

Implementasi Algoritma Random Forest untuk Klasifikasi Penerima Bantuan Bedah Rumah

Yana Iqbal Maulana

Teknologi Informasi, Universitas Bina Sarana Informatika

e-mail: yana.yim@bsi.ac.id

Abstrak

Bantuan bedah rumah merupakan suatu bentuk dukungan pemerintah bagi masyarakat yang kurang mampu untuk memperbaiki tempat tinggal. Pemerintah memiliki banyak data masyarakat yang membutuhkan bantuan bedah rumah, namun tanpa suatu sistem klasifikasi yang baik, hal ini dapat menyebabkan bantuan tersebut salah sasaran. Oleh karena itu, dibutuhkan suatu sistem klasifikasi menggunakan teknik data mining untuk membantu pemerintah menentukan sasaran yang tepat dan memastikan bahwa bantuan tersebut sesuai dengan kebutuhan masyarakat yang membutuhkan. Penelitian ini menggunakan algoritma Random Forest untuk klasifikasi penerima bantuan bedah rumah dengan jumlah data set 179 data. Tahapan penelitian yang dilakukan dimulai dari data collection dan data preprocessing. Selanjutnya citra akan diklasifikasi menggunakan algoritma Random Forest. Hasil klasifikasi akan dievaluasi menggunakan confusion matrix, kemudian dilakukan implementasi terhadap data uji. Hasil penelitian yang diperoleh yaitu nilai akurasi sebesar 93,8%, precision sebesar 89,47% dan recall sebesar 65,38 %. Dari hasil penelitian ini, diharapkan dapat membantu pemerintah dalam memastikan bahwa bantuan bedah rumah diterima oleh individu atau keluarga yang benar-benar membutuhkan. Selain itu, penelitian ini juga diharapkan dapat memberikan informasi penting bagi pemerintah dalam menentukan penerima bantuan di masa yang akan datang.

Kata kunci: Bedah rumah, Klasifikasi, Random Forest, Data Mining

Abstract

House renovation assistance is a form of government support for underprivileged people to improve their homes. The government has a lot of data on people who need house renovation assistance, but without a good classification system, this can cause this assistance to be misdirected. Therefore, a classification system is needed using data mining techniques to help the government determine the right targets and ensure that the assistance meets the needs of the people in need. This study used the Random Forest algorithm to classify beneficiaries of home improvement assistance with a total of 179 data sets. The stages of the research carried out started from data collection and data preprocessing. Furthermore, the image will be classified using the Random Forest algorithm. Classification results will be evaluated using a confusion matrix, then implementation of the test data will be carried out. The research results obtained are accuracy of 93.8%, precision of 89.47% and recall of 65.38%. From the results of this study, it is hoped that it can assist the government in ensuring that home renovation assistance is received by individuals or families who really need it. In addition, this research is also expected to provide important information for the government in determining beneficiaries of assistance in the future.

Keywords: House renovation, Classification, Random Forest, Data Mining

1. PENDAHULUAN

Bantuan bedah rumah merupakan suatu bentuk dukungan pemerintah bagi masyarakat yang kurang mampu untuk memperbaiki tempat tinggal [1]. Ini bertujuan untuk meningkatkan kualitas hidup dan kenyamanan masyarakat miskin dengan memberikan mereka akses terhadap rumah yang lebih baik dan layak huni. Melalui bantuan ini, pemerintah berupaya untuk membantu masyarakat miskin memperbaiki rumah mereka tanpa harus memikirkan biaya yang besar. Bantuan bedah rumah

merupakan suatu bentuk investasi pemerintah dalam membangun masyarakat dan meningkatkan kualitas hidup masyarakat miskin.

Pemerintah memiliki banyak data masyarakat yang membutuhkan bantuan bedah rumah, namun tanpa suatu sistem klasifikasi yang baik, hal ini dapat menyebabkan bantuan tersebut salah sasaran. Oleh karena itu, dibutuhkan suatu sistem klasifikasi menggunakan teknik data mining untuk membantu pemerintah menentukan sasaran yang tepat dan memastikan bahwa bantuan tersebut sesuai dengan kebutuhan masyarakat yang membutuhkan [2].

