
SISTEM INFORMASI KELAYAKAN BANTUAN SOSIAL MENGUNAKAN ALGORITMA *RANDOM FOREST* (STUDI KASUS: KECAMATAN SOSA)

Wahyu Fadillah Hasibuan¹, Samsudin², Fathiya Hasyifah Sibarani³
Universitas Islam Negeri Sumatera Utara, Medan
e-mail: ¹fadillahhasibuan89@gmail.com, ²samsudin@uinsu.ac.id,
³fathiyahasyifahsibarani@uinsu.ac.id

Abstract: *The background of this research stems from the need for a system capable of determining the eligibility of aid recipients more objectively, quickly, and accurately compared to manual methods that are prone to subjectivity. From a total of 2,304 community data, this study determined N Samples (Training Data) of 300 data manually labeled by the researcher as Ground Truth (answer key), consisting of 78 labels "Eligible" and 222 labels "Not Eligible". This data then went through a pre-processing stage (Feature Encoding) to convert categorical attributes to numeric, rather than score weighting. The Random Forest model was drilled using these 300 N Samples before being used to predict the remaining 2004 test data. The results of the model classification on all 2,304 data showed that 395 communities were predicted as "Eligible" and 1,909 communities were predicted as "Not Eligible". The model performance evaluation, calculated from 20% of the internal test data (60 data points from 300 N Samples), yielded 88.33% Accuracy, 90.91% Precision, 62.50% Recall, and 74.07% F1-Score. These results indicate that the model has a high level of precision (effectively avoiding mistargeting), although it trades off with lower recall (still missing some eligible cases). Thus, this system is expected to assist the Social Services Agency in improving the accuracy, transparency, and efficiency of aid recipient determination in the field.*

Keywords: *Random Forest, Social Assistance, Classification, Data Mining, Machine Learning*

Abstrak: Latar belakang penelitian ini berangkat dari kebutuhan akan sistem yang mampu menentukan kelayakan penerima bantuan secara lebih objektif, cepat, dan akurat dibandingkan metode manual yang rawan subjektif. Dari total 2.304 data masyarakat, penelitian ini menetapkan N Sampel (Data Latih) sebanyak 300 data yang dilabeli secara manual oleh peneliti sebagai *Ground Truth* (kunci jawaban), yang terdiri dari 78 label "Layak" dan 222 label "Tidak Layak". Data ini kemudian melalui tahap *pre-processing (Feature Encoding)* untuk mengubah atribut kategorikal menjadi numerik, bukan pembobotan skor. Model *Random Forest* dilatih menggunakan 300 N Sampel ini sebelum digunakan untuk memprediksi sisa 2004 data uji. Hasil klasifikasi model terhadap keseluruhan 2.304 data menunjukkan 395 masyarakat diprediksi "Layak" dan 1.909 masyarakat diprediksi "Tidak Layak". Evaluasi performa model, yang dihitung dari 20% data uji internal (60 data dari 300 N Sampel), menghasilkan Akurasi 88,33%, Presisi 90,91%, Recall 62,50%, dan F1-Score 74,07%. Hasil ini menunjukkan bahwa model memiliki tingkat presisi yang tinggi (efektif menghindari salah sasaran), meskipun dengan trade-off pada recall yang lebih rendah (masih melewatkan beberapa kasus yang layak). Dengan demikian, sistem ini diharapkan dapat membantu pihak Dinas Sosial dalam meningkatkan akurasi, transparansi, dan efisiensi penentuan penerima bantuan di lapangan.

Kata kunci: *Random Forest, Bantuan Sosial, Klasifikasi, Data Mining, Machine Learning*

PENDAHULUAN

Kecamatan Sosa adalah salah satu Kecamatan yang berada di Kabupaten Padang Lawas, Provinsi Sumatera Utara yang mempunyai RW (Rukun Warga) dari 16 desa lainnya. Oleh karena itu, penentuan kelompok masyarakat miskin yang berhak menerima bantuan sosial bersyarat dari pemerintah memerlukan analisis yang mendalam. Data kependudukan yang ada terlebih dahulu disediakan di lapangan kemudian dianalisis (Ayu Amanda & Mayana Br Tarigan, 2025). Meskipun analisis ini digunakan untuk memilih masyarakat miskin yang berhak menerima bantuan sosial bersyarat, namun kenyataannya masih ada sebagian masyarakat yang tidak memenuhi kriteria alokasi bantuan tersebut. Jika penerima bantuan sosial jelas terdata dengan baik, serta syarat kelayakan penerimaan bantuan sosial yang terdata dengan baik, maka akan mengurangi ketidaksinkronan penerima bantuan sosial (Kinanti, 2024) dan juga lebih memudahkan petugas di dalam pembagian bantuan sosial tersebut.

