

KLASIFIKASI PENDUDUK TIDAK MAMPU DESA MANDIRAJA WETAN MENGGUNAKAN ALGORITMA C4.5

Asep Muhidin¹⁾, Liswanti²⁾

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
asep.muhammad@pelitabangsa.ac.id

Disetujui, 25 Maret 2019

Abstraksi

Dalam upaya penanggulangan kemiskinan di Kecamatan Mandiraja khususnya Desa Mandiraja Wetan. Pemerintah mempunyai berbagai program bantuan sosial yang digunakan untuk masyarakat kategori miskin (tidak mampu). Bantuan tersebut diberikan berdasarkan indikator yang sudah ditentukan dengan tujuan membantu dan mempermudah petugas dalam mengklasifikasikan kesenjangan penduduk yang mampu dan tidak mampu.

Metode klasifikasi data mining yang digunakan dalam penelitian ini adalah algoritma Decision Tree C4.5. Objek data yang digunakan adalah kumpulan data per kepala keluarga yang menggunakan 6 parameter dalam pengolahan data yaitu pendapatan per bulan, pekerjaan, jumlah anggota keluarga, jenis bangunan, jenis lantai, dan keterangan. Berdasarkan uraian tersebut menghasilkan nilai akurasi dari data testing sebesar 97.22% .

Kata Kunci : Kemiskinan, Algoritma C.45, Data Mining.

Abstract

Poverty reduction efforts in the Special Mandiraja District of Mandiraja Wetan Village. The government has a variety of social assistance programs that are used for people in the poor category. The assistance is provided based on indicators that have been properly determined and help in classifying capable and incapable populations.

The data mining classification method used in this study is the Decision Tree C4.5 algorithm. The data object used is a data collection per head of family that uses 6 parameters in processing data, namely income per month, job, number of family members, type of building, name of the floor and description. Based on the description result in the value of testing data accuracy is 97.22%.

Keyword : poverty, C4.5 Algorithm, Data mining.

1. Pendahuluan

Provinsi Jawa Tengah di BPS pada bulan September 2017 tercatat 12,23 % dan terdata sekitar 4.197,49 jiwa warga kategori tidak mampu. Kabupaten Banjarnegara, dalam data statistika BPS pada tahun 2017 terdapat 0.84 % penduduk miskin. Berdasarkan pemaparan yang ada, perlu upaya untuk menuntaskan kemiskinan di Kabupaten Banjarnegara. Khususnya Desa Mandiraja. Pengetasan kemiskinan di wilayah tersebut diharapkan mampu mengatasi kesenjangan pertumbuhan perekonomian masyarakat miskin. Terkait pengetasan kemiskinan, Pemerintah mempunyai berbagai macam program untuk mengatasi kemiskinan. Akan tetapi, terdapat berbagai kendala dalam penyaluran bantuan dari pemerintah untuk mengklasifikasikan para calon penerima bantuan. Sehingga, membuat proses penyaluran bantuan berjalan lama dan tidak tepat sasaran. Pengetasan kemiskinanpun juga belum bisa dikendalikan. Pada pengumpulan data penduduk yang sudah ada di Desa, dengan memanfaatkan teknik data mining pengklasifikasian dapat ditentukan.

Data mining merupakan metode pengolahan data berskala besar , oleh karena itu data mining memiliki peranan penting dalam berbagai bidang. Secara umum kajian data mining membahas metode–metode seperti klasifikasi, clustering, regresi, seleksi variabel, dan market basket analisis. Teknik Klasifikasi meliputi Naive Bayes, Neural Networks, dan Decision Trees. Teknik klasifikasi akan membantu memudahkan lembaga pemerintahan dalam mengatasi kemiskinan antara kesenjangan penduduk yang mampu dan tidak mampu dengan mengolah data menjadi sebuah informasi.

2. Tinjauan Studi

2.1. Algoritma

Menurut (Munir, 2012), Algoritma adalah urutan logis langkah – langkah penyelesaian masalah yang disusun secara sistematis. Metode pengurutan digambarkan dalam sejumlah langkah terbatas yang mengarah pada solusi permasalahan.

