

KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES

Haditsah Annur
haditsah@gmail.com
Universitas Ichsan Gorontalo

Abstrak

Permasalahan utama dalam upaya pengurangan kemiskinan saat ini terkait dengan adanya fakta bahwa pertumbuhan ekonomi tidak tersebar secara merata. Penelitian akan melakukan klasifikasi berdasarkan data penduduk miskin yang diperoleh dari Kecamatan Tibawa dengan menggunakan teknik data mining. Atribut yang akan digunakan dalam melakukan klasifikasi penduduk adalah Umur, Pendidikan, Pekerjaan, Penghasilan, Tanggungan, Status (Kawin/Belum Kawin). Metode yang akan digunakan adalah metode *Naïve Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam *data mining*. Berdasarkan penelitian yang dilakukan dihasilkan kesimpulan bahwa, Sistem klasifikasi masyarakat miskin di wilayah pemerintahan Kecamatan Tibawa Kab. Gorontalo dapat direkayasa dan Berdasarkan hasil pengujian *confussion matrix* dengan teknik split validasi, penggunaan metode klasifikasi *naïve bayes* terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori *Good*. Sementara nilai *Precision* sebesar 92% dan *Recall* sebesar 86%.

Kata kunci: Tingkat kemiskinan, Data Mining, Klasifikasi, Naïve Bayes

Abstract

The main problem in the current poverty reduction effort is related to the fact that economic growth is not evenly distributed. The research will classify based on the data of poor people obtained from Tibawa District by using data mining technique. Attributes to be used in classifying the population are Age, Education, Work, Income, Dependent, Status (Married / Unmarried). The method to be used is the Naïve Bayes Classifier method, which is one of the classification techniques in data mining. Based on the research, it is concluded that, the classification system of the poor in the administrative area of Tibawa sub-district, Gorontalo regency can be engineered and Based on the result of confusion matrix testing with split validation technique, the use of naïve Bayes classification method to the dataset which has been taken on the research object obtained the level of accuracy 73% or included in the Good category. While the Precision value of 92% and Recall of 86%.

Keywords: Poverty Level, Data Mining, Classification, Naïve Bayes

1. Pendahuluan

Semua Masyarakat miskin merupakan suatu kondisi dimana fisik masyarakat yang tidak memiliki akses ke prasarana dan sarana dasar lingkungan yang memadai, dengan kualitas perumahan dan pemukiman yang jauh dibawah standar kelayakan serta mata pencaharian yang tidak menentu yang mencakup seluruh multidimensi. Penggolongan kemiskinan didasarkan pada suatu standart tertentu yaitu dengan membandingkan tingkat pendapatan orang atau keluarga dengan tingkat pendapatan yang diperlukan untuk memenuhi kebutuhan pokok minimum [1]. Wilayah Kecamatan Tibawa yang terbagi ke dalam 16 Desa, dan Kecamatan Tibawa termasuk Kecamatan yang memiliki penduduk yang masih di bawah taraf hidupnya, yang biasa dikatakan masyarakat miskin. Berdasarkan data masyarakat miskin yang peroleh dari Kecamatan Tibawa Kabupaten Gorontalo yaitu sebagai berikut :

Tabel 1. Data Penduduk Miskin

Tahun	Jumlah Penduduk Miskin
2015	2790
2014	2819
2013	2950

Sumber : Kecamatan Tibawa

Permasalahan utama dalam upaya pengurangan kemiskinan saat ini terkait dengan adanya fakta bahwa pertumbuhan ekonomi tidak tersebar secara merata di seluruh wilayah Indonesia, khususnya di Kecamatan Tibawa, ini dibuktikan dengan tingginya perbedaan pendapatan antar daerah. Selain itu kemiskinan juga merupakan sebuah hubungan sebab akibat (kausalitas melingkar) artinya tingkat kemiskinan yang tinggi terjadi karena rendahnya pendapatan perkapita, pendapatan perkapita yang rendah terjadi karena investasi perkapita yang juga rendah [2]. Salah satu metode *Data Mining* yang bisa digunakan adalah *Association rule discovery* merupakan tugas yang sama dalam *data mining*, dengan pengecualian bahwa tujuan utama dari klasifikasi adalah prediksi label kelas, sedangkan asosiasi aturan penemuan menggambarkan korelasi antara *item* dalam *database* transaksional [3]. Metode yang akan digunakan adalah metode *Naïve Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam *data mining* [4]. Dimana akan dilakukan analisis untuk memperoleh informasi terhadap data lama tingkat kemiskinan. Diharapkan dari penelitian yang dilakukan terhadap sampel data penduduk miskin tersebut dapat diperoleh suatu informasi yang bisa membantu pihak kecamatan untuk merancang strategi dalam meningkatkan kesejahteraan masyarakat. Penelitian akan membuat aplikasi klasifikasi berdasarkan data penduduk miskin yang diperoleh dari Kecamatan Tibawa tahun 2015 dengan menggunakan teknik data mining. Variabel inputan yang akan digunakan dalam melakukan klasifikasi penduduk miskin tahun 2016 adalah Umur, Pendidikan, Pekerjaan, Penghasilan, Tanggungan, Status (Kawin/Belum Kawin), sesuai data yang telah diambil dan sesuai dengan variabel yang akan diinputkan, maka hasil klasifikasinya nanti akan menentukan tingkat kemiskinan seperti : Miskin dan Tidak Miskin. Rumusan Masalah penelitian ini adalah menentukan cara merekayasa sistem untuk klasifikasi masyarakat miskin Menggunakan Algoritma Naive Bayes dan Hasil penerapan Algoritma Naive Bayes untuk klasifikasi masyarakat miskin di kecamatan tibawa. Adapun tujuan dari penelitian ini adalah : Klasifikasi masyarakat miskin di Kecamatan Tibawa Menggunakan Algoritma *Naive Bayes* dan Menerapkan Algoritma *Naive Bayes* klasifikasi masyarakat miskin di Kecamatan Tibawa.