Bantuan sosial merupakan program penting yang bertujuan meningkatkan kesejahteraan masyarakat, khususnya bagi mereka yang berada dalam kondisi ekonomi rendah. Di Indonesia, distribusi bantuan sosial sering kali menemui berbagai kendala, seperti data penerima yang tidak akurat dalam penentuan kelayakan, dan ketidaktepatan sasaran. Hal ini mengakibatkan bantuan tidak merata dan penggolongan kemiskinan didasarkan pada suatu standar tertentu yaitu dengan membandingkan tingkat pendapatan orang atau keluarga dengan tingkat pendapatan yang diperlukan untuk memenuhi kebutuhan pokok minimum. Tantangan ini menjadi isu utama yang menghambat efektivitas program bantuan sosial. Program ini bertujuan untuk memutus mata rantai kemiskinan antar generasi, sehingga generasi berikutnya dapat keluar dari perangkap kemiskinan.

Solusi untuk mengatasi permasalahan yang telah diuraikan yaitu

dengan memanfaatkan teknologi dan algoritma *machine learning*, seperti *Random Forest*. *Random Forest* merupakan metode yang dapat meningkatkan hasil akurasi, karena metode ini digunakan untuk membangun pohon keputusan dengan mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan (Argadianata et al., 2025). Algoritma ini dikenal dengan keakuratannya dalam menangani dataset besar dan mampu menghasilkan klasifikasi yang optimal (Hidayat et al., 2025). Dengan memanfaatkan data masyarakat secara objektif dan akurat, algoritma ini diharapkan mampu membantu proses seleksi penerima bantuan sosial secara lebih transparan dan tepat sasaran (Br Sebayang et al., 2023). Klasifikasi merupakan proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Arisandi, 2023). Tujuan dari metode ini untuk mengklasifikasikan nilai suatu variabel yang tidak diketahui dari variabel lain yang telah ditentukan sebelumnya. Fungsi klasifikasi memudahkan untuk mengidentifikasi kelayakan penerima dan bukan penerima bantuan (Riansah et al., 2025).

Dalam penelitian sebelumnya, berbagai metode klasifikasi telah digunakan untuk menentukan kelayakan penerima bantuan sosial. Misalnya, dalam jurnal mahasiswa teknik informatika Vol. 8 No. 1, Februari 2024 Aisyah Fatmawati, DKK yang berjudul “Implementasi *Naive Bayes* Untuk Klasifikasi Kelayakan Penerima Bantuan Sosial” menjelaskan bahwa hasil penelitian serta pengujian dengan teknik data mining memakai algoritma klasifikasi *Naive Bayes* KDD, diperoleh nilai akurasi keseluruhan sebesar 86,36%. Dengan demikian, metode *Naive Bayes* memiliki unjuk kerja yang baik (Fatmawati et al., 2024). Jika dilihat dari penelitian terdahulu penulis mempunyai maksud memanfaatkan

algoritma *Random Forest* untuk menentukan klasifikasi bantuan sosial serta menambahkan implementasi *web* dengan memakai Bahasa pemrograman PHP dan *database Mysql* karena dipenelitian sebelumnya hanya menggunakan berbasis desktop.

METODE

Metode Penelitian

Metode penelitian yang digunakan adalah metode *Research and Development (R&D)*. Metode *Research and Development (R&D)* merupakan metode penelitian yang digunakan untuk menghasilkan produk tertentu dan menguji keefektifan produk tersebut (Fauzi & Samsudin, 2022; Samsudin et al., 2022) Penelitian ini dilaksanakan di Kecamatan Sosa. Langkah-langkah dalam penelitian R&D menurut Borg dan Gall seperti terlihat pada gambar berikut:



Gambar 1 Langkah-Langkah Metode Penelitian R&D

Teknik Pengumpulan Data

Berikut penjelasannya metode yang digunakan untuk pengumpulan data.

