

KLASIFIKASI RUMAH TANGGA MISKIN DI KABUPATEN JOMBANG DENGAN PENDEKATAN RANDOM FOREST CART

Dian Seftiana, Oktiva Dhani Arleina, Giriesa
Kinanti S., Rizka Amalia Dewi S., Fachrunisah

1Jurusan Statistika, Fakultas Matematika dan Ilmu
Pengetahuan Alam, Institut Teknologi Sepuluh
Nopember Surabaya

email: dianadasefty10@gmail.com

Abstract

Poverty is a major cause of a number of social problems, political, economic and particularly in developing countries. Poverty reduction is the key factor that must be done to achieve the targets in the Millennium Development Goals (MDGs) by 2015 government assistance program for the reduction of poverty that exist today have not been able to reduce the number of poor households in Jombang, whereas Jombang have a positive economic growth rate from year to year. Thus, in order to support effective and efficient government then carried out a classification based on the household assistance package that is expected by the poor households in Jombang. Poor households as the object of observation is divided into two according to their poverty status, namely Poor Households (RTM) and Chronically Poor Households (RTSM). Classification is done by using the CART approach is then combined with the method of the Random Forest to obtain a more accurate and stable classification. Classification results using Random Forest CART stated that the expected assistance package by RTSM determined by income per month, place of disposing of fecal water, type of floor building, and source of drink water. While, the most important characteristic of RTM classification income per month, place of disposing of fecal water, source of drink water, and large of house building. The classification result from Random Forest CART method produces a total accuracy rate of 99.50 percent for RTSM and 98.33 percent for RTM. The mapping of households in Jombang stated that cash money, cattle, and modal were the most expected assistance package.

Keywords: CART, Chronically Poor Households, Poor Households, Random Forests, Total Accuracy Rate.

1. PENDAHULUAN

Kemiskinan merupakan salah satu persoalan mendasar yang menjadi pusat perhatian di setiap negara. Kemiskinan menjadi penyebab utama dari sejumlah problem sosial, politik, dan ekonomi yang terjadi khususnya di negara-negara berkembang termasuk Indonesia. Oleh karena itu, penanggulangan kemiskinan menjadi puncak dari solusi program Millenium Development Goals (MDGs) 2015 (Lubis, 2010). Kabupaten Jombang yang merupakan salah satu kabupaten di Provinsi Jawa Timur yang memiliki laju pertumbuhan ekonomi positif dan cenderung mengalami peningkatan. Pada tahun 2009 laju pertumbuhan ekonomi Kabupaten Jombang mencapai 5,28 persen dan meningkat menjadi 6,12 persen pada tahun 2010 (BPS, 2011). Laju pertumbuhan ekonomi di Kabupaten Jombang ini lebih cepat dibandingkan dengan wilayah lainnya di Jawa Timur. Namun, kondisi ini tidak diimbangi dengan menurunnya tingkat kemiskinan di Kabupaten Jombang. Jumlah penduduk miskin di Kabupaten Jombang tetap tinggi dengan persentase sebesar 16,43 persen dari total penduduk pada tahun 2008 dan mencapai 17,53 persen dari total penduduk pada tahun 2009 (Bappeda, 2011). Oleh karena itu, terdapat indikasi adanya permasalahan dalam implementasi program penanggulangan kemiskinan dari pemerintah. Berdasarkan hal tersebut, berbagai program bantuan pemerintah perlu dikaji lebih mendalam agar realisasinya berdampak positif bagi peningkatan kesejahteraan masyarakat miskin. Sehingga, dilakukan analisis klasifikasi rumah tangga miskin di Kabupaten Jombang menurut paket bantuan rumah tangga yang diharapkan dengan didasarkan pada faktor-faktor yang menjadi indikator kemiskinan meliputi aspek kesehatan, sosial, dan ekonomi.

