

Klasifikasi Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Kelayakan Bantuan Rehabilitasi Rumah Tidak Layak Huni Pada Desa Lenek Duren Kecamatan Aikmel Kabupaten Lombok Timur

Suhartini¹, Hariman Bahtiar²

Fakultas Teknik, Universitas Hamzanwadi
Suhartini311279@gmail.com¹, harimaob@gmail.com²

Abstrak

Rumah adalah suatu kebutuhan sangat mendasar untuk semua orang di samping sandang pangan. Rumah dapat mencerminkan tingkatan kesejahteraan dan tingkatan kesehatan tentang penduduk/penghuninya. Kepantasan suatu rumah sebagai tempat perlindungan baik dapat dilihat dari struktur dan fasilitas bangunan. Niat baik pemerintah daerah dalam pengentasan kemiskinan melalui program RTLH, harus didukung dengan tingkat akurasi data. Selain tingkat akurasi data juga dibutuhkan efisiensi waktu pengolahan data penerima bantuan. Penelitian ini memanfaatkan teknik Data Mining dengan Algoritma K-Nearest Neighbor berbasis Particle Swarm Optimization dalam mengklasifikasi kelayakan penerimaan bantuan rehabilitasi rumah tidak layak huni. Dengan metode perhitungan algoritma K-Nearest Neighbor berbasis Particle Swarm Optimization dapat diketahui nilai akurasi yang terbaik dalam prediksi kelayakan penerimaan bantuan rehabilitasi rumah tidak layak huni. Setelah dilakukan pengujian dengan menggunakan algoritma *K-NN* dan algoritma *K-NN* berbasis *Particle Swarm Optimization* maka hasil terbaik yang didapat adalah algoritma *K-NN* menghasilkan nilai akurasi sebesar 89,29% dan nilai AUC 0,786. Kemudian algoritma *K-NN* berbasis *Particle Swarm Optimization* menghasilkan nilai akurasi sebesar 95,33% dan nilai AUC 0.970. Setelah melakukan pengujian terhadap kedua model tersebut memiliki perbedaan tingkat akurasi sebesar 6,04% dan perbedaan nilai AUC 0.184.

Kata kunci : *Rumah, Algoritma K-NN, Particle Swarm Optimization*

Abstract

Home is a very basic requirement for all people besides food and clothes. Home can express prosperity level and health level about resident / its dweller. A high quality of home as good shelter can be seen from its building, facility and structure. The local government program to decrease poverty through program of RTLH, have to be supported with data accuration. Besides good data accuration, it is also required data-processing time efficiency of receiver of aid. This research used Data technique of Mining with Algorithm of K-Nearest Neighbor base on Particle Swarm Optimization in classification elegibility of acceptance of aid rehabilitate house improper dwell. By using method of calculation of algorithm of K-Nearest Neighbor base on Particle Swarm Optimization, the accuration value in prediction elegibility of acceptance of aid rehabilitate house improper dwell is calculated. After conducting a test by using algorithm of K-NN and algorithm of K-NN base on Particle Swarm Optimization, the result is algorithm of K-NN yield accuration value equal to 89,29% and value of AUC 0,786. The algorithm of K-NN base on Particle Swarm Optimization yield accuration value equal to 95,33% and value of AUC 0.970. After conducting the tests to both of the model, the difference is showed with 6,04% percentage and difference of value of AUC 0.184.

Keyword : *House, Algorithm of K-NN, Particle Swarm Optimization*

1. Pendahuluan

Rumah yang tidak layak huni dapat diketahui dari struktur bangunan rumah dan fasilitas rumah yang ada. Struktur bangunan rumah meliputi jenis lantai, jenis atap, dan jenis dinding. Informasi yang di peroleh dari tim pelaksana di lapangan bahwa terdapat fenomena ketidak tepat sasaran penerima bantuan rehabilitasi sosial rumah tidak layak huni, hal ini terjadi karena kesalahan dalam pengolahan data calon penerima bantuan oleh panitia pelaksana sehingga terjadi kesalahan dalam pengambilan keputusan.

