

## KLASIFIKASI KESEJAHTRAAN RUMAH TANGGA DENGAN PEPENERAPAN ALGORITMA DECISION TREE SELEKSI ATRIBUT GENETIK ALGORITMA

*Jauhari Maulani, S.Kom, M.Kom*  
(*jauharimaulani@gmail.com*)

### ABSTRAK

*kesejahteraan sosial menunjuk ke jangkauan pelayanan untuk memenuhi kebutuhan masyarakat. Ini adalah istilah yang digunakan dalam ide negara sejahtera. kemiskinan merupakan masalah kompleks yang dipengaruhi oleh berbagai faktor yang saling berkaitan, antara lain: tingkat pendapatan, kesehatan, pendidikan, akses terhadap barang dan jasa, lokasi, geografis, gender, dan kondisi lingkungan. Melakukan pemetaan kemiskinan yaitu langkah awal dalam upaya penanggulangan kemiskinan yaitu mengenali karakteristik dari penduduk yang miskin sehingga diperlukan pemetaan kemiskinan yang digunakan sebagai alat untuk memecahkan persoalan yang mereka alami. Pada pemetaan kemiskinan di beberapa wilayah kota, daerah atau kabupaten yang telah di lakukan survey dengan beberapa pendataan. Dari data yang telah terkumpul pada pemetaan kemiskinan tersebut maka dapat diketahui tingkat kesejahteraan rumah tangga.*

*Penelitian ini menerapkan algoritma Decision Tree C4.5 dengan Seleksi Atribut GA (Genetik Algoritma) untuk klasifikasi status kesejahteraan rumah tangga. pada data kemiskinan dalam menentukan status kesejahteraan rumah tangga miskin kedalam 2 (dua) kategori RTS (Rumah Tangga Sasaran) yaitu RTSM (Rumah Tangga Sangat Miskin/Fakir Miskin) dan RTM (Rumah Tangga Miskin). Setelah dilakukan pengujian yaitu algoritma C4.5 sendiri sudah memiliki akurasi yang cukup baik yaitu sebesar 82.08 % dengan proses seleksi atribut oleh algoritma genetika, model yang terbentuk dapat ditingkatkan lagi menjadi 82.98% dalam mengklasifikasikan status kesejahteraan rumah tangga miskin kedalam 2 (dua) kategori RTS (Rumah Tangga Sasaran) yaitu RTSM (Rumah Tangga Sangat Miskin/Fakir Miskin) dan RTM (Rumah Tangga Miskin).*

**Kata Kunci** : 2 (dua) kategori RTS (Rumah Tangga Sasaran), Algoritma Decision Tree, Seleksi Atribut Genetika Algoritma

### LATAR BELAKANG

Pada penelitian ini mencoba metode algoritma decision tree dengan seleksi atribut genetik algoritma, untuk klasifikasi status kesejahteraan rumah tangga. pada data kemiskinan dalam menentukan status kesejahteraan rumah

tangga miskin kedalam 2 (dua) kategori RTS (Rumah Tangga Sasaran) yaitu RTSM (Rumah Tangga Sangat Miskin/Fakir Miskin) dan RTM (Rumah Tangga Miskin). Masalah kemiskinan merupakan salah satu persoalan mendasar yang menjadi pusat perhatian pemerintah di negara manapun. Salah satu aspek penting untuk mendukung

strategi penanggulangan kemiskinan adalah tersedianya data kemiskinan yang akurat dan tepat sasaran. Pengukuran kemiskinan yang dapat dipercaya dapat menjadi instrumen tangguh bagi pengambil kebijakan dalam memfokuskan perhatian pada kondisi hidup orang miskin. Data kemiskinan yang baik dapat digunakan untuk mengevaluasi kebijakan pemerintah terhadap kemiskinan, membandingkan kemiskinan antar waktu dan daerah, serta menentukan target penduduk miskin dengan tujuan untuk memperbaiki posisi mereka.

### **RUMUSAN MASALAH**

Permasalahan dari latar belakang di atas bagaimana menerapkan algoritma decision tree dengan seleksi atribut genetik algoritma untuk mengklasifikasi tingkat status kesejahteraan rumah tangga, agar bisa mendapatkan hasil tingkat akurasi yang lebih tepat.

