

PERANCANGAN APLIKASI WEB PEMETAAN DATA WARGA BERBASIS ALGORITMA RANDOM FOREST UNTUK PREDIKSI KESEJAHTERAAN DI LINGKUNGAN RT 002 RW 012

Reza Pradana¹, Rengga Herdiansyah²

^{1,2}Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia, 15417
e-mail: ¹rezapradana472@gmail.com, ²dosen01101@unpam.ac.id

Abstract

Residents' well-being is a crucial factor determining the quality of life in a community. Planning targeted socioeconomic development programs is essential to achieving equitable distribution of resident welfare. However, diverse socioeconomic conditions and the manual process of mapping welfare data present challenges for decision-making at the neighborhood association/Rukun Tetangga (RT) level. This research aims to design a web application based on the random forest algorithm capable of predicting the level of welfare of residents in RT 002 RW 012. The research method involved collecting data from various welfare indicators, which were analyzed using the random forest algorithm to generate a classification of residents' welfare levels. The results showed that the random forest algorithm was able to predict welfare levels with high accuracy, with a confusion matrix test result of 90.33%. Other tests, including cross-validation, achieved 95.83% accuracy. This research also resulted in a system to simplify the management of resident data and socioeconomic conditions. This system can assist neighborhood association/Rukun Tetangga (RT) administrators in making more effective and data-driven decision-making.

Keywords: Web Application, Classification, Welfare Prediction, Random Forest.

Abstrak

Kesejahteraan warga merupakan faktor penting yang menentukan kualitas hidup di lingkungan masyarakat. Perencanaan program pembangunan sosial ekonomi yang tepat sasaran diperlukan untuk mewujudkan pemerataan kesejahteraan warga. Namun, kondisi sosial ekonomi yang beragam serta proses pemetaan data kesejahteraan yang masih dilakukan secara manual menjadi tantangan dalam pengambilan keputusan di tingkat Rukun Tetangga (RT). Penelitian ini bertujuan untuk merancang aplikasi web berbasis algoritma *random forest* yang mampu memprediksi tingkat kesejahteraan warga di lingkungan RT 002 RW 012. Metode penelitian melibatkan pengumpulan data dari berbagai indikator kesejahteraan yang dianalisis menggunakan algoritma *random forest* untuk menghasilkan klasifikasi tingkat kesejahteraan warga. Hasil penelitian menunjukkan bahwa algoritma *random forest* mampu memberikan prediksi tingkat kesejahteraan dengan akurasi tinggi dengan hasil pengujian *di confusion matrix* sebesar 90,33%. Dalam pengujian lainnya di *cross validation* sebesar 95,83%. Penelitian ini juga menghasilkan sistem untuk mempermudah pengelolaan data warga dan kondisi sosial ekonomi warga sehingga sistem yang dikembangkan dapat membantu pengurus Rukun Tetangga (RT) dalam proses pengambilan keputusan menjadi lebih efektif dan berbasis data.

Keywords: Aplikasi Web, Klasifikasi, Prediksi Kesejahteraan, *Random Forest*.

1. PENDAHULUAN

Kemiskinan masih menjadi permasalahan utama di Indonesia. Kemiskinan dapat diartikan sebagai ketidakmampuan masyarakat dalam mencukupi kebutuhan hidup minimum, khususnya terkait pendapatan dan pemenuhan kebutuhan pangan. Upaya penanganan masalah sosial yang berhubungan dengan kemiskinan dapat berhasil apabila memperoleh dukungan dari berbagai pihak. Penanggulangan kemiskinan melalui program bantuan sosial maupun pemberdayaan masyarakat diharapkan dapat mengatasi permasalahan tersebut [1].

Data Terpadu Kesejahteraan Sosial (DTKS) adalah sistem data elektronik yang berisi informasi sosial, ekonomi, dan demografi sekitar 99 juta penduduk dengan tingkat kesejahteraan terendah di Indonesia. Berdasarkan Peraturan Menteri Sosial Nomor 5 Tahun 2019 tentang Pengelolaan DTKS, istilah Basis Data Terpadu (BDT) resmi diganti menjadi DTKS. Isi DTKS mencakup Pemerlu Pelayanan Kesejahteraan Sosial (PPKS), Penerima Bantuan dan Pemberdayaan Sosial, serta Potensi dan Sumber Kesejahteraan Sosial (PSKS). DTKS merepresentasikan 40% penduduk dengan status kesejahteraan sosial terendah. Data tersebut harus diperbarui secara rutin agar selalu akurat dan sesuai dengan kondisi terbaru. Sesuai dengan lampiran UU 23/2004, pemerintah daerah berkewajiban melakukan pendataan serta pengelolaan data fakir miskin dan kelompok rentan di tingkat kabupaten/kota. Namun, menurut data Pusdatin Kemensos RI, masih terdapat kabupaten/kota di Indonesia yang belum sepenuhnya aktif dalam melakukan finalisasi DTKS [2].

Dalam pelaksanaannya, masih ditemukan berbagai tantangan di tingkat lokal, salah satunya di Kota Tanjungpinang, Provinsi Kepulauan Riau. Sejumlah laporan dari media lokal seperti Batamnews dan Tribun Batam menyebutkan bahwa bantuan kerap tidak tepat sasaran; terdapat warga tergolong mampu yang terdaftar sebagai penerima, sementara warga miskin yang layak menerima bantuan justru terabaikan karena tidak tercantum dalam DTKS. Permasalahan ini menunjukkan adanya kelemahan dalam proses verifikasi dan validasi data di lapangan. Hal ini menunjukkan adanya kesenjangan signifikan antara data administrasi dan realitas sosial di masyarakat. Ketimpangan ini diperparah oleh lemahnya mekanisme verifikasi di tingkat bawah, khususnya di level RT/RW dan kelurahan [3].