2. Tinjauan Pustaka

2.1 Penelitian Terkait

- Penelitian yang dilakukan oleh Irfan Aribowo dengan judul Analisa Penerapan *K-Nearest Neighbor* Dalam Mengklasifikasikan Penerima Dana Pemugaran Rumah Tidak Layak Huni. Pada penelitian ini, dikemukakan bahwa pemerintah memiliki program kerja untuk membantu masyarakat yang kurang mampu. Program kerjanya dinamakan Bantuan Dan Pemugaran Rumah Tidak Layak Huni (RTLH). Agar pemberian bantuan tepat sasaran maka dilakukan klasifikasi untuk penentuannya [1]
- Penelitian yang dilakukan oleh Ricky Imanuel Ndaumanu, Kusrini, M. Rudyanto Arief yang berjudul Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa Dengan Metode *K-Nearest Neighbor*. Pada

penelitian ini, dikemukakan bahwa dari jumlah pendaftaran mahasiswa baru ini, banyak juga mahasiswa yang mengundurkan diri setiap tahunnya yang disebabkan berbagai masalah. Oleh karena adanya mahasiswa yang mengundurkan diri, penulis melakukan analisis pengunduran diri mahasiswa menggunakan algoritma *K-Nearest Neighbor*. Penelitian ini bertujuan untuk memprediksi seberapa tingkat akurasi pengunduran diri mahasiswa STIKOM UYELINDO Kupang[2].

2.2 Landasan Teori

1. Data Mining

Data mining adalah proses yang menggunakan teknik *static*, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database dalam jumlah besar [3].

2. Algoritma K-Nearest Neighbor (K-NN)

K-Nearest Neighbor termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. K-NN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing. Algoritma *K-Nearest Neighbor* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dan kasus lama yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada [2].

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), yang paling sering digunakan adalah euclidean distance yaitu [4]:

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

Dimana $a = a_1, a_2, \dots, a_n$, dan $b = b_1, b_2, \dots, b_n$ mewakili n nilai atribut dari dua record.

3. Particle Swarm Optimization

Particle Swarm Optimization (PSO) merupakan algoritma pencarian berbasis populasi dan diinisialisasi dengan populasi solusi acak yang disebut partikel. PSO merupakan metode pencarian yang berasal dari penelitian untuk gerakan sekelompok burung atau ikan. Untuk PSO dapat diasumsikan sebagai kelompok burung secara acak mencari makanan disuatu daerah. Burung tersebut tidak tahu dimana makanan tersebut berada, tapi mereka tahu seberapa jauh makanan itu berada, jadi strategi terbaik untuk menemukan makanan tersebut adalah dengan mengikuti burung yang terdekat dari makanan tersebut. PSO digunakan untuk memecahkan masalah optimasi [5].

4. Rapidminer

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik [3].

2.3 Tahapan Penelitian

Tahapan dari penelitian ini meliputi :

- Pengumpulan data
Tahapan dalam penelitian ini adalah adalah pengumpulan data yang dilakukan dengan berbagai teknik seperti observasi, dokumentasi, dan studi pustaka.
- Pengolahan data awal
- Model yang di usulkan
- Eksperimen / Pengujian
- Evaluasi Hasil

3. Metode Penelitian

3.1. Analisa Kebutuhan

Data yang digunakan dalam penelitian ini merupakan data primer yang diambil dari Kantor Desa Lenek Duren Kecamatan Aikmel. Dokumentasi dilakukan pada kantor tersebut untuk mengumpulkan data penerima bantuan. Data penerima bantuan dikumpulkan pada periode 2016. Seperti pada gambar 8 data yang dimiliki pada data awal berupa data penerima bantuan rehabilitasi rumah tidak layak huni yang

terdiri dari 13 atribut serta 1 atribut akhir dikosongkan.