Metode yang umum digunakan dalam proses klasifikasi adalah analisis diskriminan dan regresi logistik multivariat. Metode ini memiliki keterbatasan dalam hal pemenuhan

asumsi dan kesederhanaan interpretasi. Oleh karena itu, untuk tujuan pengklasifikasian digunakan metode nonparametrik yang umum digunakan dan mampu mengatasi keterbatasan kemampuan klasifikasi tersebut seperti metode *Classification and Regression Trees* (CART) (Rumiati & Sutikno, 2005). Metode CART kemudian dikombinasikan dengan metode *ensemble* yaitu *Random Forest* yang mampu meningkatkan kestabilan pohon klasifikasi yang terbentuk dan akurasi prediksi yang dihasilkan. Penelitian yang ada sebelumnya yaitu Muttaqin (2013) masih menggunakan jumlah sampel yang sedikit dan Hidayanti (2013) mengkategorikan bantuan menjadi dua kategori saja dengan hasil ketepatan (akurasi) klasifikasi yang masih rendah. Oleh karena itu, dalam penelitian dilakukan analisis mengenai klasifikasi rumah tangga miskin di Kabupaten Jombang menurut paket bantuan rumah tangga yang diharapkan dengan didasarkan pada indikator-indikator kemiskinan menggunakan pendekatan *Random Forest* CART yang dapat berguna sebagai rekomendasi bagi pemerintah untuk menyusun strategi program penanggulangan kemiskinan di Kabupaten Jombang.

2. METODE

Penelitian ini menggunakan metode CART yang merupakan salah satu metode *Machine Learning* dimana metode eksplorasi data dilakukan dengan teknik pohon keputusan. Metode yang dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J. Stone ini merupakan teknik klasifikasi dengan algoritma penyekatan rekursif biner dimana pemilahan dilakukan pada sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul/node menjadi dua simpul anak (Lewis, 2000). CART akan menghasilkan pohon klasifikasi jika variabel respon berupa data kategorik, sedangkan jika variabel respon berupa data kontinu akan dihasilkan pohon regresi (Breiman, Friedman, Olshen, dan Stone, 1993). Algoritma CART secara umum melalui tiga tahapan yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi dengan menggunakan *cost complexity minimum* dengan fungsi berikut

$$R_{\alpha}(T) = R(T) + \alpha |\tilde{T}|$$

dengan $R(T)$ adalah *resubstitution estimate* (proporsi kesalahan pada sub pohon), α adalah kompleksitas parameter dan $|\tilde{T}|$ merupakan ukuran banyaknya simpul terminal pohon T . Lalu, dilakukan penentuan pohon klasifikasi optimum dengan penduga *Cross Validation 10-Fold Estimate* sehingga diperoleh hasil klasifikasi akhir yang menjadi model pohon terbaik dengan nilai kesalahan relatif yang minimum.



Gambar 1. Diagram Alir Metode Penelitian

Metode CART kemudian dikombinasikan dengan metode *ensemble* yaitu *Random Forest* yang dikembangkan dengan harapan untuk meningkatkan akurasi klasifikasi dari *classifier* tunggal. Konsep dasar *Random Forest* adalah dengan menggunakan banyak *classifier* dari metode yang sama dan kemudian mengkombinasikannya melalui proses *voting* untuk memperoleh dugaan klasifikasi akhir.

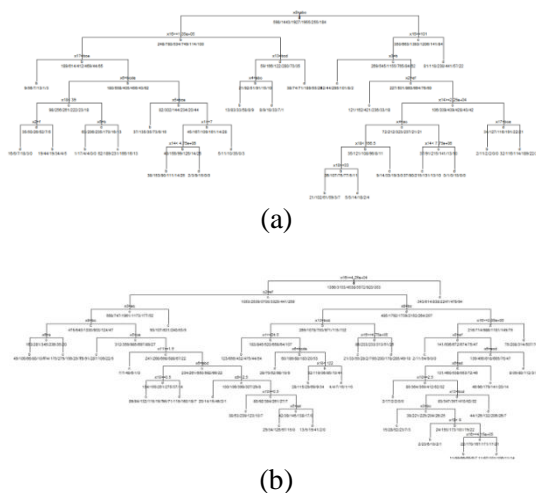
Liaw dan Wiener (2002) menyatakan bahwa *Random Forest* dikembangkan oleh Breiman pada tahun 2001 dengan ide bahwa perlu ada penambahan *layer* dari proses

pengacakan pada *bagging*. Selain data sampel yang diambil secara acak untuk membentuk pohon klasifikasi, pada *Random Forest* variabel prediktor juga diambil secara acak saat penentuan pemilah. Proses tersebut diharapkan dapat mengatasi permasalahan *overfitting*.

Adapun metode pelaksanaan yang digunakan, secara garis besar digambarkan pada diagram alir seperti Gambar 1.