Fenomena serupa juga ditemukan di lingkungan RT 002/RW 012, yang terdiri dari 199 kepala keluarga dengan total 707 jiwa. Warga yang tidak tercantum dalam Data Terpadu Kesejahteraan Sosial (DTKS) merasa layak menerima bantuan sosial, kemudian melapor ke pengurus RT untuk meminta pendampingan agar dapat diajukan sebagai penerima bantuan. Pihak RT selanjutnya meneruskan pengajuan tersebut ke Dinas Sosial (Dinsos). Namun, permasalahan yang muncul adalah penilaian awal terhadap kelayakan warga yang dilakukan masih bersifat manual tanpa dukungan sistem yang terstruktur. Hal ini menyebabkan proses penentuan sering memakan waktu lama, rawan terjadi ketidaktepatan penilaian, dan bergantung pada subjektivitas pengurus RT. Kondisi yang terjadi berakibat pada pengajuan yang dikirim ke Dinsos tetap tidak lolos verifikasi karena tidak memenuhi kriteria kesejahteraan yang ditetapkan, sehingga prosesnya menjadi kurang efisien dan berpotensi sia-sia. Kondisi ini menunjukkan bahwa diperlukan suatu sistem yang mampu melakukan penilaian awal dan pemetaan tingkat kesejahteraan warga secara objektif, agar pengajuan bantuan dapat lebih tepat sasaran dan sesuai dengan kondisi riil di lapangan.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan berbasis sistem yang mampu membantu proses pendataan dan penilaian kesejahteraan warga secara otomatis dan terukur. Pendekatan ini dapat memanfaatkan metode *machine learning* dalam melakukan klasifikasi tingkat kesejahteraan berdasarkan indikator sosial-ekonomi. Sistem yang dibuat menyediakan fitur pendataan warga yang meliputi data keluarga dan data anggota keluarga kemudian untuk data indikator sosial-ekonomi berdasarkan indikator kelayakan penerima DTKS Keputusan Menteri Sosial Nomor 146/HUK/2013 tentang kriteria fakir miskin teregister. Metode *machine learning* yang digunakan untuk melakukan klasifikasi tingkat kesejahteraan adalah algoritma *random forest*.

Random forest merupakan salah satu metode *ensemble learning* yang terdiri dari kumpulan pohon keputusan yang digunakan untuk membuat prediksi yang lebih akurat dan stabil. Dalam klasifikasi menggunakan *random forest*, metode ini menggunakan pendekatan voting untuk mengambil keputusan mayoritas berdasarkan hasil dari pohon – pohon yang telah dibentuk. Pendekatan menggunakan *ensemble learning* memiliki beberapa keuntungan, yaitu dapat digunakan baik untuk klasifikasi maupun regresi, mampu mencapai tingkat akurasi yang tinggi, dan cocok untuk menganalisis *dataset* yang besar dengan banyak dimensi [4].

Melalui penelitian ini diharapkan dapat dihasilkan sistem pemetaan data kesejahteraan warga berbasis web dengan algoritma *random forest* yang mampu menjadi sistem pendamping pengurus RT dalam melakukan verifikasi dan pembaruan data kesejahteraan secara objektif, cepat, dan akurat sekaligus menjadi pendukung bagi sistem pusat yang telah ada. Dengan demikian, penyaluran bantuan sosial dapat menjadi lebih tepat sasaran dan selaras dengan kondisi nyata masyarakat.

2. METODE

Jenis dan Desain Penelitian

Penelitian menggunakan pendekatan kuantitatif dengan menguji algoritma *random forest* untuk memprediksi kondisi kesejahteraan warga berbasis *website*. Algoritma *random forest* yang dibentuk melalui tahapan pengumpulan data, *preprocessing*, pembagian data, pelatihan model, pengujian model, dan evaluasi [5]. Penelitian ini mengintegrasikan algoritma *random forest* berbasis *website* untuk membantu pemetaan kesejahteraan warga agar meningkatkan ketepatan sasaran penyaluran bantuan sosial.

Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan di wilayah RT 002 RW 012, Kelurahan Grogol Selatan, Kecamatan Kebayoran Lama, Jakarta Selatan. Pemilihan lokasi ini berdasarkan pada kondisi di wilayah RT 002 RW 012 yang masih menggunakan metode konvensional dalam menentukan tingkat kesejahteraan warga. Waktu penelitian dilakukan selama 6 bulan, meliputi tahap pengumpulan data, analisis kebutuhan sistem, perancangan, implementasi, dan pengujian sistem.

Populasi dan Sampel Penelitian

Populasi penelitian di wilayah RT 002 RW 012 berjumlah 199 kepala keluarga dengan total 707 jiwa. Sampel yang diperoleh dalam penelitian ini berjumlah 37 kepala keluarga dengan total 127 jiwa. Indikator data yang digunakan berupa kriteria fakir miskin berdasarkan Keputusan Menteri Sosial Nomor 146/HUK/2013 berdasarkan *website* Dinas Sosial Kota Tebing Tinggi dengan judul Syarat Kelayakan Penerima Manfaat Usulan DTKS dibuat pada tanggal 15 Juni 2021 dan diakses pada tanggal 19 Agustus 2025 indikator data sebagai syarat kelayakan penerima DTKS [6].

Teknik Pengumpulan Data

Pengumpulan data yang dibutuhkan menggunakan 3 metode: wawancara, kuesioner, dan studi dokumentasi. Wawancara dilakukan dengan pihak ketua RT 002 RW 012 untuk pengumpulan informasi. Kuesioner dilakukan dengan memperoleh informasi kepada responden melalui Google Form. Studi dokumentasi pengumpulan data warga dari Kartu Keluarga, Kartu Tanda Penduduk, data demografi catatan RT sebagai sumber data yang digunakan.

Analisis Data

Berikut hasil data yang diperoleh melalui kuesioner Google Form yang kemudian diolah ke dalam Microsoft Excel:

No	Nama keluarga	jenis tanggungan keluarga	jumlah tanggungan keluarga	jumlah tanggungan keluarga	jumlah tanggungan keluarga	jumlah tanggungan keluarga	jumlah tanggungan keluarga	jumlah tanggungan keluarga
1	11.000.000.001	1	1	1	1	1	1	1
2	11.000.000.002	1	1	1	1	1	1	1
3	11.000.000.003	1	1	1	1	1	1	1
4	11.000.000.004	1	1	1	1	1	1	1
5	11.000.000.005	1	1	1	1	1	1	1
6	11.000.000.006	1	1	1	1	1	1	1
7	11.000.000.007	1	1	1	1	1	1	1
8	11.000.000.008	1	1	1	1	1	1	1
9	11.000.000.009	1	1	1	1	1	1	1
10	11.000.000.010	1	1	1	1	1	1	1
11	11.000.000.011	1	1	1	1	1	1	1
12	11.000.000.012	1	1	1	1	1	1	1
13	11.000.000.013	1	1	1	1	1	1	1
14	11.000.000.014	1	1	1	1	1	1	1
15	11.000.000.015	1	1	1	1	1	1	1
16	11.000.000.016	1	1	1	1	1	1	1
17	11.000.000.017	1	1	1	1	1	1	1
18	11.000.000.018	1	1	1	1	1	1	1
19	11.000.000.019	1	1	1	1	1	1	1
20	11.000.000.020	1	1	1	1	1	1	1
21	11.000.000.021	1	1	1	1	1	1	1
22	11.000.000.022	1	1	1	1	1	1	1
23	11.000.000.023	1	1	1	1	1	1	1
24	11.000.000.024	1	1	1	1	1	1	1
25	11.000.000.025	1	1	1	1	1	1	1
26	11.000.000.026	1	1	1	1	1	1	1
27	11.000.000.027	1	1	1	1	1	1	1
28	11.000.000.028	1	1	1	1	1	1	1
29	11.000.000.029	1	1	1	1	1	1	1
30	11.000.000.030	1	1	1	1	1	1	1
31	11.000.000.031	1	1	1	1	1	1	1
32	11.000.000.032	1	1	1	1	1	1	1
33	11.000.000.033	1	1	1	1	1	1	1
34	11.000.000.034	1	1	1	1	1	1	1
35	11.000.000.035	1	1	1	1	1	1	1
36	11.000.000.036	1	1	1	1	1	1	1
37	11.000.000.037	1	1	1	1	1	1	1

Gambar 1. Hasil data kuesioner

Data yang dikumpulkan kemudian diubah bentuknya dengan *label encoding*. *Label encoding* merupakan teknik dalam pemrosesan data yang digunakan untuk mengubah nilai-nilai kategorikal menjadi nilai numerik. Berikut hasil perubahan data dengan *label encoding*:

No.	Penghasilan	Pendidikan	Dinding rumah	Lantai Rumah	Atap Rumah	Penerangan	Luas Lantai Rumah	Sumber Air Minum	Pengeluaran Makanan	Akses Kesehatan	Pakaian	Kelas/label
1	1	5	4	3	3	4	1	4	2	2	3	1
2	0	2	3	3	3	4	0	4	2	2	3	0
3	1	0	3	3	3	3	1	4	3	2	3	1
4	0	3	4	3	3	4	0	4	0	2	3	0
5	1	3	4	3	3	4	1	2	1	2	3	1
6	1	4	2	2	2	4	1	2	0	1	2	0
7	1	5	2	2	2	4	1	2	1	2	3	1
8	0	0	2	3	3	3	1	2	2	3	3	0
9	0	5	2	1	1	4	1	2	2	1	2	0
10	1	3	2	1	2	4	0	2	0	1	2	0
11	1	4	4	3	3	4	0	4	2	3	4	1
12	0	5	4	3	3	3	1	3	1	2	3	0
13	0	3	4	3	3	4	0	4	2	2	3	0
14	1	2	4	3	3	4	0	4	1	2	3	1
15	1	5	2	1	3	4	1	4	2	2	3	1
16	1	4	4	3	3	4	0	4	1	2	3	1
17	1	4	4	3	3	4	1	2	1	2	3	1
18	0	4	2	3	3	3	1	4	2	2	3	0
19	0	5	4	3	3	4	1	4	2	2	2	0
20	1	3	3	3	3	4	0	4	1	3	3	1
21	1	4	4	3	3	4	1	4	1	2	3	1
22	0	1	4	3	3	4	1	2	2	2	3	0
23	1	4	4	3	3	4	1	4	1	2	3	1
24	0	5	4	3	3	4	1	2	2	2	2	0
25	1	2	4	3	3	4	1	4	1	2	3	1
26	1	2	4	3	3	4	1	4	1	2	3	1
27	1	3	4	3	3	4	1	4	2	2	3	1
28	0	1	4	3	3	4	0	4	2	2	3	0
29	0	0	4	3	3	4	0	4	2	2	3	0
30	0	1	4	3	3	4	0	4	2	2	3	0
31	0	2	4	3	3	4	0	4	2	2	3	0
32	0	0	4	3	3	4	1	2	2	2	3	0
33	1	0	4	3	3	4	1	4	2	2	4	1
34	0	5	4	3	3	4	1	4	3	2	2	0
35	0	4	4	3	3	3	0	4	1	2	3	0
36	0	2	4	3	3	4	0	4	2	1	2	0
37	0	0	4	3	3	4	1	2	2	2	3	0

Gambar 2. Dataset hasil perubahan dengan *label encoding*

Algoritma Random Forest

Berikut merupakan rumus – rumus perhitungan yang digunakan dalam pembentukan pohon keputusan di algoritma random forest [7]:

Rumus Entropy:

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

Keterangan:

- S : himpunan data (dataset) yang sedang dianalisis.
- n : jumlah kelas (kategori/label) dalam dataset.
- p_i : probabilitas (proporsi) dari kelas ke-i dalam himpunan S.

Rumus Information Gain (IG):

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

Keterangan:

- S : himpunan data keseluruhan.
- A : atribut/fitur yang diuji.
- Entropy(S) : tingkat ketidakpastian atau *impurity* dari himpunan data S.
- S_i : subset dari S yang terbentuk ketika data dipartisi berdasarkan nilai atribut A.
- $\frac{|S_i|}{|S|}$: proporsi jumlah data di subset S_i terhadap total data S.
- Entropy(S_i) : entropi dari subset data S_i

Rumus Split Info:

$$SplitInfo(S, A) = \sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (3)$$

Keterangan:

- S : himpunan data keseluruhan.
- A : atribut yang sedang diuji.
- c : jumlah partisi/himpunan hasil pembagian data berdasarkan atribut A.
- S_i : jumlah data dalam partisi ke-i.
- $\frac{|S_i|}{|S|}$: proporsi data di partisi ke-i terhadap total data.

Rumus *Gain Ratio*:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (4)$$

Keterangan:

S : himpunan *data training*.