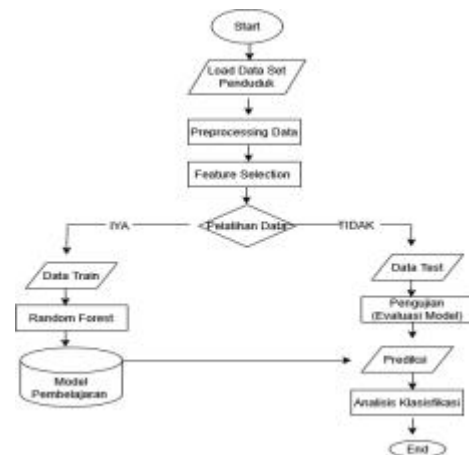
1. Observasi
 Observasi adalah teknik pengumpulan data dengan pengamatan langsung dan pencatatan secara teliti dan sistematis (Romdona et al., 2025). Peneliti melakukan observasi di Kecamatan Sosa Kabupaten Padang Lawas dengan maksud untuk mendapatkan informasi dan memperoleh gambaran mengenai sistem yang berjalan.
2. Wawancara
 Wawancara adalah teknik pengumpulan data yang digunakan untuk mendapatkan data yang dapat

diperlukan dalam penelitian ini dengan cara tanya jawab dengan orang yang dinilai memiliki kaitan erat dalam proses perancangan Aplikasi *Random Forest* Dalam Klasifikasi Kelayakan Masyarakat Mendapatkan Bantuan.

3. Studi Pustaka

Studi Pustaka yang dilakukan adalah teknik pengumpulan data dan informasi dalam menggali pengetahuan dan ilmu dari sumber sumber yang ada serta menjadi referensi dan pendukung dalam pembuatan dan penulisan penelitian ini.

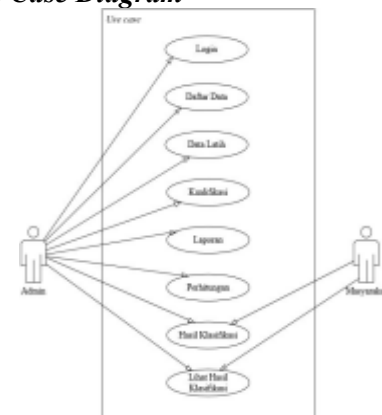
Flowchart Metode



Gambar 2 Flowchart Algoritma Random Forest

HASIL DAN PEMBAHASAN

Use Case Diagram



Gambar 3 Use Case Diagram

Pre-Processing Data

Tahap *pre-processing* data dilakukan untuk memastikan bahwa data yang digunakan dalam proses klasifikasi bersih, konsisten, dan valid. Validasi data pertama dilakukan sejak tahap input maupun impor data masyarakat, misalnya dengan menolak data yang memiliki NIK duplikat atau NIK yang tidak sesuai panjang digit resmi. Hal ini bertujuan agar data mentah yang masuk ke dalam sistem benar-benar valid.

Tahap *pre-processing* kedua dan yang paling krusial dalam metodologi ini adalah *Feature Encoding* (Encoding

Fitur). Proses ini dilakukan secara otomatis oleh skrip Python di *backend* (menggunakan *library Scikit-learn LabelEncoder*) saat "Proses Klasifikasi" dijalankan. Setiap nilai teks yang unik dalam sebuah fitur (kolom) akan ditransformasi dan diberi label angka yang unik. Sebagai contoh:

1. Pada fitur *jenis_kelamin*, nilai 'Laki-laki' dapat diubah menjadi 0 dan 'Perempuan' menjadi 1.
2. Pada fitur *dinding*, nilai 'Permanen' diubah menjadi 1, 'Semi Permanen' menjadi 2, dan 'Tidak Layak' menjadi 0.

Hasil Klasifikasi menggunakan *Random Forest*

Tabel 1 Hasil Klasifikasi menggunakan *Random Forest*

No	Nama	Kelurahan	Label (data latih)	Prediksi	Status
1	Peri Pebrianto Lubis	Aer Bale	-	Tidak Layak	-
2	Muhammad Lan Harahap	Aer Bale	-	Tidak Layak	-
3	Panerangan Harahap	Aer Bale	-	Tidak Layak	-
4	Siti Alam Hasibuan	Aer Bale	-	Tidak Layak	-
5	Ali Yasbi	Aer Bale	-	Tidak Layak	-
6	Rahmat Pardede	Aer Bale	Layak	Tidak Layak	Salah
7	Nuraminah Hasibuan	Aer Bale	-	Tidak Layak	-
8	Mahammad Ilham Pardede	Aer Bale	-	Tidak Layak	-
9	Diana Santi Pardede	Aer Bale	-	Tidak Layak	-
10	Ardina Khoirunnisa Hasibuan	Aer Bale	-	Tidak Layak	-
...
2303	Rosmin Hasibuan	Janjiraja	-	Tidak Layak	-
2304	Middin Hasibuan	Janjiraja	-	Tidak Layak	-