2.2. Data mining

Menurut Chilton, Data mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola – pola baru dari kumpulan – kumpulan data sangat besar, meliputi mode – mode yang merupakan irisan – irisan dari artificial intelligence, machine learning, statistics, dan database system (Suyanto, 2017).

Data mining ditujukan untuk mengekstrak (mengambil intisari) pengetahuan dan sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basisdata dan management data, pemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi dan online updating (Suyanto, 2017).

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut Knowledge Discovery Databases (KDD) (Vulandari, 2017).

2.3. Klasifikasi

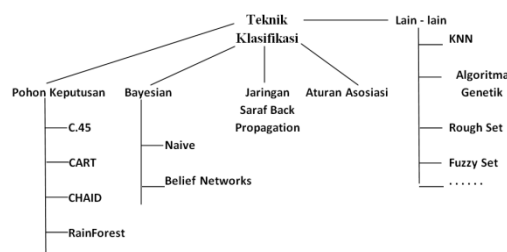
Menurut Zaki et al (2013) klasifikasi didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan . klasifikasi dapat digu nakan dalam berbagai aplikasi, di antaranya adalah deteksi kecurangan (fraud detection), pengelolaan pelanggan, diagnosis medis, prediksi penjualan, dan sebagainya (Suyanto, 2017).

Classification adalah sebuah model dalam data mining dimana classifier dikonstruksi untuk memprediksi categorical label, seperti “aman” atau “beresiko” untuk data aplikasi peminjaman uang;”ya” atau “tidak” untuk data marketing atau “treatment A”, “treatment B”, “treatment C” untuk data medis. Kategori tersebut dapat dipresentasikan dengan nilai yang sesuai dengan kebutuhannya. Dimana pengaturan dari nilai tersebut tidak memiliki arti tertentu (Vulandari, 2017).

Klasifikasi data adalah suatu proses yang menemukan properti - properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya.

Klasifikasi dalam data mining dikelompokkan ke dalam teknik pohon keputusan, Bayesian (Naïve Bayesian dan Bayesian Belief Networks), Jaringan Saraf Tiruan (Backpropagation), teknik yang berbasis konsep dari penambangan aturan-aturan asosiasi, dan teknik lain (k-Nearest Neighbor, algoritma genetik, teknik dengan pendekatan himpunan rough dan fuzzy).

Setiap teknik pohon keputusan memiliki kelebihan dan kekurangannya sendiri, berikut gambar pengelompokan teknik klasifikasi (Jamhur, 2016).



Gambar 2. 1 Pengelompokan Teknik Klasifikasi (Jamhur, 2016)

2.4. Algoritma C45

Algoritma C4.5 adalah algoritma pembentukan pohon keputusan yang merupakan pengembangan dari algoritma pohon keputusan ID3 (Literative Dichotomiser 3) yang diciptakan oleh J.Ross Quinlan pada 1993. Menurut (Whitten, et al, 2011), Algoritma C.45 memiliki keunggulan dibandingkan dengan ID3 yaitu mampu mengatasi nilai yang hilang (missing value), mengatasi data bertipe kontinu dan melakukan

pemangkasan pohon (pruning trees). Dalam konstruksi pohon keputusan data terbagi menjadi sampel pelatihan (training sample) dan sampel pengujian (testing sample). Sampel pelatihan digunakan untuk mengkonstruksi pohon keputusan dan sampel pengujian untuk menguji hasil konstruksi.

Pohon keputusan dalam analisis pemecahan masalah pengambil keputusan adalah pemetaan mengenai alternatif – alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon tersebut juga memperlihatkan faktor–faktor kemungkinan atau probabilitas yang akan mempengaruhi alternatif–alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil keputusan tersebut (Vulandari, 2017).

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Vulandari, 2017) :

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.

2.5. Penduduk Tidak Mampu

Penduduk tidak mampu atau warga miskin adalah suatu kondisi ketidakmampuan secara ekonomi untuk memenuhi standar hidup rata – rata masyarakat di suatu daerah. Kondisi ketidakmampuan ini ditandai dengan rendahnya kemampuan pendapatan untuk memenuhi kebutuhan pokok baik berupa pangan, sandang, maupun papan. Kemampuan untuk memenuhi standar hidup rata – rata seperti standar masyarakat dan standar pendidikan (Suryawati, 2004).