3. HASIL DAN PEMBAHASAN

Hasil deskripsi terhadap 41.181 rumah tangga miskin di Jombang adalah sebagian besar rumah tangga miskin (33,1 persen) memiliki lantai tanah, 45,9 persen memiliki dinding tembok, 35,1 persen tidak memiliki jamban, 55,3 persen mengkonsumsi air sumur, 25,8 persen membuang air tinja di septictank, 43,4 persen menggunakan listrik bukan meteran (menumpang), 47,8 persen masih menggunakan kayu bakar, 74,4 persen makan tiga kali sehari, 50,2 persen tidak pernah mengkonsumsi protein, 61,1 persen membeli pakaian hanya sekali dalam setahun, 66,5 persen berobat ke rumah sakit/puskesmas, 41,5 persen kepala rumah tangga tidak memiliki ijazah pendidikan, 79,2 persen memiliki atap rumah dari genteng, dan 66,5 persen telah memiliki rumah sendiri.



Gambar 2. (a) Konstruksi Pohon Klasifikasi Optimal untuk RTSM (b) Konstruksi Pohon Klasifikasi Optimal untuk RTM

Analisis Klasifikasi CART diawali dengan proses pembentukan pohon klasifikasi maksimal. Hasil analisis CART menyatakan

bahwa pohon klasifikasi maskimal yang diperoleh memiliki tingkat kedalaman (*depth*) sebesar 35 dengan jumlah *terminal node* sebanyak 271 *nodes* untuk klasifikasi RTSM, sedangkan pohon klasifikasi maksimal hasil klasifikasi RTM memiliki tingkat kedalaman sebesar 60 dengan jumlah *terminal node* sebanyak 692 *nodes*. Pohon klasifikasi yang dihasilkan tersebut berukuran sangat besar sehingga dilakukan pemangkasan pohon klasifikasi maksimal menggunakan metode *10-fold cross-validation estimate* untuk menghindari terjadinya kasus *under/over fitting* dan untuk memudahkan dalam proses interpretasi pohon klasifikasi. Proses ini menghasilkan pohon klasifikasi optimal yang memiliki *error (cost)* yang paling kecil. Konstruksi pohon klasifikasi optimal untuk RTSM dan RTM tampak dalam Gambar 2.

Pohon klasifikasi untuk RTSM yang terpilih adalah pohon optimal dengan *terminal nodes* berjumlah 22 simpul, nilai *complexity parameter* sebesar 0,001588, *crossval error* paling minimum sebesar 0,851827, dan *relative error* sebesar 0,820967. Sedangkan, pohon klasifikasi RTM yang terpilih adalah pohon dengan *terminal nodes* berjumlah 28 simpul, nilai *complexity parameter* sebesar 0,000917, *crossval error* paling minimum sebesar 0,898166, dan *relative error* sebesar 0,870753.

Konstruksi pohon klasifikasi yang diperoleh berguna untuk mengetahui penciri dari suatu hasil pengklasifikasian. Berdasarkan konstruksi pohon klasifikasi optimal diketahui bahwa penciri utama yang dapat digunakan untuk menentukan paket bantuan yang dibutuhkan oleh RTSM adalah jenis bahan bakar memasak yang digunakan, sedangkan untuk RTM adalah penghasilan tiap bulan. Dengan menelusuri pohon klasifikasi dimulai dari simpul awal (*parent node*) hingga simpul akhir (*terminal node*) diperoleh suatu hasil karakteristik rumah tangga miskin pada masing-masing paket bantuan yang dibutuhkan. Hasil karakteristik RTSM dan RTM dari *terminal node* yang paling kuat mengklasifikasikan dijelaskan dalam Tabel 1.

