga penerima bantuan PKH pada penelitian ini terdiri dari 60 data kemudian dari data yang ada akan dibagi menjadi data *training* dan *testing* untuk klasifikasi.

Tabel 3. 1 Kategori dataset

No	Atribut	Variabel	Keterangan
1	Pekerjaan	0	Honorar
		1	Pedangang
		2	Buruh
		3	Petani
2	Penghasilan	0	> Rp. 3.500.001
		1	Rp. 2.500.001 – Rp. Rp. 3.500.000
		2	Rp. 1.500.001 – Rp. 2.500.000
		3	< Rp. 1.500.000
3	Status Rumah	0	Memiliki
		1	Belum Memiliki
4	Jenis Lantai	0	Keramik
		1	Semen
		2	Tanah
5	Jenis Dinding	0	Permanen
		1	Non Permanen
6	Bumil (Ibu Hamil)	0	Tidak Ada
		1	Ada
7	Balita	-	Jumlah anak
8	SD	-	Jumlah anak
9	SMP	-	Jumlah anak
10	SMA	-	Jumlah anak
11	Lansia	0	Tidak Ada
		1	Ada
12	Disabilitas	0	Tidak Ada
		1	Ada
13	Status	0	Tidak diterima
		1	Diterima

Dari dataset masyarakat penerima bantuan PKH selanjutnya dilakukan pengujian sehingga diperoleh hasil pengujian terhadap data keluarga penerima bantuan PKH menggunakan pemodelan ANN dan CART dengan software dan Bahasa Pemrograman Phyton. Dibawah ini merupakan tabel tahap transformasi data masyarakat penerima bantuan PKH

Tabel 3. 2 Tabel Transformasi Data

No	Pekerjaan	Penghasilan	Status Rumah	Jenis Lantai	Jenis Dinding	Bumil	Balita	SD	SMP	SMA	LANSIA	DISABILITAS	STATUS
1	2	2	0	1	0	0	0	2	0	0	0	0	0
2	2	2	0	2	0	0	2	0	0	0	0	0	0
3	3	2	0	1	0	0	1	1	0	0	0	0	1
4	2	2	0	2	1	0	0	1	1	0	0	0	1
5	0	2	1	1	0	0	1	1	0	0	0	0	1

#### 3.2 Artificial Neural Network (ANN)

##### a. Proses Pelatihan (*Training*) Data

Pada proses pelatihan data bertujuan untuk memperoleh nilai posterior dari masing-masing atribut pada masing masing kelas. Dalam penelitian ini digunakan 80% data *training* dan 20% data *testing*. Hasil dari pengujian ini diketahui :

*Total params* (jumlah total parameter dalam model) : 13,697 (53.50 KB)

*Trainable params* (jumlah parameter yang dilatih) : 13,313 (52.00 KB)

*Non-trainable params* (jumlah parameter yang tidak dapat dilatih) : 384 (1.50 KB)

Berdasarkan nilai dari parameter model memberikan gambaran tentang ukuran model dan kompleksitasnya, selain itu juga memberikan informasi model yang dapat beradaptasi dan belajar dari data, selain itu juga menginformasikan bagian dari model yang tetap konstan dan tidak diperbaharui selama pelatihan.

### b. Proses Pengujian (*Testing*) Data

Proses dalam ANN selanjutnya melakukan pengujian terhadap model yang sudah dilatih dengan data uji, untuk mendapatkan matrik kinerja akhir model setelah pelatihan. Dalam penelitian ini hasil dari pengujian data :