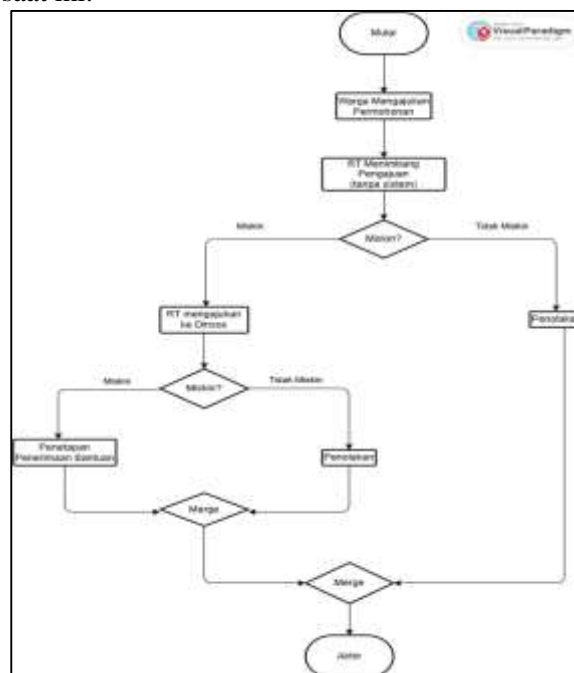
A : atribut (fitur) yang sedang diuji.

$Gain(S,A)$: nilai *Information Gain* dari atribut A terhadap data S.

$SplitInfo(S,A)$: ukuran seberapa besar atribut A membagi data S ke dalam *subset-subset*.

Analisis Sistem Berjalan

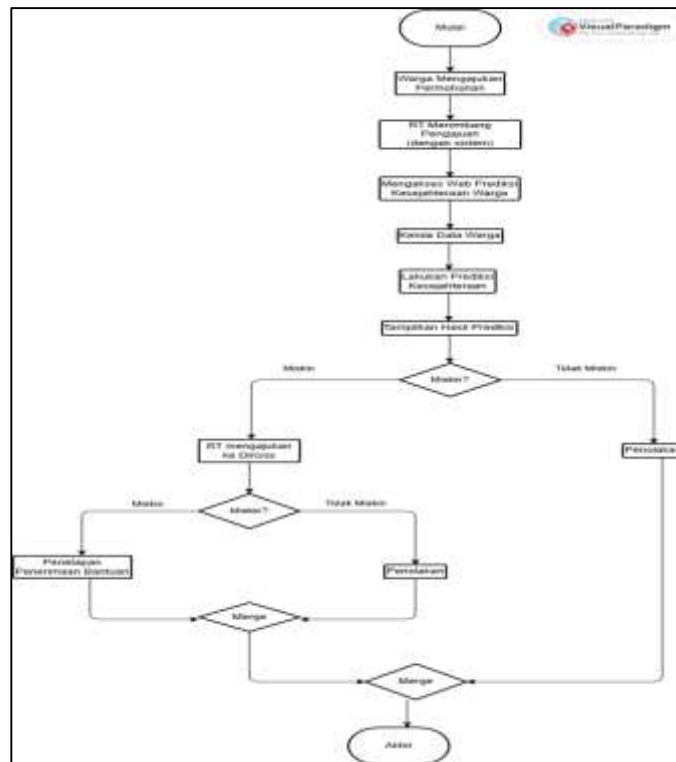
Analisis sistem berjalan merupakan alur yang saat ini masih digunakan oleh wilayah RT 002 RW 012 dalam penentuan tingkat kesejahteraan warga. Analisis ini dilakukan untuk memahami permasalahan yang terdapat dalam sistem saat ini.



Gambar 3. Analisis Sistem Berjalan

Analisis Sistem Usulan

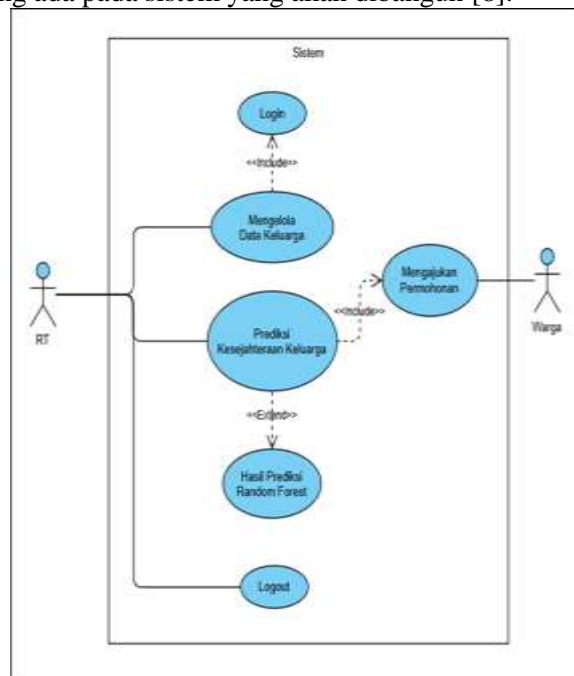
Analisis sistem usulan merupakan rancangan untuk mengatasi permasalahan di analisis sistem berjalan. Pada analisis ini dirumuskan solusi teknologi yang mampu memberikan prediksi akurat berdasarkan variabel-variabel kesejahteraan warga.



Gambar 4. Analisis Sistem Usulan

Use Case Diagram

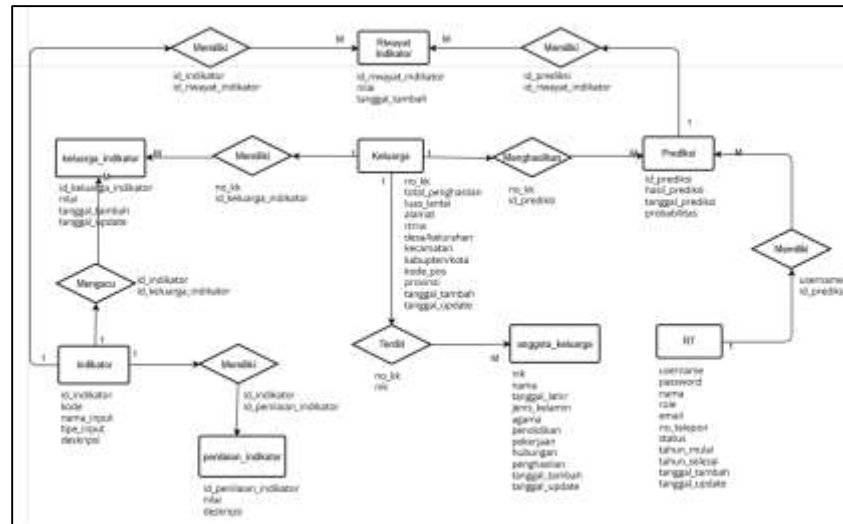
Use case diagram dipakai untuk menjelaskan perancangan sistem kepada user dan melakukan perancangan semua fitur yang ada pada sistem yang akan dibangun [8].