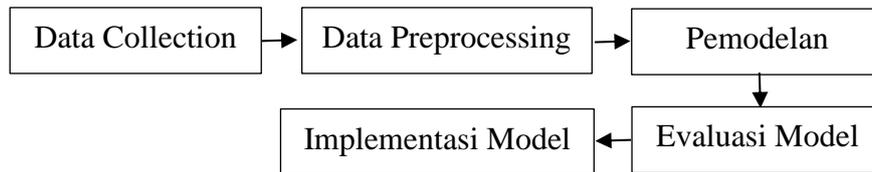
Metode klasifikasi data sering digunakan dalam berbagai bidang, termasuk dalam membantu pemerintah dalam menentukan penerima bantuan. Dalam hal ini, Random Forest dapat digunakan untuk menentukan penerima bantuan bedah rumah. Random Forest adalah algoritma yang fleksibel dan mudah digunakan menghasilkan hasil yang bagus di sebagian besar waktu. Random Forest terdiri dari sejumlah pohon, dengan setiap pohon tumbuh menggunakan beberapa bentuk pengacakan dan tugas lain itu beroperasi dengan membangun banyak pohon keputusan di waktu pelatihan dan mengeluarkan kelas yang merupakan mode dari kelas atau prediksi rata-rata dari pohon individu [3]. Tujuan dari penelitian ini adalah untuk mengklasifikasikan penerima bantuan bedah rumah menggunakan algoritma Random Forest dan mengevaluasi akurasi dari klasifikasi tersebut.

Beberapa peneliti sebelumnya telah melakukan penelitian sejenis, dengan algoritma dan data berbeda. Damuri, dkk [4] telah melakukan penelitian dengan judul “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako” dengan jumlah dataset 175 data. Hasil yang diperoleh yakni akurasi sebesar 86%, recall 85%, dan presisi 88%. Fitriani [5] melakukan penelitian “Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan”. Tools yang digunakan dalam penelitian ini adalah Rapid miner. Hasil dari penelitian ini adalah perhitungan dengan algoritma C4.5 menghasilkan nilai akurasi sebesar 91,25% dan nilai AUC sebesar 0,930 dengan tingkat diagnosa Excellent Classification sedangkan algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 87,11% dan nilai AUC sebesar 0,923 dengan tingkat diagnosa Excellent Classification. Penelitian selanjutnya dilakukan oleh Rosid, dkk [6] dengan judul “Klasifikasi Penerima Bantuan Sosial Dengan Algoritma Random Forest Untuk Penanganan Covid 19”. Penelitian ini menggunakan tools Rapid miner. Nilai akurasi yang diperoleh dari penelitian ini sebesar 93.44%

Dalam penelitian ini, data yang digunakan adalah data penerima bantuan bedah rumah yang berisi informasi tentang individu atau keluarga yang membutuhkan bantuan. Data ini akan diolah dan dibersihkan sebelum diterapkan algoritma Random Forest. Setelah itu, model Random Forest akan dibangun dan diuji dengan menggunakan data uji. Akurasi klasifikasi akan diukur dan dianalisis untuk mengevaluasi hasil klasifikasi yang diperoleh.

Dari hasil penelitian ini, diharapkan dapat membantu pemerintah dalam memastikan bahwa bantuan bedah rumah diterima oleh individu atau keluarga yang benar-benar membutuhkan. Selain itu, penelitian ini juga diharapkan dapat memberikan informasi penting bagi pemerintah dalam menentukan penerima bantuan di masa yang akan datang.

2. METODE PENELITIAN



Gambar 1. Research Method

Tahapan penelitian yang dilakukan dimulai dari Data Collection dan Data Preprocessing. Selanjutnya citra akan diklasifikasi menggunakan algoritma random forest. Hasil klasifikasi akan dievaluasi dan dianalisis, kemudian dilakukan implementasi terhadap data uji. Berikut adalah metode penelitian untuk klasifikasi penerima bantuan bedah rumah menggunakan algoritma Random Forest:

- a. Data Collection (Pengumpulan data): Dalam tahap ini, data tentang penerima bantuan bedah rumah akan dikumpulkan dan disiapkan untuk proses pemodelan. Data yang dibutuhkan meliputi informasi demografi, ekonomi, dan kondisi rumah dari masing-masing penerima bantuan.
- b. Data Preprocessing: Setelah data dikumpulkan, data tersebut akan diolah dan dipersiapkan untuk proses pemodelan. Langkah-langkah ini meliputi imputasi data hilang, normalisasi data, dan pembuatan dummies jika diperlukan.
- c. Pemodelan: Dalam tahap ini, algoritma Random Forest akan diterapkan untuk memodelkan data penerima bantuan bedah rumah.
- d. Evaluasi Model: Setelah proses pemodelan selesai, model akan dievaluasi untuk menentukan akurasi dan stabilitas model. Metode evaluasi yang dapat digunakan meliputi akurasi, confusion matrix, precision dan recall. Dalam proses evaluasi model, hasil dapat dikategorikan menjadi 5 kelompok yaitu [5]
 1. Akurasi bernilai 0.90 – 1.00 = Excellent classification
 2. Akurasi bernilai 0.80 – 0.90 = Good classification
 3. Akurasi bernilai 0.70 – 0.80 = Fair classification
 4. Akurasi bernilai 0.60 – 0.70 = Poor classification
 5. Akurasi bernilai 0.50 – 0.60 = Failure
- e. Implementasi Model: Setelah model divalidasi, model tersebut akan dilakukan implementasi untuk membantu dalam proses pengambilan keputusan untuk penerima bantuan bedah rumah.