Tabel 1. di atas menyajikan hasil akhir dari keseluruhan proses klasifikasi yang dilakukan oleh model *Random Forest* terhadap total 2304 data masyarakat. Kolom "Prediksi" adalah hasil utama dari penelitian ini, yang menampilkan hasil klasifikasi dari model untuk setiap data. Tanda strip (—) yang muncul di dua kolom lainnya bukanlah data yang hilang atau error, melainkan representasi visual dari metodologi penelitian. Kolom "Label (data latih)" berfungsi sebagai "Kunci Jawaban" (*Ground Truth*) yang ditentukan oleh peneliti pada tahap N Sampel (misal 300 data). Tanda (—) di kolom ini secara spesifik menandakan bahwa baris data

tersebut adalah bagian dari Data Uji (sisa 2004 data) yang tidak memiliki "Kunci Jawaban". Sebaliknya, baris yang terisi (seperti "Layak" atau "Tidak Layak") adalah bagian dari Data Latih yang digunakan untuk melatih model. Konsekuensinya, kolom "Status" yang merupakan perbandingan evaluasi antara "Label" dan "Prediksi" juga akan menampilkan tanda (—) untuk Data Uji, karena status "Benar" atau "Salah" tidak dapat dihitung tanpa adanya "Kunci Jawaban" sebagai pembanding.

Interface

Tampilan Halaman Menu Perhitungan

Tampilan halaman Menu

Perhitungan pada sistem ini dirancang untuk menyajikan hasil evaluasi performa algoritma klasifikasi *Random Forest* secara menyeluruh. Pada bagian utama, ditampilkan metrik evaluasi berupa *Akurasi*, *Precision*, *Recall*, dan *F1-Score*, yang masing-masing disajikan dalam kartu berwarna agar mudah dikenali dan dipahami oleh pengguna. Perhitungan metrik ini didasarkan pada *confusion matrix* yang ditampilkan dalam tabel, yang membandingkan data aktual (layak dan tidak layak) dengan hasil prediksi sistem. Dengan demikian, pengguna dapat melihat secara langsung jumlah prediksi benar maupun salah yang dihasilkan oleh model, sehingga evaluasi dapat dilakukan secara lebih transparan.



Gambar 4 Tampilan Halaman Perhitungan

Berdasarkan hasil perhitungan *confusion matrix* dari data uji (20% dari N Sampel), model *Random Forest* menghasilkan akurasi (*Accuracy*) sebesar 88,33%, yang berarti bahwa sebagian besar data uji (53 dari 60 data) dapat diprediksi dengan benar oleh sistem. Analisis lebih lanjut menunjukkan adanya *trade-off* (pertukaran) yang signifikan antara presisi dan recall. Model ini mencapai presisi (*Precision*) yang sangat tinggi, yaitu 90,91%, yang mengindikasikan bahwa ketika model memprediksi seseorang "Layak", prediksinya hampir selalu benar (hanya 1 *False Positive*). Namun, hal ini dicapai dengan mengorbankan recall, yang hanya mencapai 62,50%. Nilai recall ini menegaskan bahwa model melewatkan 37,5% dari total masyarakat yang sebenarnya "Layak", dan salah mengklasifikasikannya sebagai "Tidak Layak" (terdapat 6 kasus *False Negative*).

Dengan kata lain, hasil ini menunjukkan bahwa model yang dikembangkan sangat 'hati-hati' dan efektif dalam menghindari kesalahan pemberian bantuan kepada yang tidak berhak (presisi tinggi), namun belum mampu memenuhi aspek kelengkapan data (recall rendah) karena masih ada penerima yang berhak yang terlewat.