2. Metode

1.1 Klasifikasi

Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [5]. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees*, *Bayesian classifiers/ Naïve Bayes classifiers*, *Neural networks*, *Analisa Statistik*, *Algoritma Genetika*, *Rough sets*, *k-nearest neighbor*, *Metode Rule Based*, *Memory based reasoning*, dan *Support vector machines (SVM)* [6].

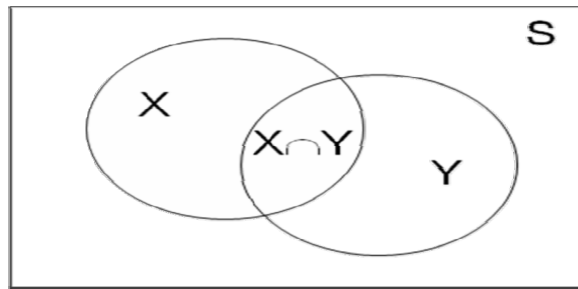
1.2 Algoritma Naïve Bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [7]. Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam Data Mining. Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan :

- X = Data dengan class yang belum diketahui
- H = Hipotesis data X merupakan suatu class spesifik
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)
- P(H) = Probabilitas hipotesis H (prior prob.)
- P(X|H) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas dari X



Gambar 1. Teorema Bayes

3. Hasil dan Pembahasan

3.1 Manual Penggunaan Metode Klasifikasi

Klasifikasi dengan *Naïve Bayes* Menggunakan Data Latih

Berdasarkan dataset/data latih, maka akan dilakukan proses klasifikasi terhadap data baru berikut :

Umur = Tua

Status = Kawin

Pendidikan = SLTP

Tanggungannya = 1 Orang

Pekerjaan = Pedagang

Penghasilan = Tinggi

Proses naïve bayes

Probabilitas Kelas Miskin :

$$P(\text{Miskin}) = 145/171 = 0.847$$

Probabilitas Kelas Tidak Miskin :

$$P(\text{Tidak Miskin}) = 26/171 = 0.152$$

Menghitung kemungkinan termasuk kategori miskin :

$$P(\text{Class.Miskin} \mid \text{Umur.Tua}) = 86 / 145 = 0.59$$

$$P(\text{Class. Miskin} \mid \text{Status.Kawin}) = 122 / 145 = 0.84$$

$$P(\text{Class.Miskin} \mid \text{Pendidikan.SLTP}) = 121 / 145 = 0.83$$

$$P(\text{Class.Miskin} \mid \text{Tanggungannya.1 Org}) = 19 / 145 = 0.13$$

$$P(\text{Class.Miskin} \mid \text{Pekerjaan.Pedagang}) = 1 / 145 = 0.006$$

$$P(\text{Class.Miskin} \mid \text{Penghasilan.Tinggi}) = 0 / 145 = 0$$

Maka kemungkinan termasuk kategori miskin adalah

$$\text{Class.Miskin} = 0.847 \times 0.59 \times 0.84 \times 0.83 \times 0.13 \times 0.006 \times 0$$

$$\text{Class.Miskin} = 0$$

Menghitung kemungkinan termasuk kategori tidak miskin :