Dengan mengikuti metode penelitian ini, diharapkan dapat diperoleh model yang dapat membantu dalam pengambilan keputusan untuk penerima bantuan bedah rumah secara akurat dan efisien.

3. HASIL DAN ANALISA

3.1. Dataset

Dalam penelitian ini penulis mengambil populasi sebanyak 200 untuk mengklasifikasi penduduk yang layak untuk mendapatkan bantuan bedah rumah. Atribut yang digunakan dalam penelitian ini adalah NIK, Nomor rumah, Nama kepala keluarga, jumlah tanggungan per rumah, lantai rumah, atap rumah, dinding rumah, penghasilan per bulan, sarana air bersih, listrik, kategori (label awal). Berikut adalah sample untuk dataset yang digunakan:

Tabel 1. Sample Data Penelitian

No	NIK	No Rumah	Nama Kepala Keluarga	Jumlah tanggungan per rumah	Lantai Rumah	Atap Rumah
1	123456	01	Agus	3	Keramik	Genteng
Penghasilan/ bulan	Dinding rumah	Sarana air bersih	Listrik	Label Awal		
3.500.000	Tembok	Ada	Ada	Tidak Layak		

3.2. Data Preprocessing

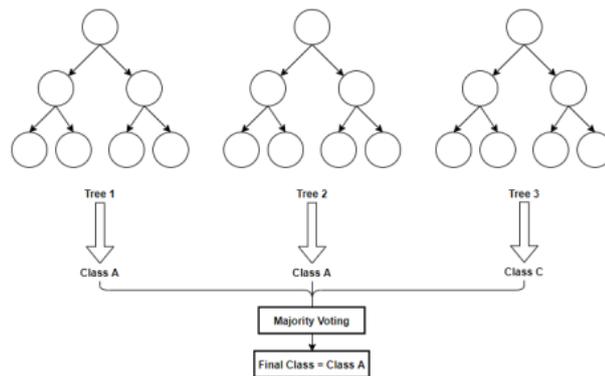
Dalam tahapan data preprocessing, beberapa hal yang perlu dilakukan antara lain:

- Pembersihan Data:** Ini termasuk menghapus duplikat, memperbaiki nilai yang hilang atau tidak valid, dan mengatasi masalah data yang tidak konsisten. Setelah proses pembersihan data, data yang diperoleh berjumlah 179 data. Sedangkan atribut yang dihilangkan yaitu nomor rumah dan nama kepala keluarga. Atribut tersebut dihilangkan karena dianggap tidak berpengaruh terhadap hasil penelitian.
- Enkoding Fitur:** Fitur diskrit seperti lantai rumah, atap rumah, dinding rumah, penghasilan per bulan, sarana air bersih, listrik harus diubah menjadi representasi numerik untuk diterima oleh algoritma Random Forest.
- Normalisasi Fitur:** Dalam beberapa kasus, skala fitur dapat mempengaruhi hasil model. Oleh karena itu, fitur harus dalam skala yang sama untuk menghindari bias.
- Pemilihan Fitur:** Dalam banyak kasus, beberapa fitur mungkin tidak relevan atau memiliki korelasi yang kuat dengan fitur lain. Pemilihan fitur yang tepat dapat meningkatkan performa model.
- Pembagian Data:** Data harus dibagi menjadi data latih dan data uji sebelum dilatih dan diuji dengan algoritma Random Forest. Dalam proses pembagian data, penulis

menggunakan perbandingan 70:30 yaitu 70% untuk data latih dan 30% untuk data uji. Jumlah data latih yaitu 125 data, sedangkan jumlah data uji sebesar 54 data

3.3. Pemodelan

Setelah data dipersiapkan, algoritma Random Forest dapat diterapkan pada data latih untuk membuat model klasifikasi. Berikut adalah ilustrasi proses klasifikasi dengan random forest [7]



Gambar 2. Ilustrasi Proses Random Forest

3.4. Evaluasi Model

Proses evaluasi model dalam penelitian ini menggunakan Confusion Matrix. Berikut adalah hasil Confusion Matrix yang diperoleh

Tabel 2. Hasil Confusion Matrix

	Tidak Layak	Layak	
Tidak Layak	151	9	94,38%
Layak	2	17	89,47%
	98,69%	65,38%	

Jumlah True Positive (TP) adalah 17 record diklasifikasikan sebagai penduduk LAYAK dapat bantuan dan False Negative (FN) sebanyak 9 record diklasifikasikan sebagai penduduk LAYAK tetapi TIDAK LAYAK dapat bantuan. Berikutnya 151 record untuk True Negative (TN) diklasifikasikan sebagai penduduk KURANG LAYAK dapat bantuan, dan 2 record False Positive (FP) diklasifikasikan sebagai penduduk TIDAK LAYAK tetapi LAYAK dapat bantuan bedah rumah.