SIMPULAN

Berdasarkan penelitian yang dilakukan, maka dapat ditarik kesimpulan yaitu penelitian ini berhasil mengembangkan sistem klasifikasi kelayakan penerima bantuan sosial berbasis web yang secara efektif mengimplementasikan algoritma *supervised machine learning Random Forest*. Melalui implementasi alur kerja baru, yang mencakup fitur Penentuan N Sampel Data Latih dan Pelabelan (*Labelling*) Manual, sistem ini terbukti mampu memproses data masyarakat secara objektif, transparan, dan sistematis. Dengan demikian, metodologi penelitian yang digunakan untuk membangun model klasifikasi ini dapat dipertanggungjawabkan secara ilmiah. Berdasarkan 2.304 total data masyarakat, penelitian ini menerapkan metodologi N-Sample (Data Latih) sebanyak 300 data yang dilabeli secara manual oleh peneliti sebagai *Ground Truth*. Model *Random Forest* kemudian dilatih menggunakan 300 data ini dan diaplikasikan untuk memprediksi keseluruhan data. Hasil klasifikasi akhir pada keseluruhan 2.304 data menunjukkan bahwa 395 masyarakat diprediksi 'Layak' dan 1909 masyarakat diprediksi 'Tidak Layak'.

Hasil evaluasi model, yang dihitung dari 20% data uji internal (60 data dari 300 N Sampel), menunjukkan kinerja yang baik dan realistis. Model ini mencapai Akurasi 88,33%, Presisi 90,91%, *Recall* 62,50%, dan *F1-Score* 74,07%. Metrik ini membuktikan bahwa model sangat efektif dalam memprediksi secara tepat (presisi tinggi), namun

memiliki *trade-off* di mana model cenderung 'hati-hati' dan melewatkan beberapa kasus yang seharusnya 'Layak' (recall lebih rendah). Dengan demikian, sistem ini terbukti efektif dan layak digunakan untuk membantu pemerintah dalam menentukan kelayakan secara objektif, dengan catatan analisis *trade-off recall* tersebut.

DAFTAR PUSTAKA

- Argadianata, A. P., Fatah, D. A., & Sukri, H. (2025). Klasifikasi Kualitas Buah Apel Menggunakan Metode Random Forest. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 2016–2022.
- Arisandi, R. (2023). Perbandingan Model Klasifikasi Random Forest Dengan Resampling Dan Tanpa Resampling Pada Pasien Penderita Gagal Jantung. *Jurnal Gaussian*, 12(1), 136–145.
- Ayu Amanda, R., & Mayana Br Tarigan, N. (2025). Sistem Pendukung Keputusan Pemberian Bantuan Sembako Kepada Masyarakat Kurang Mampu Di Desa Sei Buluh Dengan Metode Analytical Hierarchy Process. *Jdsp Jurnal Data Science Penusa Jdsp*, 2(2), 605–612.
- Br Sebayang, E. R., Chrisnanto, Y. H., & Melina. (2023). Klasifikasi Data Kesehatan Mental DI Industri Teknologi Menggunakan Algoritma Random Forest. *Ijespg Journal*, 1(3), 237–253.
- Fatmawati, A., Purnamasari, A. I., & Ali, I. (2024). Implementasi Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sosial. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 745–750.
- Fauzi, M. S., & Samsudin, S. (2022). Smart School Berbasis Web Interaktif DI Sd Swasta Amaliyah Sunggal DENGAN Algoritma K-Means Cluster. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(3), 332–341.
- Hidayat, R., Tri Saputra, H., Husnah, M., Nabila, N., Hidayatullah, M. B., Naufal Nazhmi, M., Azra, J., & Rana, A. (2025). Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi DI Supermarket. *Simkom*, 10(1), 101–109.
- Kinanti, R. (2024). Penerapan Metode Clustering K-Means UNTUK Menentukan Prioritas Penerima Bantuan Program Beras UNTUK Rakyat Miskin (Raskin) Studi Kasus: Kecamatan Siulak. *Jurnal Informatika Dan Rekayasa Komputer (Jakakom)*, 4(2).
- Riansah, A., Nurdiawan, O., & Herdiana, R. (2025). Penerapan Algoritma Random Forest Dan Decision Tree Untuk Meningkatkan Akurasi Klasifikasi Penjualan Pada Toko Bangunan. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4242–4249.
- Romdona, S., Senja Junista, S., & Gunawan, A. (2025). Teknik Pengumpulan Data: Observasi, Wawancara Dan Kuesioner. *Jisosepol: Jurnal Ilmu Sosial Ekonomi Dan Politik*, 3(1), 39–47.
- Samsudin, S., Nurhalizah, N., & Fadilah, U. (2022). Sistem Informasi Pendaftaran Magang Dinas Pemuda Dan Olahraga Provinsi Sumatera Utara. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 4(2), 324–332.