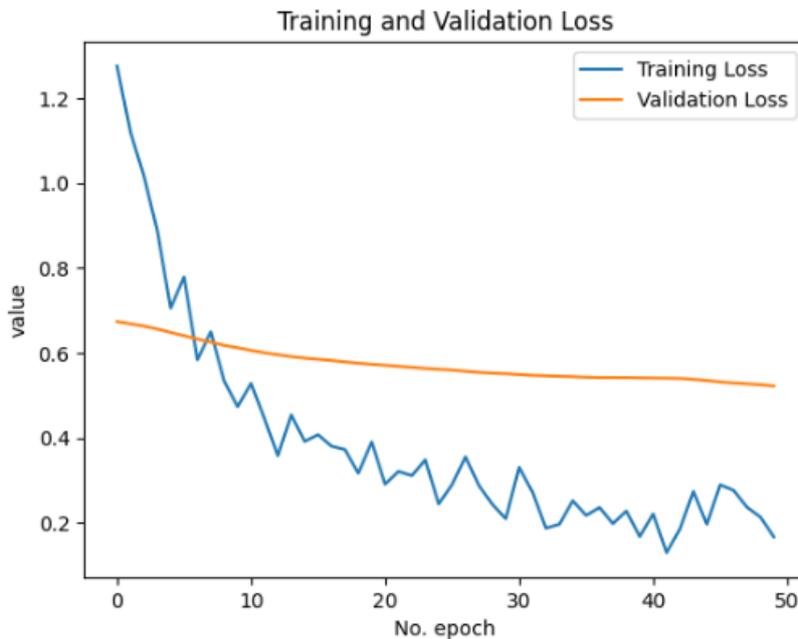
*Test loss testing*: 0.588708221912384

*Test accuracy testing*: 0.75

Berdasarkan *testing* yang dilakukan pada model menunjukkan bahwa model memiliki beberapa tingkat kesalahan dalam prediksinya dan dalam rentang yang wajar (*accuracy testing* > *loss testing*). Selain itu, nilai akurasi menunjukkan lebih dari 75% yang artinya nilai yang cukup baik dalam klasifikasi dan dapat diandalkan dalam banyak kasus.

### c. Visualisasi Data

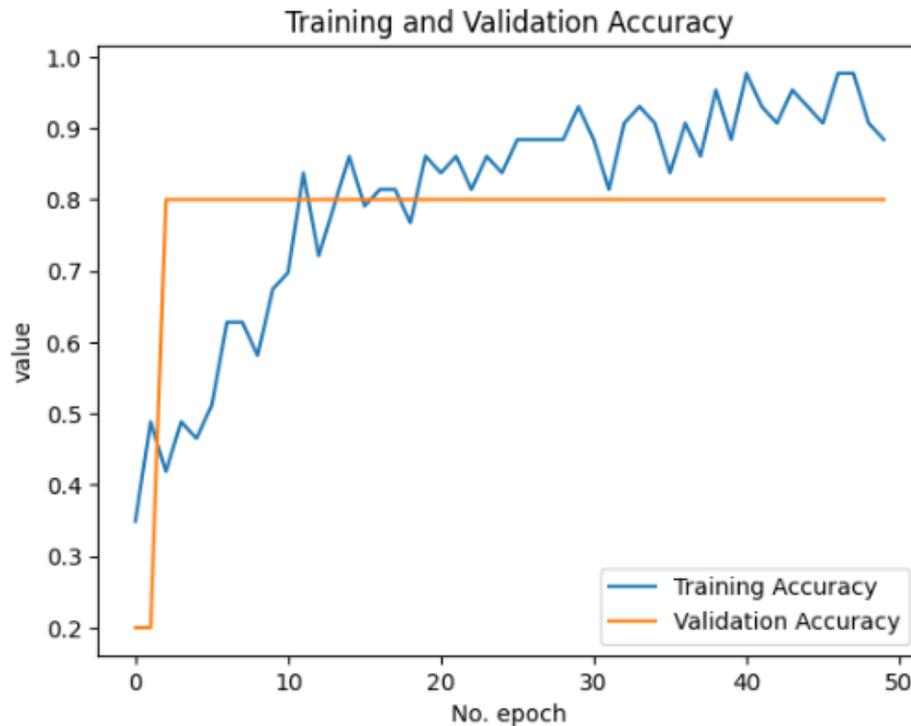
Dibawah ini merupakan grafik untuk memvisualisasikan data pelatihan dan validasi dalam setiap epoch pelatihan.