Perhitungan akurasi

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \\
 \text{Accuracy} &= \frac{17+151}{17+2+9+151}
 \end{aligned}$$

$$\text{Accuracy} = 93,8\%$$

Perhitungan recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{17}{17 + 9}$$

$$\text{Recall} = 65,38 \%$$

Perhitungan precision

$$\text{Presicion} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Presicion} = \frac{17}{17 + 2}$$

$$\text{Presicion} = 89,47\%$$

Hasil perhitungan yang dilakukan dengan algoritma Random Forest dapat dikategorikan sebagai Excellent classification

3.5. Implementasi Model

Pada proses klasifikasi menggunakan algoritma Random Forest diperoleh sebuah model. Kemudian model tersebut diimplementasikan terhadap data uji. Berikut adalah sample untuk implementasi model pada data uji.

Tabel 3. Sample Implementasi Model

NIK	Jumlah tanggungan per rumah	Lantai Rumah	Atap Rumah	Penghasilan/ bulan	Dinding rumah	Sarana air bersih	Listrik	Label Awal	Hasil
123456	3	2	3	2	1	1	1	Tidak Layak	Tidak Layak

Dari implementasi terhadap sample data uji penduduk dengan NIK 123456 dengan variabel jumlah tanggungan per rumah 3 orang, kondisi lantai rumah = 2 (Keramik), kondisi atap rumah = 3 (Genteng), penghasilan per bulan 2 = 3.500.000, dinding rumah = 1 (Tembok), sarana air bersih = 1 (ada), listrik = 1 (ada), label awal tidak

layak menerima bantuan dan setelah diklasifikasikan menggunakan algoritma Random Forest mendapatkan hasil tidak layak menerima bantuan.

4. CONCLUSION

Berdasarkan hasil penelitian, dapat ditarik kesimpulan bahwa algoritma Random Forest memiliki potensi untuk digunakan sebagai metode untuk menentukan penerima bantuan bedah rumah. Algoritma ini dapat memproses data dengan cepat dan memiliki tingkat akurasi yang cukup tinggi, sehingga dapat memberikan hasil yang dapat diandalkan dalam menentukan penerima bantuan.

Rincian hasil penelitian klasifikasi penerima bantuan bedah rumah dengan bahwa algoritma Random Forest yaitu nilai akurasi sebesar 93,8%, precision sebesar 89,47% dan recall sebesar 65,38 %.

Namun, perlu diingat bahwa hasil dari penelitian ini masih terbatas dan memerlukan penelitian lebih lanjut untuk memvalidasi hasil dan meningkatkan akurasi dari algoritma. Oleh karena itu, perlu adanya penelitian lanjutan untuk mengevaluasi dan meningkatkan efektivitas dari implementasi algoritma Random Forest untuk klasifikasi penerima bantuan bedah rumah. Selain itu, diharapkan penelitian selanjutnya bisa menciptakan aplikasi untuk penerapan model yang didapat dari penelitian.

REFERENCES

- [1] D. Emmi Rahmawita Nasution1), “Pelaksanaan Program Bantuan Stimulan Perumahan Swadaya Kepada Masyarakat Miskin Terkait Undang-Undang Nomor 1 Tahun 2011 Tentang Perumahan Dan Kawasan Permukiman (Studi Di Kota Tanjungbalai),” *Citra Justicia Maj. Huk. dan Din. Masy.*, vol. 23, no. 1, pp. 1–13, 2022.
- [2] E. Widodo and A. Jaelani, “Penerapan Data Mining untuk Prediksi Penerimaan Bantuan Pangan Non Tunai (BPNT) di Desa Wanacala Menggunakan Metode Naive Bayes,” *J. Teknol. Pelita Bangsa*, vol. 13, no. 3, pp. 153–158, 2022.
- [3] N. Khasanah, R. Komarudin, N. Afni, Y. I. Maulana, and A. Salim, “Skin Cancer Classification Using Random Forest Algorithm,” *Sisfotenika*, vol. 11, no. 2, p. 137, 2021, doi: 10.30700/jst.v11i2.1122.
- [4] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,” *J. Ris. Komput.*, vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [5] E. Fitriani, “Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan,” *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.
- [6] G. D. Abdur Rosid1)*, Odi Nurdiawan2), “KLASIFIKASI PENERIMA BANTUAN SOSIAL DENGAN ALGORITMA RANDOM FOREST UNTUK PENANGANAN COVID 19,” vol. 10, no. 2, 2022.
- [7] Q. Iman and A. W. Wijayanto, “Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode Random Forest dan Support Vector Machine,” *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 178, 2021, doi: 10.26418/justin.v9i2.44137.