Gambar 5. Use Case Diagram

Entity Relationship Diagram (ERD)

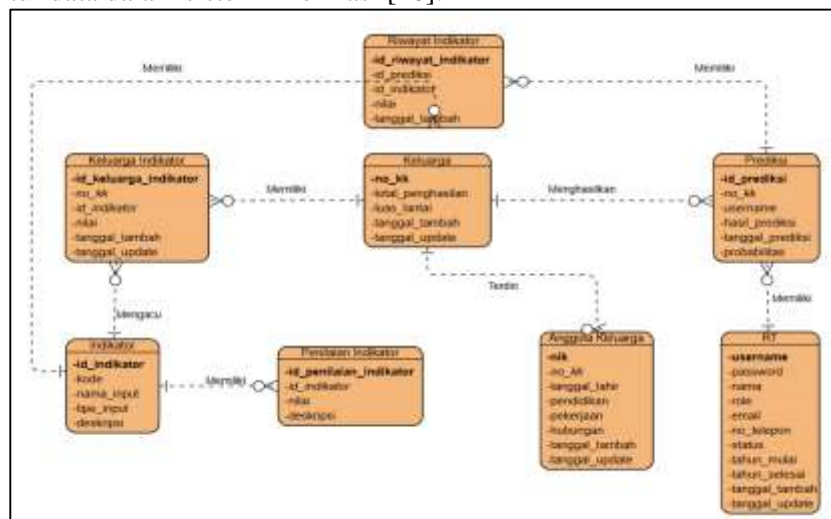
Entity Relationship Diagram (ERD) adalah model teknik pendekatan yang menyatakan atau menggambarkan hubungan suatu model. Di dalam hubungan ini tersebut dinyatakan yang utama dari ERD adalah menunjukkan objek data (Entity) dan hubungan (Relationship), yang ada pada Entity berikutnya [9].



Gambar 6. Entity Relationship Diagram (ERD)

Logical Record Structure (LRS)

Logical Record Structure (LRS) digunakan untuk menentukan struktur record pada tabel yang terbentuk melalui relasi antara entitas. Logical Record Structure (LRS) digunakan sebagai model dan representasi struktur data dalam sistem informasi [10].



Gambar 7. Logical Record Structure (LRS)

3. HASIL

Spesifikasi Perangkat Lunak

Spesifikasi perangkat lunak yang digunakan dalam pembuatan aplikasi yang adalah sebagai berikut:

1. Microsoft Windows 10
2. phpMyAdmin version 5.2.0
3. PHP 8.1.6
4. Python 3.10.2
5. Laravel Framework 10.48.29
6. 12 tabel di database dengan total size awal 318 Kib.

Spesifikasi Perangkat Keras

Berdasarkan spesifikasi pembuatan di atas, maka perangkat keras yang diusulkan untuk menjalankan program sebagai berikut:

1. CPU Dual-core 1.5 Ghz
2. RAM 4 GB
3. Storage 128 GB
4. Smartphone RAM 2 GB.

Perhitungan Manual Dan Sistem Algoritma *Random Forest*

Berikut hasil perhitungan manual algoritma *random forest* dengan pengambilan 19 data secara *bootstrap sampling*:

Node 1	Jumlah	Miskin	Tidak Miskin	Entropy	Gain	Split Info	Gain Ratio
Total	19	9	10	0,998			
Air Minum					0,0904	1,0200037	0,0886
2 (Air mesin bor pompa)	4	1	3	0,8113		0,4721963	
3 (Air PDAM)	1	1	0	0		0,2235751	
4 (Air isi ulang/kemasan)	14	7	7	1		0,3242323	
Penghasilan					0,998	0,998	1
0 (<= GKM Jakarta)	9	9	0	0		0,5106	
1 (> GKM Jakarta)	10	0	10	0		0,4874	

Gambar 8. Tabel Perhitungan *Entropy*, *Gain*, *Split Info*, *Gain Ratio*

Berikut hasil perhitungan sistem algoritma *random forest* dengan bahasa pemrograman Python dan dataset yang sama dengan perhitungan manual:

```

PS C:\Users\USER\OneDrive\Desktop\KULIAH\SKRIPSI\TRAINING_PYTHON> python maindot.py

Total Entropy: 0.9980

Fitur: Penghasilan
Information Gain: 0.9980
Split Info: 0.9980
Gain Ratio: 1.0000

Fitur: Air Minum
Information Gain: 0.0904
Split Info: 1.0215
Gain Ratio: 0.0885
    
```

Gambar 9. Hasil perhitungan secara sistem

Evaluasi Model *Confusion Matrix*

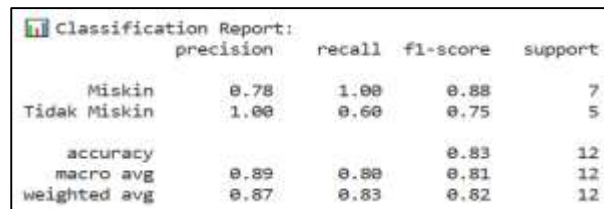
Merupakan metode yang digunakan untuk mengevaluasi kinerja suatu algoritma klasifikasi. Metode ini menampilkan hasil klasifikasi dalam empat kategori utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* [11].

Berikut skenario pengujian model *confusion matrix* pertama dengan pembagian data untuk 80% data latih dan 20% data uji. Berikut hasil *confusion matrix* yang didapatkan:

Classification Report:				
	precision	recall	f1-score	support
Miskin	1.00	1.00	1.00	4
Tidak Miskin	1.00	1.00	1.00	4
accuracy			1.00	8
macro avg	1.00	1.00	1.00	8
weighted avg	1.00	1.00	1.00	8

Gambar 10. Hasil evaluasi *confusion matrix* 1

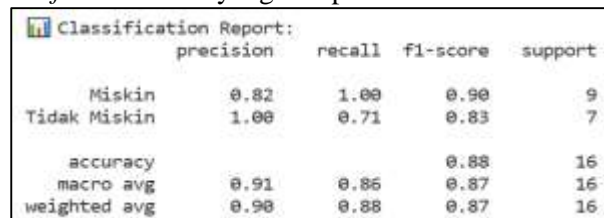
Skenario pengujian model *confusion matrix* kedua dengan pembagian data untuk 70% data latih dan 30% data uji. Berikut hasil *confusion matrix* yang didapatkan:



Classification Report:				
	precision	recall	f1-score	support
Miskin	0.78	1.00	0.88	7
Tidak Miskin	1.00	0.60	0.75	5
accuracy			0.83	12
macro avg	0.89	0.80	0.81	12
weighted avg	0.87	0.83	0.82	12

Gambar 11. Hasil evaluasi *confusion matrix 2*

Skenario pengujian model *confusion matrix* ketiga dengan pembagian data untuk 60% data latih dan 40% data uji. Berikut hasil *confusion matrix* yang didapatkan:



Classification Report:				
	precision	recall	f1-score	support
Miskin	0.82	1.00	0.90	9
Tidak Miskin	1.00	0.71	0.83	7
accuracy			0.88	16
macro avg	0.91	0.86	0.87	16
weighted avg	0.90	0.88	0.87	16

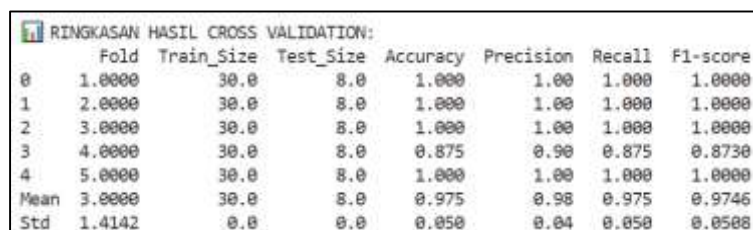
Gambar 12. Hasil evaluasi *confusion matrix 3*

Pada pengujian pertama model menunjukkan kinerja yang sangat baik dengan mendapatkan akurasi sebesar 100%. Pengujian kedua model mendapatkan akurasi sebesar 83% dan pengujian ketiga model mendapatkan akurasi sebesar 88%. Berdasarkan ketiga pengujian yang telah dilakukan menunjukkan bahwa hasil akurasi rata – rata yang didapatkan mencapai 90,33%.

Cross Validation

Cross validation adalah metode yang membagi *dataset* ke dalam beberapa *subset*, kemudian secara bergantian menggunakan sebagian *subset* untuk proses pelatihan (*training*) dan sisanya untuk pengujian (*testing*) [12].

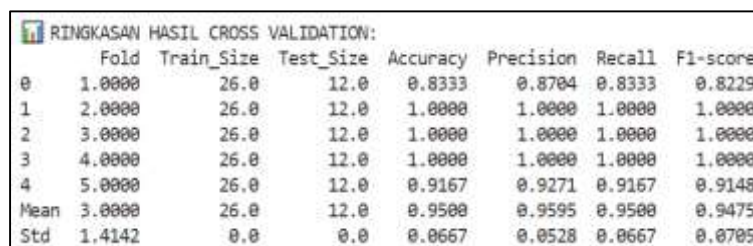
Berikut pengujian model *cross validation* pertama dengan pembagian data 80% data latih dan 20% data uji:



RINGKASAN HASIL CROSS VALIDATION:							
	Fold	Train_Size	Test_Size	Accuracy	Precision	Recall	F1-score
0	1.0000	30.0	8.0	1.000	1.00	1.000	1.0000
1	2.0000	30.0	8.0	1.000	1.00	1.000	1.0000
2	3.0000	30.0	8.0	1.000	1.00	1.000	1.0000
3	4.0000	30.0	8.0	0.875	0.90	0.875	0.8730
4	5.0000	30.0	8.0	1.000	1.00	1.000	1.0000
Mean	3.0000	30.0	8.0	0.975	0.98	0.975	0.9746
Std	1.4142	0.0	0.0	0.050	0.04	0.050	0.0508

Gambar 13. Pengujian *cross validation 1*

Berikut pengujian model *cross validation* kedua dengan pembagian data 70% data latih dan 30% data uji:



RINGKASAN HASIL CROSS VALIDATION:							
	Fold	Train_Size	Test_Size	Accuracy	Precision	Recall	F1-score
0	1.0000	26.0	12.0	0.8333	0.8704	0.8333	0.8229
1	2.0000	26.0	12.0	1.0000	1.0000	1.0000	1.0000
2	3.0000	26.0	12.0	1.0000	1.0000	1.0000	1.0000
3	4.0000	26.0	12.0	1.0000	1.0000	1.0000	1.0000
4	5.0000	26.0	12.0	0.9167	0.9271	0.9167	0.9148
Mean	3.0000	26.0	12.0	0.9500	0.9595	0.9500	0.9475
Std	1.4142	0.0	0.0	0.0667	0.0528	0.0667	0.0705

Gambar 14. Pengujian *cross validation 2*

Berikut pengujian model *cross validation* ketiga dengan pembagian data 60% data latih dan 40% data uji:

RINGKASAN HASIL CROSS VALIDATION:							
	Fold	Train_Size	Test_Size	Accuracy	Precision	Recall	F1-score
0	1.0000	22.0	16.0	0.8750	0.8977	0.8750	0.8708
1	2.0000	22.0	16.0	1.0000	1.0000	1.0000	1.0000
2	3.0000	22.0	16.0	1.0000	1.0000	1.0000	1.0000
3	4.0000	22.0	16.0	1.0000	1.0000	1.0000	1.0000
4	5.0000	22.0	16.0	0.8750	0.9028	0.8750	0.8750
Mean	3.0000	22.0	16.0	0.9500	0.9601	0.9500	0.9492
Std	1.4142	0.0	0.0	0.0612	0.0489	0.0612	0.0623

Gambar 15. Pengujian *cross validation* 3

Pada pengujian *cross validation* pertama model mendapatkan akurasi sebesar 97,5%. Pengujian kedua *cross validation* mendapatkan akurasi sebesar 95% dan di pengujian ketiga mendapatkan akurasi sebesar 95%. Berdasarkan ketiga pengujian *cross validation* yang telah dilakukan model mendapatkan hasil akurasi rata – rata sebesar 95,83%.