2.6. 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 menggunakan teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat memberi keputusan yang paling baik. RapidMiner kurang lebih memiliki 500 operator data mining termasuk operator untuk input, output, dan preprocessing atau visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner dituliskan dengan Bahasa java sehingga dapat bekerja di semua sistem operasi (Aprilia, dkk, 2012).

RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment) dimana versi awalnya dimulai dikembangkan pada tahun 2001 RalfKlinKemberg, Ingo Mierswa, dan Simon Fircher di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan dibawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai software data mining pada polling oleh Knuggets, sebuah portal datamining pada 2010 – 2011 (Aprilia, dkk, 2012)..

3. Kerangka Konsep

Mandiraja No.16, Kecamatan Mandiraja, Kabupaten Banjarnegara. Adapun jenis data yang digunakan dalam penelitian proposal penelitian ini adalah sebagai berikut :

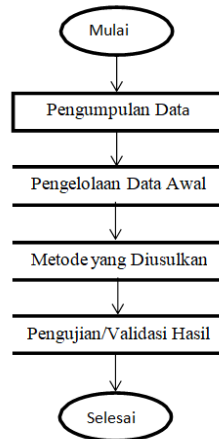
1. Data kualitatif yaitu data yang diperoleh berupa keterangan-keterangan seperti permasalahan kesenjangan kemiskinan yang belum dikelola dengan baik serta solusi untuk mengatasi permasalahan
2. Data kuantitatif yaitu data yang berupa laporan-laporan secara tertulis, seperti data penduduk Desa Mandiraja Wetan.
3. Sumber Data Primer, Dalam penelitian ini peneliti melakukan sumber penelitian langsung kepada kepala desa setempat untuk memperoleh bahan beserta data yang dibutuhkan untuk melakukan penelitian.
4. Sumber Data Sekunder, Data sekunder sebagai pelengkap data primer yang diperoleh dari bentuk file dari kantor kepala desa sebagai teori bahan dataset yang diperoleh, buku-buku, jurnal ilmiah, publikasi, laporan penelitian dan internet

Pada metode ini kegiatan yang dilakukan adalah mempelajari, mencari dan mengumpulkan data-data yang berhubungan dengan penentuan warga tidak mampu. Data yang telah diperoleh kemudian akan diolah menggunakan metode klasifikasi algoritma C4.5 dengan mengambil nilai dari setiap atribut pada data untuk menentukan calon warga yang berhak dikategorikan mampu atau tidak mampu. Berikut proses pengumpulan data yang didapatkan dari kantor Desa Mandiraja Wetan.

4. Desain penelitian/ Metodologi

Pada tahapan penelitian ini dilakukan peneliti dalam mengumpulkan informasi atau data yang akan digunakan, kemudian dianalisa dan diklasifikasi untuk memecahkan permasalahan yang telah dirumuskan dengan tahapan penelitian yang ada.

- 1) Pengumpulan Data
Data penelitian diperoleh dari kantor Kepala Desa Mandiraja Wetan yang berupa data penduduk per kepala keluarga.
- 2) Pengelolaan Data Awal
Pengelolaan awal pada penelitian ini akan mencakup semua kegiatan untuk mempersiapkan data sebelum masuk proses pemodelan . Dalam melakukan pengolahan data awal, akan dilakukan beberapa tahapan agar pada akhirnya akan didapatkan data yang bisa digunakan pada tahap berikutnya. Kemudian dilakukan pembersihan data, menghilangkan menghapus data yang tidak lengkap (outer/noise) untuk memberikankanhasil data yang berkualitas dan informatif. Tahapan tersebut antara lain: cleaning, reduction, transformation serta akan dilakukan split validation.
- 3) Metode yang Diusulkan
Pada tahapan ini dijelaskan pemilihan dan penggunaan metode algoritma C.45.
- 4) Pengujian/Validasi Hasil
Memberikan hasil penentuan warga mampu dan tidak mampu dengan perhitungan atribut yang digunakan.