Spesifikasi perangkat keras dan perangkat lunak yang digunakan dapat dilihat pada tabel 5.

Tabel 5 Spesifikasi Perangkat Keras dan Perangkat lunak

Perangkat Keras	Perangkat lunak
Processor 2.41 GHz	Sistem Operasi Windows 7 32 Bit
HDD Storage 500 GB	Rapidminer 5
RAM Memory 2 GB	

3.2. Pemahaman Data (Data Understanding)

Dataset penerimaan bantuan rehabilitasi rumah tidak layak huni yang didapatkan dari Kantor Desa Lenek Duren berupa dokumen excel sejumlah 149 data.

- Pengumpulan data awal

Adapun sumber data utama yang digunakan dalam penelitian ini adalah *dataset* penerimaan bantuan rehabilitasi rumah tidak layak huni pada tahun 2016 dengan format .xlsx karena data yang diberikan berupa dokumen *excel*.

- Mendeskripsikan data

Dataset penerimaan bantuan rehabilitasi rumah tidak layak huni terdiri dari 13 atribut antara lain Nama Penerima, Jenis Kelamin, NIK, Umur, Pekerjaan, Alamat, Penghasilan, Jumlah Tanggungan, Jenis Dinding, Jenis Lantai, Jenis Atap, Jenis bangunan dan Keterangan.

- Evaluasi kualitas data

Hasil evaluasi terhadap kualitas data yaitu terdapat nilai kosong / null yang disebut

dengan missing value pada atribut dalam dataset penerima bantuan.

- Pemilihan Atribut

Atribut yang digunakan adalah Nama Penerima, Pekerjaan, Penghasilan, Jenis Dinding, Jenis Lantai, Jenis Atap, Jenis bangunan dan Keterangan.

3.3. Persiapan Data (Data Preparation)

Persiapan data mencakup semua kegiatan untuk membangun *dataset* penerima rehabilitasi rumah tidak layak huni yang akan diterapkan ke dalam alat pemodelan, dari data mentah awal berupa *dataset* penerima rehabilitasi rumah tidak layak huni dan selanjutnya akan melakukan proses *data mining* sebagai berikut :

- Pembersihan Data

Pada tahap ini, dilakukan pembersihan terhadap data yang tidak lengkap, data yang double atau sama dan data yang tidak konsisten. Pada tahap ini data akan dibersihkan dengan cara dihapus secara manual, dan akan dilakukan penghapusan atribut ID serta akan mengganti pembersihan data dilakukan setelah integrasi dan seleksi data dilakukan. Proses integrasi akan memudahkan proses pencarian data, sementara seleksi data akan mengurangi jumlah data yang akan dibersihkan.

- Data integration

Pada tahap ini akan dilakukan penggabungan data. Data integration

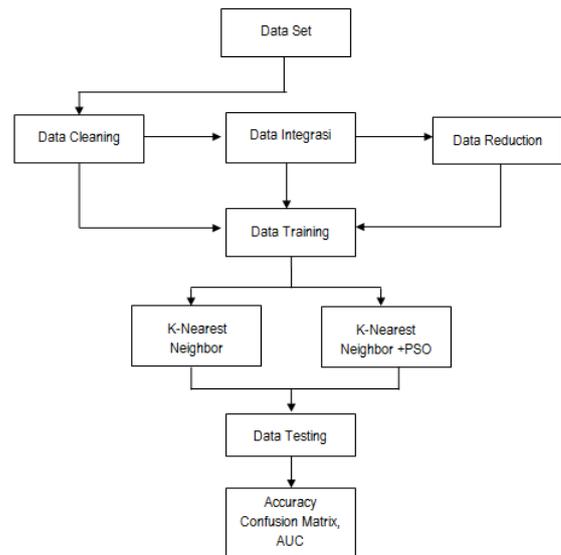
berfungsi menyatukan tempat penyimpanan yang berbeda ke dalam satu data. Dalam penelitian ini hanya ada satu tempat penyimpanan data.