### **TUJUAN PENELITIAN**

Penelitian ini bertujuan untuk penerapan algoritma decision tree dengan seleksi atribut genetik algoritma, agar dapat mengklasifikasi tingkat status kesejahteraan rumah tangga agar bisa menganalisa hasil yang lebih tepat. serta menambah pengalaman serta khasanah ilmu pengetahuan para peneliti dan pembaca hasil penelitian mengenai masalah klasifikasi tingkat status kesejahteraan rumah tangga dengan penerapan algoritma decision tree dengan seleksi atribut genetik algoritma.

### **MANFAAT PENELITIAN**

Penelitian pendataan status kesejahteraan rumah tangga dianggap sebagai tugas penting dan harus dilakukan dengan tepat

dan efisien. Maka penelitian ini diharapkan bermanfaat sebagai berikut:

1. Agar dapat digunakan oleh pemerintah dalam pengklasifikasian status kesejahteraan rumah tangga untuk analisa yang lebih tepat.
2. Dari penelitian yang dilakukan diharapkan dapat memberikan sumbangan untuk *data mining* khususnya algoritma decision tree dengan seleksi atribut genetic algoritma.

### **ANALISA KEBUTUHAN**

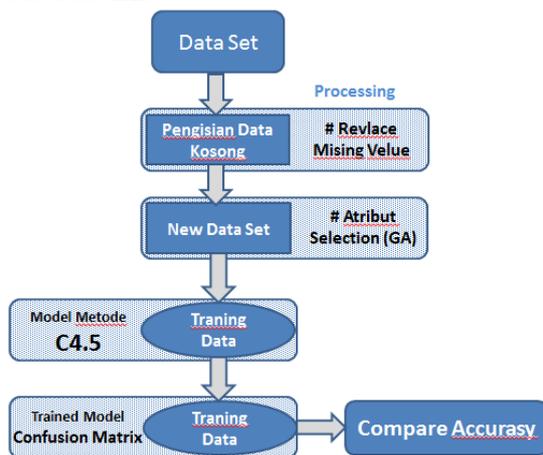
Analisis tingkat kesejahteraan merupakan bagian penting untuk ketepatan penentuan klasifikasi kesejahteraan rumah tangga, yang mana dari klasifikasi tersebut bermanfaat untuk pemerintah dalam menanggulangi tingkat kemiskinan. Oleh karena itu, sistem pengklasifikasian menggunakan metode algoritma decision tree dengan seleksi atribut genetik algoritma mungkin akan sangat bermanfaat dengan membawa semua hal itu. Sesuai informasi berbasis computer atau sistem pendukung keputusan dapat membantu dalam mencapai pengujian klinis. Algoritma decision tree merupakan pengklasifikasian yang paling sederhana, mudah diimplementasikan. Namun masih terdapat kesulitan dalam menangani data dimensi tinggi juga untuk menentukan atribut dalam akurasi prediksi masih dianggap kurang. Pencarian untuk subset yang optimal akan sangat mahal terutama ketika atribut makin meningkat dengan jumlah data yang tersedia. Terkadang tidak layak untuk dipergunakan. Oleh karena itu, dibutuhkan metode yang lebih akurat dalam optimasi klasifikasi tersebut.

Dengan adanya seleksi atribut dapat mengurangi dimensi data, hal ini memungkinkan lebih efektif dalam operasi yang lebih cepat dari beberapa algoritma data mining. Dengan adanya

seleksi atribut membuat algoritma data mining lebih cepat dan lebih efektif karena dapat membuang beberapa informasi yang tidak relevan dan berlebihan. Dalam beberapa kasus, hasilnya adalah representasi, lebih ringkas mudah diinterpretasikan dari target suatu konsep data. Terlebih lagi hasil dengan seleksi atribut memungkinkan dapat meningkatkan akurasi dalam pengklasifikasian data, sehingga representasi terlihat lebih padat atau ringkas dan dapat memudahkan penginterpretasi dari target secara konseptual.

**METODE YANG DIUSULKAN**

Menggambarakan alur metode yang diusulkan serta menjelaskan cara kerja model yang diusulkan. Metode ini akan digambarkan secara skematik dan disertai dengan formula perhitungan. Model akan dibentuk dari data yang sudah diolah, dan hasil pengolahan model akan diukur dengan model yang ada saat ini.