Gambar 3. 1 Visualisasi data *Training Loss* dan *Validation Loss*

*Training Loss*: Ini adalah ukuran seberapa baik model melakukan prediksi pada data pelatihan (*training data*). Semakin rendah nilai training loss, semakin baik model dalam memprediksi data pelatihan. Pada gambar 3.1 menunjukkan bahwa training loss semakin rendah sehingga pada penelitian ini model semakin baik dalam memprediksi pelatihan.

*Validation Loss*: Ini adalah ukuran seberapa baik model melakukan prediksi pada data yang tidak digunakan selama pelatihan, yaitu data validasi (*validation data*). *Validation loss* membantu mengukur kemampuan generalisasi model pada data baru yang belum pernah dilihat sebelumnya. Jika *validation loss* tetap rendah atau menurun selama pelatihan, ini menandakan bahwa model tidak hanya belajar dengan baik pada data pelatihan tetapi juga dapat menggeneralisasi dengan baik ke data baru. Dari gambar 3.1 diketahui jika *Validation Loss* pada penelitian ini menurun selama pelatihan dan ini menandakan model belajar dengan baik data pelatihan dan juga dapat menggeneralisasi dengan baik ke data baru.



Gambar 3. 2 Visualisasi data *Training Accuracy* dan *Validarion Accuracy*

*Training Accuracy*: Ini mengukur seberapa sering model membuat prediksi yang benar pada data pelatihan (*training data*). *Training accuracy* dihitung dengan membandingkan jumlah prediksi yang benar dengan jumlah total prediksi pada data pelatihan.[19] Tingginya *training accuracy* menunjukkan bahwa model dapat memprediksi data pelatihan dengan baik. Berdasarkan gambar 3.2 diketahui jika nilai *training accuracy* pada penelitian semakin naik dan ini menunjukkan bahwa model memprediksi data dengan baik.

*Validation Accuracy*: Ini mengukur seberapa baik model memprediksi data yang tidak digunakan selama pelatihan, yaitu data validasi (*validation data*) dan berdasarkan hasil dari visualiasasi dapat dilihat bahwa akurasi validasi data pada penelitian ini tetap, selama *traning accuracy* yang terus naik.

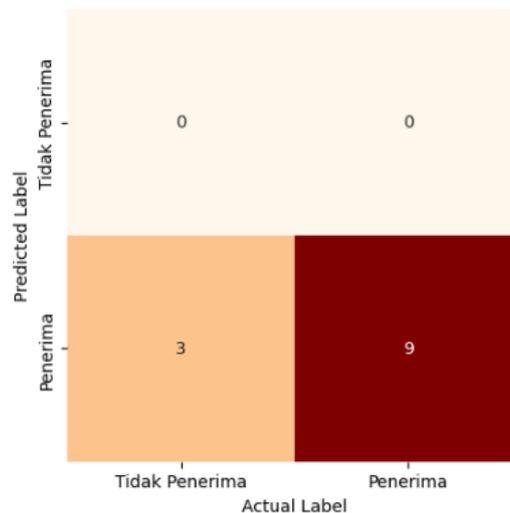
Selanjutnya mengidentifikasi area model yang bekerja dengan baik dan mengetahui ruang untuk perbaikan.

Tabel 3. 3 *Classification Report ANN*

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Tidak Penerima	0.00	0.00	0.00	3
Penerima	0.75	1.00	0.86	9
<i>accuracy</i>	-	-	0.75	12
<i>macro avg</i>	0.38	0.50	0.43	12
<i>weighted avg</i>	0.56	0.75	0.64	12

*Accuracy* dalam penggunaan metode ANN Terhadap data keluarga penerima bantuan PKH mencapai 0.75.

Menggunakan visualiasasi data dengan heatmap dalam memvisualisasikan matriks kebingungan dan memudahkan analisis kinerja model klasifikasi, selain itu visualiasasi dengan heatmap untuk mengindetifikasikan area model yang bekerja dengan baik dan yang memerlukan perbaikan.