Implementasi Program

Login

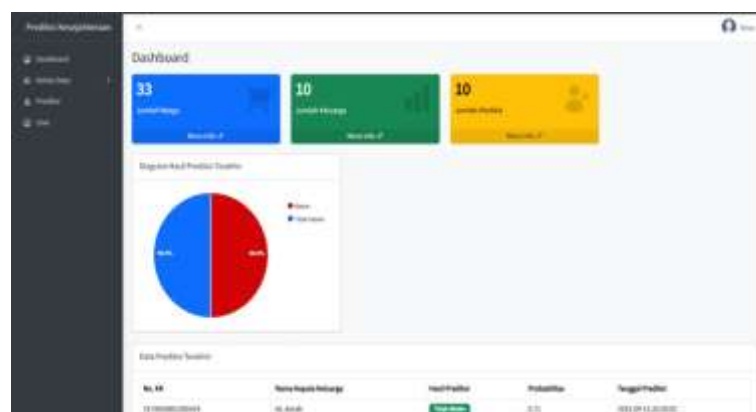
Halaman *login* berisi *input username*, *password*, dan *remember token*. *Username* merupakan data unik yang dimiliki setiap pengguna, sedangkan *password* adalah kata sandi yang berisi kombinasi huruf, simbol dan angka. *Remember token* berisi *token* untuk memvalidasi *user* agar setiap kembali ke *website* pengguna tidak perlu *login* kembali.



Gambar 16. Halaman *login*

Dashboard

Halaman *dashboard* adalah halaman pertama yang ditampilkan setelah pengguna berhasil *login*. Pada halaman ini terdapat informasi data seperti jumlah warga, jumlah keluarga, jumlah prediksi, diagram perbandingan hasil prediksi, dan data prediksi terakhir.



Gambar 17. Halaman *dashboard*

Kelola Data Keluarga

Halaman kelola data keluarga berisi detail data keluarga seperti nomor kartu keluarga, nama kepala keluarga, jumlah anggota keluarga, total penghasilan dan aksi. Fitur ini digunakan oleh pihak RT untuk menambahkan data keluarga ketika warga mengajukan permohonan bantuan. Dalam prosesnya, RT menerima pengajuan dari warga dan menginput data keluarga berdasarkan informasi serta kondisi aktual yang dimiliki oleh warga tersebut.

No. KK	Nama Kepala Keluarga	Status Indikator	Jumlah/Prediksi	Aktif
1179010010001744	Muti (Karyati Wt Sari)	11.0001.1.1	0	<input checked="" type="checkbox"/>
1179010010000000	Mulyani	11.0001.1.1	0	<input checked="" type="checkbox"/>
1179010010000000	Chandani	11.0001.1.1	0	<input checked="" type="checkbox"/>
1179010010000000	Ahmad	11.0001.1.1	0	<input checked="" type="checkbox"/>
1179010010004700	Sulisti	11.0001.1.1	0	<input checked="" type="checkbox"/>
1179010010000000	Ruswati	11.0001.1.1	0	<input checked="" type="checkbox"/>
1179010010000000	Yulita	11.0001.1.1	0	<input checked="" type="checkbox"/>
1179010010000000	Sarif Wahyuni Yuliani	11.0001.1.1	0	<input checked="" type="checkbox"/>
1179010010000000	Mahmuti Diani	11.0001.1.1	0	<input checked="" type="checkbox"/>

Gambar 21. Halaman prediksi

Pengujian Sistem Sistem *Black Box*

Black box testing adalah jenis testing yang memperlakukan perangkat lunak tanpa kinerja internal yang diketahui sehingga pengujian melihat perangkat lunak seperti *black box* yang tidak penting untuk melihat isinya, tetapi cukup tunduk pada proses testing luarnya [13]. Pengujian terdiri dari 16 skenario dari fitur *login*, halaman keluarga, halaman keluarga indikator, halaman indikator, dan halaman prediksi. Pengujian yang dilakukan mendapatkan keberhasilan kesesuaian sebesar 100%.

Sistem *White Box*

White box disebut dengan pengujian kotak kaca atau pengujian terstruktur dimana pengujian dikembangkan berdasarkan kode – kode sistem. Pengujian *white box* memiliki pengetahuan tentang kode dengan menggunakan parameter yang sesuai [14]. Pengujian dilakukan di *class* fungsi prediksi yang mendapatkan 3 jalur dari perhitungan *cyclomatic complexity* berdasarkan *flowgraph* yang terbentuk. Pengujian hasil kesesuaian dari 3 jalur mendapatkan skor 100%.

4. PEMBAHASAN

Hasil perhitungan algoritma *random forest* yang dilakukan pada perhitungan manual dan sistem menunjukkan konsistensi nilai yang mendekati sama. Hal ini menunjukkan bahwa perhitungan manual dapat digunakan untuk memvalidasi kebenaran algoritma yang dijalankan oleh sistem.

Evaluasi model yang dilakukan dengan pengujian *confusion matrix* dan *cross validation* menunjukkan hasil yang berbeda – beda sesuai dengan kondisi yang di uji. Pembagian data latih dan data uji menunjukkan pengaruh pada hasil pengujian yang dilakukan di model.

Pada pengujian model tingkat akurasi pada *confusion matrix* mendapatkan rata – rata akurasi sebesar 90,33% dan pada pengujian *cross validation* mendapatkan rata – rata akurasi sebesar 95,83%. Hasil ini menunjukkan perhitungan yang dilakukan pada penelitian kali ini sejalan dengan hasil penelitian terdahulu dimana pada penelitian “Perbandingan Tingkat Akurasi Algoritma *Decision Tree* Dan *Random Forest* Dalam Mengklasifikasi Penerima Bantuan Sosial Bpnt Di Desa Slangit” evaluasi untuk *random forest* mendapatkan tingkat akurasi sebesar 97,86% [15]. Pada penelitian lainnya “Analisa Prediksi Kesejahteraan Masyarakat Nelayan Lombok Timur Menggunakan Algoritma *Random Forest*” pada pengujian *confusion matrix* mendapatkan nilai akurasi sebesar 93,37% [11].

5. KESIMPULAN

Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada penelitian mengenai “Perancangan Aplikasi Web Pemetaan Data Warga Berbasis Algoritma *Random Forest* Untuk Prediksi Kesejahteraan Di Lingkungan RT 002 RW 012” ini dapat diambil kesimpulan sebagai berikut:

1. Algoritma *random forest* berhasil diterapkan untuk memprediksi tingkat kesejahteraan warga RT 002 RW 012 dengan hasil pengujian di *confusion matrix* memberikan tingkat akurasi rata – rata sebesar 90,33%. Dalam pengujian lainnya di *cross validation* memberikan tingkat akurasi rata – rata sebesar 95,83% dan di pengujian perbandingan hasil aktual dan prediksi sistem mendapatkan akurasi sebesar 89%. Hal ini menunjukkan bahwa algoritma *random forest* mampu melakukan prediksi dengan efektif dan mengatasi masalah subjektivitas pada metode konvensional.
2. Penelitian ini berhasil merancang sebuah sistem berbasis *website* sebagai sistem pendamping yang menyediakan fitur prediksi kesejahteraan warga, pengelolaan data warga, pengelolaan indikator warga, serta visualisasi data yang memudahkan proses verifikasi di tingkat RT. Dengan adanya sistem ini, diharapkan penyaluran bantuan sosial dapat menjadi lebih tepat sasaran, karena hasil prediksi dapat digunakan sebagai bahan pendukung keputusan sebelum data warga diusulkan ke sistem pusat.