Gambar 1 Metode yang diusulkan

Pada gambar 3.1 model yang di usulkan pada pengolahan data awal, data set dilakukan preprocessing dengan menggunakan replace missing value untuk mengisi data yang kosong, setelah itu data tersebut dilakukan lagi dengan

menyeleksi atribut menggunakan Genetika Algoritma.

Data yang telah diseleksi atribut menggunakan GA diproses kembali dengan metode algoritma C4.5. Confusion Matrix melakukan evaluasi hasil dari penerapan algoritma C4.5 untuk mendapatkan nilai akurasi, precision, dan recall serta mendapatkan nilai AUC dan ROC Curva.

**HASIL PENGUJIAN**

Hasil dari uji coba yang dilakukan yaitu untuk menghasilkan nilai *accuracy*, dan nilai AUC (*Area Under Curve*). Hasil akurasi yang dihasilkan dengan algoritma klasifikasi C4.5 yang di lakukan dengan 10 K-Validation sebesar 82.10 % sudah cukup baik, maka pengujian dicoba ditingkatkan dengan menambah seleksi atribut *Genetika Algoritma* maka mendapatkan hasil sebagai berikut :

Population Size	Minimum Number Of Fitures											
	5		6		7		8		9		10	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
1	82.88%	0.842	82.78%	0.841	82.88%	0.842	82.78%	0.839	82.69%	0.845	82.69%	0.845
2	82.98%	0.846	82.98%	0.846	82.78%	0.840	82.78%	0.841	82.59%	0.846	82.98%	0.846
3	82.98%	0.842	82.98%	0.842	82.98%	0.842	82.98%	0.842	82.98%	0.842	82.98%	0.848
4	82.98%	0.842	82.98%	0.842	82.98%	0.842	82.98%	0.842	82.98%	0.848	82.98%	0.846
5	82.98%	0.842	82.98%	0.842	82.98%	0.842	82.98%	0.842	82.98%	0.848	82.98%	0.846

Gambar 2 Hasil Pengujian C 4.5 + GA

terlihat penggunaan algoritma genetic yang melakukan seleksi atribut dengan nilai minimum number of fitur dan population size menghasilkan tingkat akurasi tertinggi 82.98 %. Hasil tersebut kemudian digunakan untuk membandingkan seberapa besar kenaikan tingkat akurasi C 4.5 dengan C4.5 berbasis Genetik Algoritma.

**Analisis Evaluasi dan Validasi Model**

Dari hasil pengujian diatas, dengan dilakukan evaluasi baik secara *confusion matrix* maupun *ROC curve* ternyata terbukti bahwa pengujian yang dilakukan algoritma klasifikasi C4.5 berbasis GA memiliki nilai akurasi yang lebih tinggi dibanding hanya menggunakan algoritma klasifikasi C4.5. Nilai akurasi untuk model algoritma klasifikasi C4.5 sebesar 82.10 % dan nilai akurasi algoritma klasifikasi C4.5 berbasis GA sebesar 82.98 % dengan selisih akurasi sebesar 0.88 %, dapat dilihat pada Tabel dibawah ini.

*Tabel Pengujian Algoritma klasifikasi C4.5 dan C4.5 berbasis GA*

	Accuracy	AUC
C 4.5	82.10 %	0.848
C 4.5 + GA	82.98 %	0.848

Untuk evaluasi menggunakan *ROC curve* sehingga menghasilkan nilai *AUC (Area Under Curve)* untuk model algoritma klasifikasi C4.5 berbasis GA (*Genetik Algoritma*) menghasilkan nilai 0.848 dengan nilai diagnosa *Good Classification*.