Gambar 3. 3 Visualiasasi Data Heatmap

Dari hasil uji 20% yang dilakukan pada penelitian ini, didapatkan hasil bahwa terdapat 9 data penerima yang diprediksi benar oleh model. Dan terdapat 3 data tidak penerima yang diprediksi salah oleh model. Dalam hal ini model memprediksi data benar lebih banyak dari prediksi salah dari model.

### 3.3 Classification And Regression Tree (CART)

Pada tahapan ini melakukan proses untuk melakukan manipulasi data, membangun model klasifikasi dan mengevaluasi kinerja model.

#### 1) Melakukan data *preprocessing*

Analisis data dilakukan menggunakan perangkat lunak Python. Sebelum analisis data, dilakukan proses pra-proses data yang mencakup penanganan data yang hilang, transformasi data, dan penanganan ketidakseimbangan data. Dalam penelitian ini terdapat 13 data dengan status tidak diterima dan 47 status data diterima sebagai penerima bantuan PKH. Data yang akan digunakan dalam penelitian ini meliputi data pekerjaan, penghasilan, status rumah, jenis lantai, jenis dinding, bumil dan anak atau lansia dan penyandang disabilitas yang berada dalam keluarga tersebut. Selanjutnya data data tersebut dilakukan transformasi data dengan menggunakan kategorikal seperti yang terdapat pada tabel 3.1 diatas.

#### 2) Membagi data training dan data testing

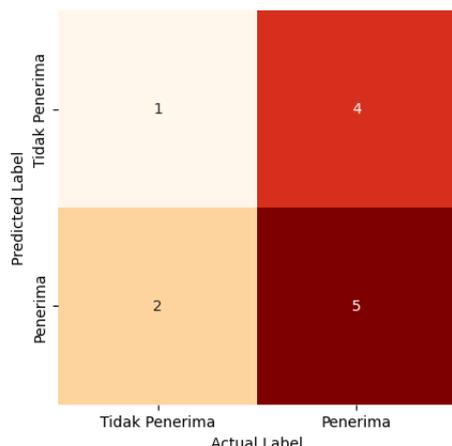
Dalam penelitian proporsi data yang akan digunakan sebagai set pengujian. Dalam hal ini, 20% dari total data akan digunakan sebagai set pengujian, sementara 80% sisanya akan digunakan sebagai dataset pelatihan yang dalam hal ini terdapat 12 data sebagai pengujian (*testing*) dan 48 data yang digunakan sebagai data latih (*training*).

#### 3) Melakukan klasifikasi menggunakan algoritma CART

Dari hasil split yang dilakukan diperoleh hasil nilai *accuracy* : 0.9583 atau akurasi 95.83% menunjukkan bahwa model memiliki performa yang sangat baik pada data *training*. Model ini mampu mempelajari pola dalam data dengan cukup baik untuk melakukan prediksi yang tepat pada sebagian besar contoh. Selain itu nilai akurasi yang mencapai 0,42. Akurasi 42% menunjukkan bahwa model hanya mampu memprediksi dengan benar sekitar 42% dari total sampel data uji. Dengan kata lain terdapat 58% dari prediksi model tersebut salah.