Saran

Berdasarkan penelitian yang telah dilakukan, berikut saran – saran yang dapat peneliti ajukan:

1. Bagi pengelola RT 002 RW 012, diharapkan sistem yang telah dirancang dapat dipergunakan sebagai alat bantu verifikasi data DTKS di tingkat RT untuk memastikan penyaluran bantuan sosial lebih tepat sasaran.
2. Bagi peneliti selanjutnya, diharapkan dapat menggunakan *dataset* yang lebih besar dan dapat menambahkan variabel-variabel indikator baru yang lebih beragam.

6. UCAPAN TERIMAKASIH

Ucapan terima kasih penulis kepada pihak yang membantu ataupun memberikan dukungan terkait dengan penelitian yang dilakukan kepada:

1. Bapak Rengga Herdiansyah, A.Md., S.Kom., M.Kom., selaku pembimbing skripsi pada Program Studi Teknik Informatika di Universitas Pamulang.
2. Ibu Nurasih, S.Pd.I, selaku Ketua RT 002 RW 012.
3. Bapak dan Ibu tercinta yang telah mendukung penulis baik spirit maupun materi.
4. Saudara dan sahabat-sahabatku.

DAFTAR PUSTAKA

- [1] T. Sulistyarini, "Implementasi Kebijakan Peraturan Bupati Ponorogo Nomor 60 Tahun 2020 Tentang Percepatan Penanggulangan Kemiskinan Berbasis Data Terpadu Kesejahteraan Sosial (DtkS) Di Kecamatan Ngebel Kabupaten Ponorogo Propinsi Jawa Timur," *Ji@P*, vol. 10, no. 2, pp. 139–155, 2023, doi: 10.33061/jp.v10i2.5911.
- [2] R. Dwiarto, "Inovasi Penyaluran Jaminan Sosial Tepat Sasaran Melalui Kebijakan Pengelolaan Anggaran Data Terpadu Kesejahteraan Sosial (DTKS) Dan Pemanfaatan Aplikasi " Cek Bansos ", " *Pros. Semin. Nas. UNIMUS*, pp. 204–215, 2023.
- [3] R. Nursuhayla, Melati, Nursyidah, and W. Pangestoeti, "JMEB: Jurnal Management, Ekonomi Dan Bisnis," vol. 3, no. 1, pp. 11–16, 2025.
- [4] N. N. Sholihah and A. Hermawan, "Implementation of Random Forest and Smote Methods for Economic Status Classification in Cirebon City," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1387–1397, 2023, doi: 10.52436/1.jutif.2023.4.6.1135.
- [5] E. R. B. Sebayang, Y. H. Chrisnanto, and Melina, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," *IJESPG J.*, vol. 1, no. 3, pp. 237–253, 2023.
- [6] Dinas Sosial Kota Tebing Tinggi, "Syarat Kelayakan Penerima Manfaat Usulan DTKS." Accessed: Aug. 19, 2025. [Online]. Available: <https://dinsos.tebingtinggikota.go.id/news/syarat-kelayakan-penerima-manfaat-usulan-dtkS#:~:text=1.,mempunyai kemampuan memenuhi kebutuhan dasar.&text=c. Mempunyai dinding rumah terbuat,listrik atau listrik tanpa meteran.&text=h. Mempunyai sumber air minu>
- [7] N. Ismail, "Algoritma Random Forest," pp. 1–15, 2023.
- [8] F. N. Hasanah, *Buku Ajar Rekayasa Perangkat Lunak*. 2020. doi: 10.21070/2020/978-623-6833-89-6.
- [9] L. A. Budiman, A. R. Hakim, D. Pratama, I. E. Tsalatsah, and P. Rosyani, "Perancangan Sistem Informasi Nilai Siswa Berbasis Website," *J. Kreat. Mhs. Inform.*, vol. 2, no. 1, pp. 1–6, 2021.

- [10] R. Indah Melyani, R. Rosita, and S. Aji, “Pengembangan Sistem Informasi Penggajian Berbasis Web Menggunakan Framework Laravel dengan Metode Agile Software Development,” *J. Sist. Inf. Akunt.*, vol. 3, no. 1, pp. 31–36, 2023, doi: 10.31294/jasika.v3i01.2195.
- [11] A. Sandi, K. Kusrini, and K. Kusnawi, “Analisa Prediksi Kesejahteraan Masyarakat Nelayan Lombok Timur Menggunakan Algoritma Random Forest,” *Infotek J. Inform. dan Teknol.*, vol. 6, no. 2, pp. 238–248, 2023, doi: 10.29408/jit.v6i2.10104.
- [12] D. Haidar, B. Irawan, and A. Bahtiar, “Penerapan Deep Learning Model Random Forest Untuk Prediksi Penerima Bantuan Program Keluarga Harapan (Pkh),” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3564–3571, 2024, doi: 10.36040/jati.v7i6.8250.
- [13] D. Wintana, D. Pribadi, and M. Y. Nurhadi, “Analisis Perbandingan Efektifitas White-Box Testing dan Black-Box Testing,” *J. Larik Ldng. Artik. Ilmu Komput.*, vol. 2, no. 1, pp. 8–16, 2022, doi: 10.31294/larik.v2i1.1382.
- [14] A. Praniffa *et al.*, “Pengujian Black Box Dan White Box Sistem Informasi Parkir Berbasis Web Black Box And White Box Testing Of Web-Based Parking Information System,” *J. Test. Dan Implementasi Sist. Inf.*, vol. 1, no. 1, pp. 1–16, 2023.
- [15] A. Aldiyansyah, A. Irma Purnamasari, and I. Ali, “Perbandingan Tingkat Akurasi Algoritma Decision Tree Dan Random Forest Dalam Mengklasifikasi Penerima Bantuan Sosial Bpnt Di Desa Slangit,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 127–132, 2024, doi: 10.36040/jati.v8i1.8290.