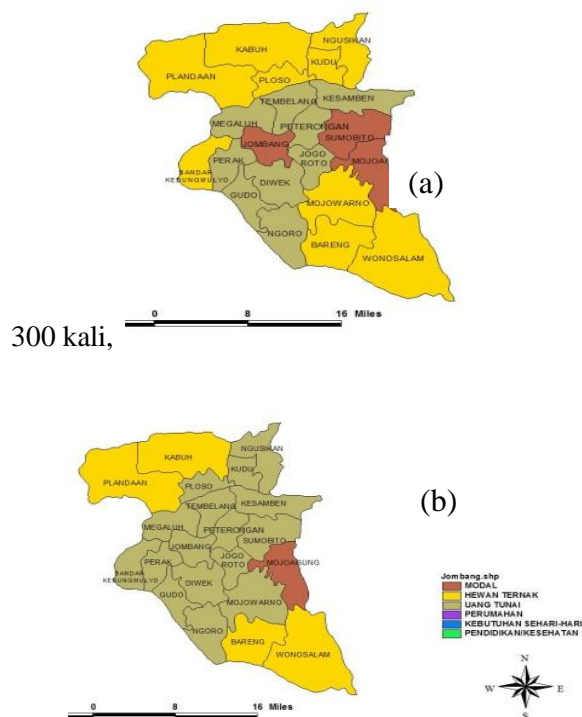
Tabel 1. Karakteristik RTSM dan RTM Menurut Paket Bantuan

Bantuan	Karakteristik Rumah Tangga
RTSM	
Modal	Memasak dengan listrik/elpiji/minyak tanah, penghasilan ≥ 135.000 , rumah milik sendiri/sewa/ rumah dinas, Air tinja dibuang selain di <i>septic tank</i> , luas kavling ≥ 39 m ² , air minum dari ledeng
Hewan Ternak	Memasak menggunakan arang kayu atau kayu bakar, penghasilan tiap bulan ≥ 101 , dinding rumah dari kayu
Uang Tunai	Memasak menggunakan arang kayu atau kayu bakar, penghasilan tiap bulan ≥ 101 , dinding rumah dari tembok atau bambu, lantai rumah dari keramik/ubin/ semen/kayu, memiliki aset $\geq 7.750.000$, tidak memiliki jamban/memiliki jamban atas kepemilikan bersama
RTM	
Modal	Penghasilan ≥ 205.000 , lantai dari keramik/ubin/semen /kayu, memasak menggunakan listrik/arang/kayu bakar, air minum dari air dalam kemasan/ledeng/pompa/sumur, air tinja dibuang di septic tank/sungai/lubang tanah, dalam seminggu mengkonsumsi protein lebih dari 2 kali
Hewan Ternak	Penghasilan ≥ 205.000 , lantai dari keramik/ubin/semen /kayu, memasak menggunakan listrik/arang/kayu bakar, air minum dari mata air/air sungai
Uang Tunai	Penghasilan ≥ 42.500 , lantai dari bambu/tanah, dinding dari tembok/bambu, memasak menggunakan listrik/arang/kayu bakar, air minum dari ledeng/air sungai, membeli pakaian dalam setahun maksimal sekali, air tinja dibuang di sungai/tanah lapang(kebun), makan dalam sehari lebih dari dua kali, dalam seminggu tidak pernah mengkonsumsi protein

Hasil klasifikasi dengan CART tidak stabil (Sutton, 2005), sehingga perlu dilakukan kombinasi dengan metode *ensemble* seperti *Random Forest*. Metode *Random Forest* melakukan pembentukan pohon dari data hasil *resampling* sebanyak

kemudian prediksi klasifikasi dari masing-masing pohon dikombinasikan untuk mendapatkan klasifikasi akhir yang merupakan hasil *voting* mayoritas.

Hasil *Random Forest* CART menyatakan bahwa penentuan bantuan yang dibutuhkan RTSM berdasarkan pada penghasilan tiap bulan, tempat pembuangan air tinja, jenis lantai bangunan, dan sumber air minum. Sedangkan, karakteristik paling penting untuk klasifikasi RTM adalah penghasilan tiap bulan, tempat pembuangan air tinja, sumber air minum, dan luas kavling bangunan. Klasifikasi yang dihasilkan dari penggunaan *Random Forest* CART memiliki akurasi (*total accuracy rate* atau 1-APER) sebesar 99,50 persen untuk RTSM dan sebesar 98,33 persen untuk RTM. Hasil pemetaan dari prediksi klasifikasi seluruh rumah tangga di Jombang ditunjukkan dalam Gambar 3. Persebaran RTM di Kabupaten Jombang tidak memiliki perbedaan yang besar dan bantuan yang paling banyak diinginkan adalah bantuan uang tunai, hewan ternak, dan modal.