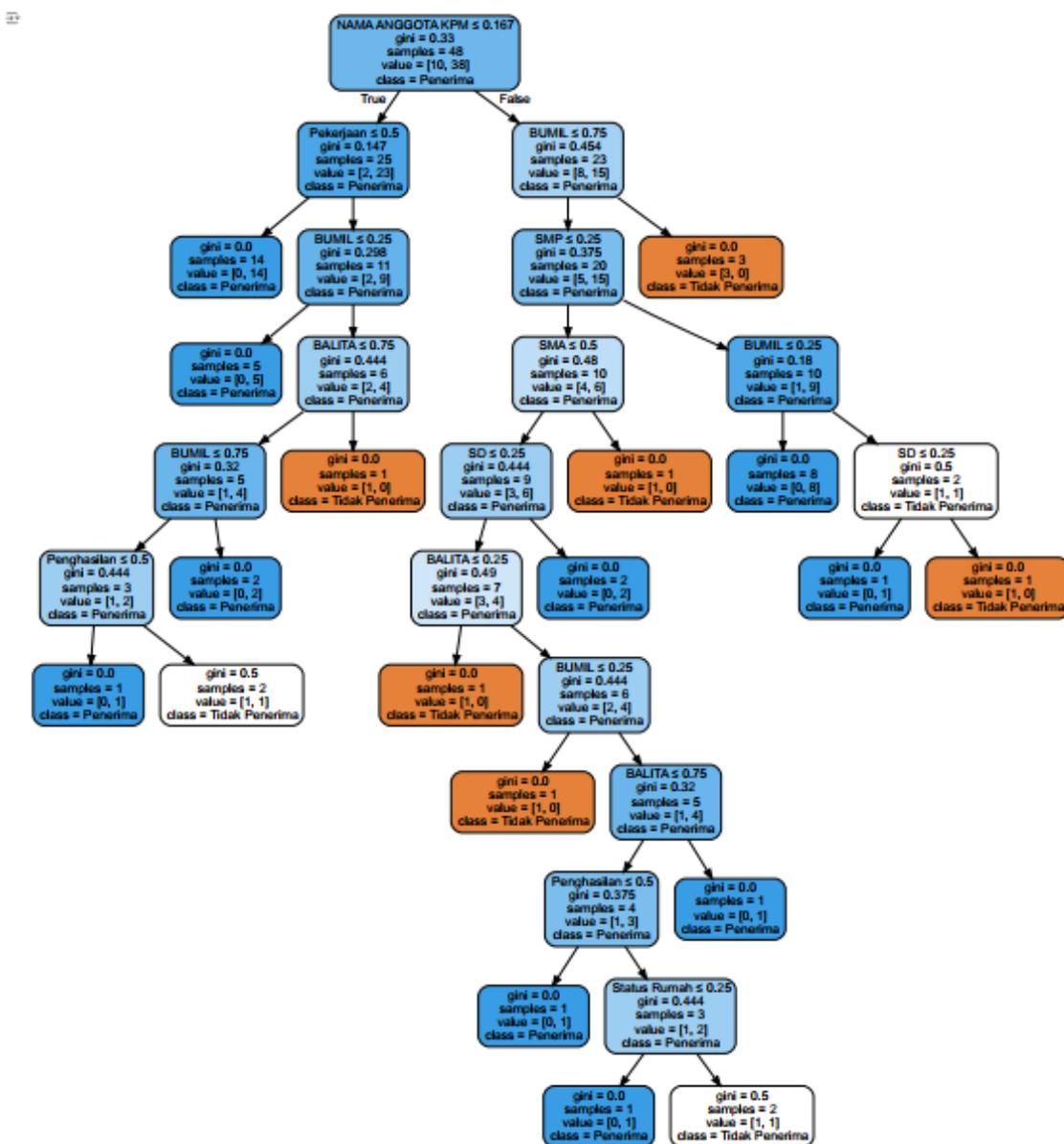
Tabel 3.4 Classification Report CART

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Tidak Penerima	0.17	0.33	0.22	3
Penerima	0.67	0.44	0.53	9
<i>accuracy</i>	-	-	0.42	12
<i>macro avg</i>	0.42	0.39	0.38	12
<i>weighted avg</i>	0.54	0.42	0.46	12



Gambar 3. 4 Visualiasasi Data Heatmap

Dari gambar 3.4 diketahui jika pada terdapat 1 data tidak penerima yang diprediksi benar oleh model, terdapat 5 data penerima yang diprediksi benar oleh model. Terdapat 2 data penerima yang diprediksi salah oleh model dan terdapat 4 data tidak penerima yang diprediksi salah oleh model.



Gambar 3. 5 Pohon CART

Analisis CART menggunakan Python dilakukan dengan mengimpor pustaka `DecisionTreeClassifier` dari `sklearn.tree`. Kriteria yang diterapkan adalah nilai Gini. Dalam proses ini, pohon keputusan CART yang dibentuk dengan membagi data pelatihan dan pengujian dalam berbagai rasio menghasilkan bentuk pohon yang berbeda. Pohon keputusan yang dihasilkan memberikan informasi nilai gini serta kelas pada setiap cabang.