Gambar 4. 1 *Flowchart* Tahapan Penelitian

5. Hasil Penelitian dan Pembahasan

Hasil pengukuran data accuracy yang diperoleh dari data testing mencapai 97.22%. Dari data tersebut diketahui prediksi Mampu dengan true Mampu mencapai 25 penduduk dan true Tidak Mampu sebanyak 1 penduduk, dengan pencapaian class precision 96.15%. Sedangkan untuk prediksi Tidak Mampu dengan true Mampu sebanyak 0 penduduk dan untuk true Tidak Mampu mencapai 10 penduduk dengan pencapaian class precision 100%. Untuk class recall dengan true Mampu mencapai 100% sedangkan untuk class recall dengan true Tidak Mampu mencapai 90.91%.

Hasil analisa data testing dengan data training pada RapidMiner dapat dilihat pada gambar di bawah ini.

Gambar 5.1 Hasil Prediksi RapidMiner

Hasil penerapan data testing terhadap data training sebagai berikut:

- Jumlah data testing 36
- Jumlah data yang diprediksi benar 35
- Jumlah data yang diprediksi salah 1.

6. Kesimpulan

Dari uraian penelitian yang telah dilakukan sebelumnya dapat ditarik kesimpulan:

1. Dari hasil pengujian data training penduduk Desa Mandiraja Wetan menggunakan algoritma C4.5 maka diperoleh akurasi mencapai 92.22%.
2. Dari hasil analisis data training diperoleh sebuah pohon keputusan yang memiliki *rule-rule* di dalamnya sehingga memudahkan lembaga pemerintahan mengatasi kesenjangan penduduk dalam mengambil keputusan dengan melihat *rule* yang sudah ada.

Daftar Pustaka

Aprilia, dkk. (2012). *Belajar Data Mining Dengan Rapidminer*. Jakarta: Gramedia.

Astuti, & Haryanto. (2015). *Metode Keputusan Menggunakan Algoritma C.45 Untuk Mengelompokan Data Penduduk Pada Tingkatan Kesejahteraan Keluarga*.

Bramditya, S. (2016). *Identifikasi Lama Studi Berdasarkan Karakteristik Mahasiswa Menggunakan Algoritma C.45 (Studi Kasus Lulusan Fakultas Sains Matematika Universitas Diponegoro Tahun 2013/2014)*.

Derick Iskandar, dkk. (2013). *Perbandingan Klasifikasi Kemiskinan Antara Algoritma C.45 dan Naive Bayes Classifier*. Java Jurnal of Electrical and Electronics Engineering.

Dhika, dkk. (2016). *Implementasi Algoritma C.45 Terhadap Kepuasan Pelanggan*. SNaPP.

E. E. (2016). *Penerapan Algoritma C.45 Pada Sistem Pakar Untuk Penentuan Penerima Raskin (Beras Masyarakat Miskin)*.

Elsa, E. (2017). *Analisa dan Penerapan Algoritma C.45 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Konstruksi PT.Arupdhatu Adisesanti*.

Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Penerbit Andi.

Jamhur, A. I. (2016). *Penerapan Data Mining Untuk Menganalisa Jumlah Pelanggan Aktif Dengan Menggunakan Algoritma C.45*. Majalah Ilmiah, 12-20.

Lilik, dkk. (2016). *Metode Keputusan Menggunakan C.45 Untuk Mengelompokan Data Penduduk Pada Tingkatan Kesejahteraan Keluarga*.

Munir. (2012). *Multimedia Konsep & Aplikasi dalam Pendidikan*. Bandung: Alfabeta.

- Nikulin, & M. L. (2008). *Classification Of Imbalanced Marketing Data With Balanced Random Sets*. Australia.
- Suryawati. (2004). *Teori Ekonomi Micro*. Yogyakarta: Jarnasy.
- Suyanto. (2017). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Penerbit Informatika.
- Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta.
- Wahono,R.S. (2017). *Data Mining*. Retrieved from.. Online. [Tersedia]
<http://nugroho.stiemj.ac.id/materi-rpl-2-data-mining/>.
- Wahyudi, D., & Rejekingsih, T. W. (2016). Analisis Kemiskinan di JawaTengah.
- Whitten, et al. (2011). *Data Mining : Practical Machine Learning Tools and Techniques, Third Edition*. Burlingthon: Kaufman Publisher.
- Zaki, & et al. (2013). *Data Mining and Analysis : Fundamental Conceptsand Algorithms*.