- Seleksi Data

Pada tahap ini akan dilakukan penyeleksian data untuk mengurangi data yang tidak relevan dan redundant. Atribut yang tidak relevan adalah atribut yang berisi informasi yang tidak berguna untuk melakukan penambangan data, sedangkan atribut yang berlebihan (redaundant) adalah atribut yang menduplikasi banyak atau semua informasi yang terdapat didalam satu atau lebih atribut lain[2]. Pada data penerimaan bantuan rehabilitasi rumah tidak layak huni berisi 13 atribut, dan tidak semua dipakai untuk mengklasifikasikan siapa yang layak dan tidak layak mendapatkan rehabilitasi rumah tidak layak huni. Maka dari itu dilakukan penghapusan atribut yang tidak terpakai, atribut tidak terpakai adalah atribut jenis kelamin, NIK, umur, alamat dan jumlah tanggungan.

3.4. Model yang diusulkan

Pemodelan adalah fase yang secara langsung melibatkan teknik *data mining* yaitu dengan melakukan pemilihan teknik *data mining* dan menentukan algoritma yang akan digunakan. Berikut gambar model yang diusulkan.



Gambar 1. Model yang diusulkan

Keterangan :

- Data Set yaitu kumpulan data mentah.
- Data *Cleaning* yaitu membersihkan data yang kosong.
- Data *Integration* yaitu menyatukan data yang berbeda kedalam satu file.
- Data *Reduction* yaitu penentuan atribut yang kita gunakan.
- Data *training* yaitu data yang sudah diolah kemudian dimasukkan untuk mengetahui hasil dari *K-Nearest Neighbor*.
- K-Nearest Neighbor* yaitu melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.
- Particle Swarm Optimization* (PSO) yaitu digunakan untuk memecahkan masalah optimasi.
- Data *testing* yaitu mencari akurasi menggunakan validasi dimana data testing terdiri dari algoritma *K-Nearest Neighbor* dan PSO. Kemudian data testing terdiri dari

operator apply model dan performance, kemudian menentukan hasil akurasiya menggunakan *confusion matrix*.

4. Hasil dan Pembahasan

4.1. Hasil Pengujian Algoritma K-Nearest Neighbor

Pengujian dilakukan untuk mengetahui pengaruh nilai k terhadap tingkat akurasi. Nilai k yang terbaik pada algoritma ini tergantung pada jenis data yang digunakan. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Dalam kasus khusus, klasifikasi diprediksi berdasarkan training data yang paling dekat (dengan kata lain, $k=1$) sehingga disebut sebagai algoritma Nearest Neighbor. Nilai k (ketetanggaan paling dekat) yang digunakan dalam penelitian ini adalah 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 dengan *cross validation* 10. Berikut hasil pengujian yang telah dilakukan.

Tabel Hasil Pengujian Algoritma K-NN

k	Accuracy	Precision	Recall	AUC
1	89,29%	88,81%	92,92%	0,500
2	89,29%	93,15%	87,08%	0,786
3	88,57%	86,70%	94,03%	0,858
4	89,24%	87,81%	94,03%	0,897
5	85,24%	82,23%	95,14%	0,946
6	88,57%	87,43%	95,14%	0,961
7	85,90%	83,43%	95,14%	0,961
8	86,57%	84,32%	95,14%	0,953
9	85,24%	81,37%	97,64%	0,953
10	85,24%	82,62%	95,14%	0,956

Dari proses uji coba yang dilakukan, didapatkan hasil pengujian dengan nilai tertinggi terdapat pada nilai $k = 2$, yaitu *accuracy* 89,29%,

precision 93,15%, *recall* 87,08% dan nilai *AUC* 0,786.