#### Perbandingan hasil ANN dan CART terhadap nilai penerima bantuan PKH

Tabel 3. 4 Hasil Perbandingan ANN dan CART Terhadap data Penerima Bantuan PKH

Metode	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
CART	0.42	0.67	0.44	0.53
ANN	0.75	0.75	1.00	0.86

#### 4. KESIMPULAN

Kesimpulan pada penelitian ini :

- 1) Metode ANN menunjukkan akurasi prediksi yang dilakukan ANN pada penelitian ini cukup baik dengan nilai akurasi 0.75 yang berarti model ANN berhasil memprediksi dengan benar 75% dari total contoh yang ada dalam dataset, dibandingkan dengan CART yang memperoleh 0.50 akurasi untuk dataset.
- 2) Metode ANN yang digunakan pada penelitian ini memiliki kinerja yang cukup baik pada data pengujian, dengan nilai *precision*, *recall*, dan *f1-score* 0,75, 1.00 dan 0.86. Nilai 0.86 untuk *f1-score* menunjukkan keseimbangan yang cukup baik antara *Precision* (75%) dan *Recall* (100%), dengan sedikit penekanan pada *Recall* karena nilainya lebih tinggi. Ini menunjukkan bahwa model ANN memiliki performa yang solid, meskipun mungkin masih ada ruang untuk perbaikan.
- 3) Perbandingan metode yang dilakukan terhadap metode ANN dan CART untuk melakukan klasifikasi terhadap penerima bantuan Program Keluarga Harapan (PKH) dapat disimpulkan bahwa metode *Artificial Neural Network* (ANN) lebih baik dalam memprediksi penerima bantuan Program Keluarga Harapan dalam melakukan kalsifikasi terhadap penerima bantuan Program Keluarga Harapan (PKH).

#### SARAN

Peneliti memiliki beberapa masukan untuk peneltian selanjutnya yang berhubungan dengan klasifikasi data :

- Mencoba membandingkan algoritma yang lain agar lebih akurat.
- Data yang digunakan lebih banyak.
- Menambahakan atribut yang lebih spesifik lagi dalam penelitian.

#### UCAPAN TERIMA KASIH

Terimakasih kepada keluarga dan rekan-rekan Universitas Aisyah Pringsewu dan bapak ibu Dosen IIB Darmajaya dan temen kelas IIB Darmajaya MTI Angkatan 27 yang telah memberikan banyak motivasi dan dukungan hingga selesainya penelitian ini.

#### REFERENSI

- [1] B. P. Statistik, "Persentase Penduduk Miskin September 2022 naik menjadi 9,57 persen," 2023. <https://www.bps.go.id/id/pressrelease/2023/01/16/2015/persentase-penduduk-miskin-september-2022-naik-menjadi-9-57-persen.html> (accessed Dec. 11, 2023).
- [2] Kominfo, "Program Bantuan Sosial Untuk Rakyat," 2018. [https://www.kominfo.go.id/index.php/content/detail/15708/program-bantuan-sosial-untuk-rakyat/0/artikel\\_gpr](https://www.kominfo.go.id/index.php/content/detail/15708/program-bantuan-sosial-untuk-rakyat/0/artikel_gpr) (accessed Dec. 11, 2023).
- [3] A. Sucipto, "Credit Prediction With Neural Network Algorithm," *Pros. Semin. Nas. Multi Disiplin Ilmu Call Pap. Unisbank*, vol. 978-979-36, no. 15, pp. 1-10, 2012.
- [4] N. Hadianto, H. B. Novitasari, and A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163-170, 2019, doi: 10.33480/pilar.v15i2.658.
- [5] R. D. Putri, S. Amin, M. Akli, and others, "Pengaplikasian Artificial Neural Network (ANN) dalam Memprediksi Curah Hujan Menggunakan Python," *SinarFe7*, vol. 4, no. 1, pp. 369-373, 2021.
- [6] M. Anwar Sadat, P. Pujiono, A. Pambudi, and S. Ibad, "Comparison of Algorithm Between Classification & Regression Trees and Support Vector Machine in Determining Student Acceptance in State Universities," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1589-1604, 2024, doi:

- 10.52436/1.jutif.2023.4.6.1565.
- [7] F. E. Pratiwi and I. Zain, "Klasifikasi pengangguran terbuka menggunakan CART (Classification and regression tree) di Provinsi Sulawesi Utara," *J. Sains dan Seni ITS*, vol. 3, no. 1, pp. D54–D59, 2014.
- [8] E. S. Novianto, E. Ervianto, and I. Yasri, "Studi Penerapan ANN (Artificial Neural Network) Untuk Menghilangkan Harmonisa Pada Gedung Pusat Komputer," *J. Online Mhs. Fak. Tek. Univ. Riau*, vol. 3, no. 2, pp. 1–6, 2016, [Online]. Available: <https://www.neliti.com/publications/184148/studi-penerapan-ann-artificial-neural-network-untuk-menghilangkan-harmonisa-pada#cite>
- [9] Kusriani and E. T. Luthfi, "Algoritma Data Mining.pdf," in *Algoritma Data Mining*, T. A. Prabawati, Ed., C.V Andi Offset (Penerbit Andi), 2009. [Online]. Available: <https://books.google.co.id/books?id=-Ojclag73O8C&printsec=frontcover&hl=id#v=onepage&q&f=false>
- [10] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [11] D. D. Novita, A. B. Sesunan, M. Telaumbanua, S. Triyono, and T. W. Saputra, "Identifikasi Jenis Kopi Menggunakan Sensor E-Nose Dengan Metode Pembelajaran Jaringan Syaraf Tiruan Backpropagation," *J. Ilm. Rekayasa Pertan. dan Biosist.*, vol. 9, no. 2, pp. 205–217, 2021, doi: 10.29303/jrpb.v9i2.241.
- [12] M. S. Hasibuan and R. A. Aziz, "Detection of learning styles with prior knowledge data using the SVM, K-NN and Naïve Bayes algorithms," *J. Infotel*, vol. 14, no. 3, pp. 209–213, 2022, doi: 10.20895/infotel.v14i3.788.
- [13] M. Rivki, A. M. Bachtiar, T. Informatika, F. Teknik, and U. K. Indonesia, *No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title*, no. 112.
- [14] F. T. Admojo and Y. I. Sulisty, "Analisis Performa Algoritma Stochastic Gradient Descent (SGD) Dalam Mengklasifikasi Tahu Berformalin," *Indones. J. Data Sci.*, vol. 3, no. 1, pp. 1–8, 2022, doi: 10.56705/ijodas.v3i1.42.
- [15] A. Dwi, B. Program, S. Ilmu, and K. Masyarakat, "Nomer 2 Edisi Desember 2015, hal," *Ilm. Kedokt.*, vol. 4, pp. 1–12.
- [16] V. R. Joseph, "Optimal ratio for data splitting," *Stat. Anal. Data Min.*, vol. 15, no. 4, pp. 531–538, 2022, doi: 10.1002/sam.11583.
- [17] E. L. Ika Pradita, K. P. Kartika, and F. Febrinita, "Payment Application Design Using Cart (Classification And Regression Tree) Method," *Procedia Eng. Life Sci.*, vol. 2, no. 1, 2021, doi: 10.21070/pels.v2i0.1165.
- [18] Indah Lestari, Dina Fitria, Syafriandi Syafriandi, and Admi Salma, "Comparison of the C5.0 Algorithm and the CART Algorithm in Stroke Classification," *UNP J. Stat. Data Sci.*, vol. 2, no. 1, pp. 90–98, 2024, doi: 10.24036/ujsds/vol2-iss1/144.
- [19] R. Nurpambudi, E. S. P. Wulandari, and R. A. Aziz, "Prediction of flood events in the city of Bandar Lampung using the artificial neural network," *J. Infotel*, vol. 15, no. 1, pp. 34–45, 2023, doi: 10.20895/infotel.v15i1.878.