Gambar 3. (a) Peta Paket Bantuan yang Diinginkan RTSM di Jombang (b) Peta Paket Bantuan yang Diinginkan RTM di Jombang

4. KESIMPULAN

Kesimpulan yang dapat diambil adalah kondisi ekonomi rumah tangga miskin di Jombang masih rendah karena sebagian besar rumah tangga masih menggunakan kayu bakar untuk memasak, lantai dari tanah, tidak memiliki jamban, tidak pernah mengkonsumsi protein, dan kepala rumah

tangga tidak memiliki ijazah. Selain itu, berdasarkan hasil klasifikasi dengan *Random Forest* CART dapat diketahui bahwa kriteria penentuan rumah tangga yang membutuhkan bantuan kemiskinan hendaknya didasarkan pada indikator penghasilan tiap bulan, tempat pembuangan air tinja, jenis lantai bangunan, dan sumber air minum untuk RTSM dan indikator penghasilan tiap bulan, tempat pembuangan air tinja, sumber air minum, dan luas kavling bangunan untuk RTM.

Hasil *Random Forest* CART memberikan *total accuracy rate* sebesar 99,50 persen untuk klasifikasi RTSM dan sebesar 98,33 persen untuk klasifikasi RTM. Sementara itu, hasil pemetaan merekomendasikan pemerintah untuk lebih memprioritaskan program penanggulangan kemiskinan berupa pembagian uang tunai, pemberian hewan ternak, dan penciptaan lapangan kerja, serta pemberian modal kerja bagi rumah tangga miskin karena bantuan tersebut merupakan bantuan yang paling banyak dibutuhkan oleh rumah tangga miskin di Kabupaten Jombang.

UCAPAN TERIMAKASIH

Selama pengerjaan penelitian ini, banyak pihak yang terlibat dan turut membantu. Oleh karena itu, kami mengucapkan terima kasih kepada:

1. Dr. Bambang Widjanarko Otok, S.Si., M.Si. selaku dosen pendamping PKM yang telah banyak memberikan bimbingan dan pengarahan kepada kami selama penulisan laporan akhir ini.
2. Orang tua dan saudara kami yang telah banyak memberikan dukungan maupun doa untuk kelancaran pengerjaan penelitian ini.
3. Teman-teman Statistika, ITS yang telah memberikan dorongan dan motivasi untuk menyelesaikan penelitian ini.

5. REFERENSI

- [Bappeda] Badan Perencanaan Pembangunan Daerah. 2011. *Profil Daerah Kabupaten/Kota di Jawa Timur Tahun 2007-2011*. Jombang : Bappeda.
- [BPS] Badan Pusat Statistik. 2011. *Produk Domestik Regional Bruto Kabupaten Jombang 2000-2010*. Jombang : BPS.

- Breiman, L., Friedman, J., Olshen, R., & Stone, C. 1993. *Classification and Regression Trees*. New York : Chapman Hall.
- Hidayanti, A. A. 2013. *Boosting Multivariate Adaptive Regression Spline (MARS) Binary Response untuk Klasifikasi Kemiskinan di Kabupaten Jombang*. Tesis Statistika Institut Teknologi Sepuluh Nopember, Surabaya
- Lewis, R. J. 2000. *An Introduction to Classification and Regression Trees (CART) Analysis*. Annual Meeting of the Society for Academic Emergency Medicine. California, UCLA Medical Center.
- Liaw, A. & Wiener, M. 2002. Classification and Regression by Random Forests. *R News*, 2, 18-22.
- Lubis, M. M. 2010. Zuhriatul : Kemiskinan Puncak Solusi Program MDGs. Waspada Online. Diakses pada 20 oktober 2013 dari http://waspada.co.id/index.php?option=com_content&view=article&id=91692:kemiskinan-puncak-solusi-program-mdgs&catid=166:jurnal-puteri-indonesia&Itemid=99
- Muttaqin, M. J. 2013. *Metode Ensemble pada CART untuk Perbaikan Klasifikasi Kemiskinan di Kabupaten Jombang*. Tesis Statistika Institut Teknologi Sepuluh Nopember, Surabaya.
- Rumiati, A.T & Sutikno. 2005. *Studi Pbandingan Metode Klasifikasi Pohon Binary dan Regresi Logistik pada Data Kasus Kesehatan*. Institut Teknologi Sepuluh Nopember, Surabaya.
- Sutton, C. D. 2005. Classification and Regression Trees, Bagging, and Boosting. *Handbook of Statistics*, 24, 303-329.
- [UNDP] United Nations Development Programme (2008). *Millenium Development Goals*. Jakarta : UNDP.