**KESIMPULAN**

Algoritma genetika dapat diterapkan untuk proses pengklasifikasian status kesejahteraan rumah tangga. Dengan adanya seleksi atribut akan lebih mudah dipecah oleh algoritma *Decision Tree C4.5*. Model yang diperoleh pun menjadi lebih akurat dalam mengklasifikasikan status kesejahteraan. Dari penelitian yang dilakukan model yang terbentuk dengan algoritma *Decision Tree C4.5* sendiri sudah memiliki akurasi yang cukup baik yaitu sebesar 82.08 % dengan proses seleksi atribut oleh algoritma genetika, model yang terbentuk dapat ditingkatkan lagi menjadi 82.98% dalam

mengklasifikasikan status kesejahteraan rumah tangga miskin kedalam 2 (dua) kategori RTS (*Rumah Tangga Sasaran*) yaitu *RTSM (Rumah Tangga Sangat Miskin/Fakir Miskin)* dan *RTM (Rumah Tangga Miskin)*. Nilai *AUC (Area Under Curve)* untuk model algoritma klasifikasi *Decision Tree C4.5* dengan seleksi atribut GA (*Genetik Algoritma*) menghasilkan nilai 0.848 dengan nilai diagnosa *Good Classification*.

**SARAN**

Proses penyeleksian attribut dengan menggunakan algoritma genetika terbukti dapat meningkatkan akurasi dari algoritma *Decision Tree C4.5*. Namun ada beberapa factor yang dapat dicoba untuk penelitian selanjutnya, agar algoritma *Decision Tree C4.5* dapat menghasilkan model yang lebih baik lagi:

1. Untuk menghasilkan prediksi yang lebih akurat dari penelitian ini, diperlukan pembersihan (*data cleansing*) dari masukan data yang tidak konsisten dan data yang rusak atau yang disebut dengan data sampah, pada tahap pengolahan awal.
2. Untuk dapat dipergunakan oleh bagian badan pusat statistik sebaiknya dilakukan pembuatan aplikasi yang dapat melakukan pengklasifikasian status kesejahteraan rumah tangga tersbeut, sehingga memudahkan para *user* dalam menganalisa.

### DAFTAR PUSTAKA

- [1] Afandi, Weri Nova (2011). Identifikasi Karakteristik Rumah Tangga Miskin di Kabupaten Padang Pariaman. Jakarta :Universitas Andalas.
- [2] Amir Hamzah, [http://repository.akprind.ac.id/sites/files/conferenceproceedings/2012/hamzah\\_15430.pdf](http://repository.akprind.ac.id/sites/files/conferenceproceedings/2012/hamzah_15430.pdf) [Accessed 11 Desember2014].
- [3] Andreas Daniel Arifin, Isye Arieshanti, Agus Zainal Arifin, <http://digilib.its.ac.id/public/ITS-paper-20008-5108100132-Paper.pdf> [Accessed 23 Januari 2015]
- [4] Arvian Pandu Pratama, Dimas dan Bambang Widjanarko (2013). Klasifikasi Kesejahteraan Rumah Tangga di Jawa Timur dengan Pendekatan Multivariate Adaptive Regression Spline Bootstrap Aggregating (MARS Bagging). Surabaya:ITS.
- [5] Bekti Maryuni Susanto, <http://lppm3.bsi.ac.id/jurnal/index.php/biangmatika/article/viewFile/1/12> [Accessed 11 Desember2015].
- [6] Dicky Surya Dwi Putra (2012) Analisis dan Komparasi Metode Naive Bayes dan Logistic Regression dengan Seleksi Variabel Berbasis Genetic Algorithm untuk Prediksi *Software Defect*. Jakarta : Stimik Eresha.
- [7] Edy. (2012). Penerapan Algoritma C4.5 Dengan Seleksi Atribut Berbasis Algoritma Genetika Dalam Diagnosa Penyakit Jantung. Jakarta : Stimik Ersha.
- [8] Evicienna.(2011). Penerapan Algoritma C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Hasil Pemilihan Legislatif DPRD DKI Jakarta. Jakarta :Stimik Nusa Mandiri
- [9] Henny Leidiyana (2011) Komparasi algoritma Klasifikasi *Data Mining* dalam Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. Jakarta : Stimik Eresha
- [10] Klasifikasi dan Jenis jenis Kemiskinan, <http://ahmadefendy.blogspot.com/2010/04/klasifikasi-dan-jenis-jenis-kemiskinan.html> [Accessed 8 Desember2014].
- [11] Kesejahteraan Adalah Kondisi, <http://www.slideshare.net/janroi/kesejahteraan-adalah-kondisi> [Accessed 1 Maret 2015].
- [12] Kemiskinan lengkap, <http://abstraksiekonomi.blogspot.com/2013/12/kemiskinan-lengkap-devinisi-pola-jenis.html> [Accessed 8 Desember2014].