4.2. Hasil Pengujian Algoritma K-Nearest Neighbor berbasis PSO

Berikut ini adalah hasil dari pengujian algoritma *K-Nearest Neighbor* berbasis *Particle Swarm Optimization*.

Tabel Hasil Pengujian Algoritma K-NN berbasis PSO

K	Accuracy	Precision	Recall	AUC
1	94,67%	96,07%	95,00%	0,500
2	94,67%	97,89%	92,64%	0,862
3	92,62%	93,14%	94,31%	0,885
4	94,67%	97,75%	92,78%	0,930
5	93,33%	95,89%	92,64%	0,978
6	95,33%	98,89%	92,78%	0,970
7	93,33%	95,25%	92,78%	0,969
8	94,62%	98,00%	92,46%	0,982
9	91,95%	93,25%	92,92%	0,982
10	93,33%	95,21%	92,92%	0,987

Dari hasil pengujian pada tabel, dapat diketahui bahwa nilai akurasi tertinggi terdapat pada saat nilai $k = 6$ yaitu *accuracy* 95,33%, *precision* 98,89%, *recall* 92,78% dan *AUC* 0,970.

4.3. Pembahasan

Hasil dari pengujian yang telah dilakukan dengan melakukan evaluasi menggunakan *Confusion Matrix* dan *ROC Curve*, ternyata terbukti bahwa pengujian yang dilakukan algoritma *K-Nearest Neighbor* (K-NN) berbasis *Particle Swarm Optimization* (PSO) lebih tinggi nilai akurasiya dibandingkan dengan algoritma *K-Nearest Neighbor* saja.

Nilai akurasi untuk algoritma *K-Nearest Neighbor* adalah 89,29% dan nilai akurasi algoritma *K-*

Nearest Neighbor (K-NN) berbasis *Particle Swarm Optimization* (PSO) adalah 95,33%. Untuk evaluasi menggunakan *ROC Curve* akan menghasilkan nilai *AUC* untuk model algoritma *K-NN* menghasilkan nilai *AUC* 0,786 dengan nilai *Fair Classification* dan *K-NN* berbasis *PSO* menghasilkan nilai 0,970 dengan nilai *Excellent Classification*.

Dari hasil pengujian *Confusion Matrix* dan *AUC* ternyata dapat memberikan nilai akurasi yang terbaik. Berdasarkan pengujian tersebut, algoritma *K-NN* berbasis *PSO* mampu memberikan nilai akurasi yang terbaik.

5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan bahwa hasil untuk nilai akurasi algoritma *K-nearest Neighbor* adalah 89,29%, sedangkan hasil akurasi algoritma *K-Nearest Neighbor* berbasis *PSO* adalah 95,33% sehingga selisih untuk nilai akurasi keduanya adalah 6,04%. Dari hasil akurasi dan nilai *AUC* tersebut, dapat disimpulkan bahwa penerapan teknik *Particle Swarm Optimization* (*PSO*) dapat meningkatkan nilai akurasi pada algoritma *K-Nearest Neighbor*.

6. Daftar Pustaka

- [1] I. Aribowo, "Analisa Penerapan *K-Nearest Neighbor* Dalam Mengklasifikasikan Penerima Dana Pemugaran Rumah Tidak Layak Huni," pp. 0–1, 2016.
- [2] R. I. Ndaumanu and M. R. Arief, "Analisis

Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode *K-Nearest Neighbor*," vol. 1, no. 1, 2014.

- [3] I. W. S. Wicaksana, "Belajar Data Mining Dengan Rapidminer." 2013.
- [4] Sumarin, "Implementasi Algoritma *K-Nearest Neighbor* Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM," vol. 01, pp. 52–62, 2015.
- [5] A. Rakhman, "PREDIKSI KETEPATAN LULUS MAHASISWA MENGGUNAKAN METODE *DECISION TREE* BERBASIS *PARTICLE SWARM OPTIMATION* (*PSO*). 6, no. 1, pp. 193–197, 2017.