$$P(\text{Class.Tidak Miskin} \mid \text{Umur,Tua}) = 13 / 26 = 0.5$$

$$P(\text{Class.Tidak Miskin} \mid \text{Status.Kawin}) = 20 / 26 = 0.77$$

$$P(\text{Class.Tidak Miskin} \mid \text{Pendidikan.SLTP}) = 3 / 26 = 0.11$$

$$P(\text{Class.Tidak Miskin} \mid \text{Tanggungannya.1 Org}) = 2 / 26 = 0.08$$

$$P(\text{Class.Tidak Miskin} \mid \text{Pekerjaan.Pedagang}) = 4 / 26 = 0.15$$

$$P(\text{Class.Tidak Miskin} \mid \text{Penghasilan.Tinggi}) = 8 / 26 = 0.31$$

Maka kemungkinan termasuk kategori tidak miskin :

$$\text{Class.Tidak Miskin} = 0.152 \times 0.5 \times 0.77 \times 0.11 \times 0.08 \times 0.15 \times 0.31$$

$$\text{Class.Tidak Miskin} = 0.00023$$

Karena nilai probabilitas Class.Miskin lebih kecil dari nilai probabilitas Class.Tidak Miskin, maka dapat disimpulkan bahwa data baru diatas termasuk dalam kategori TIDAK MISKIN

3.2 Pengujian Metode Klasifikasi Menggunakan Data Testing

1. Adapun pengujian terhadap metode klasifikasi naïve bayes yang digunakan, dilakukan dengan menggunakan teknik *split validation* dengan *confussion matrix*, dimana dataset yang disajikan diatas akan dibagi kedalam dua bagian yakni 90% (171 record) dari dataset akan dijadikan sebagai data *training* atau latih dan 10% (19 record) sisanya akan dijadikan sebagai data *testing* atau uji.

2. Hasil proses klasifikasi menggunakan metode naïve bayes pada data testing yang berjumlah 19 record dihasilkan proses klasifikasinya sebagai berikut :

Tabel 2. Hasil Klasifikasi Data *Testing*

id	Umur	Status	Pendidikan	Tanggungan	Pekerjaan	Penghasilan	Actual Class	Predicted Class
1	Tua	Kawin	SLTP	1	Pedagang	Tinggi	TIDAK MISKIN	TIDAK MISKIN
2	Muda	Belum Kawin	SD	0	Buruh Lepas	Rendah	MISKIN	MISKIN
3	Tua	Kawin	Sarjana	3	Aparatur Negara	Tinggi	TIDAK MISKIN	TIDAK MISKIN
4	Tua	Kawin	SLTP	4	Petani	Rendah	MISKIN	MISKIN
5	Muda	Kawin	Sarjana	4	Aparatur Negara	Tinggi	TIDAK MISKIN	TIDAK MISKIN
6	Muda	Kawin	SD	3	Petani	Rendah	MISKIN	MISKIN
7	Muda	Kawin	Tidak Sekolah	2	Petani	Sedang	MISKIN	TIDAK MISKIN
8	Tua	Kawin	SLTP	3	Buruh Lepas	Rendah	MISKIN	MISKIN
9	Tua	Cerai	SLTA	2	Tiada	Tiada	MISKIN	MISKIN
10	Muda	Kawin	SD	3	Petani	Rendah	MISKIN	MISKIN
11	Muda	Kawin	SLTP	3	Petani	Sedang	TIDAK MISKIN	TIDAK MISKIN
12	Muda	Belum Kawin	SD	0	Buruh Lepas	Rendah	TIDAK MISKIN	MISKIN
13	Tua	Kawin	SD	1	Tiada	Tiada	MISKIN	MISKIN
14	Muda	Kawin	SD	3	Buruh Lepas	Rendah	MISKIN	MISKIN
15	Muda	Kawin	SD	5	Petani	Sedang	MISKIN	TIDAK MISKIN
16	Tua	Kawin	SD	3	Petani	Rendah	MISKIN	MISKIN
17	Tua	Cerai	SLTA	1	Tiada	Tiada	MISKIN	MISKIN
18	Tua	Cerai	SD	2	Tiada	Tiada	MISKIN	MISKIN
19	Tua	Kawin	SD	3	Buruh Lepas	Rendah	MISKIN	MISKIN

Dari hasil proses klasifikasi yang disajikan pada tabel diatas maka dapat dikonversi kedalam bentuk tabel *confussion matrix* seperti dibawah :

Tabel 3. Pengujian *Confussion Matrix*

19 Record	Tidak Miskin	Miskin	
Actual: Tidak Miskin	4	1	5
Actual: Miskin	2	12	14
	6	13	

Berdasarkan tabel *confussion matrix* diatas maka kinerja dari penggunaan metode klasifikasi naïve bayes dapat diukur dengan menghitung nilai *accuracy*, *precision* dan *recall*.

$$\begin{aligned} \text{Accuracy} &: (\text{TP} + \text{TN}) / \text{Total} \\ &: (12 + 4) / 19 = 73 \% \\ \text{Precision} &: \text{TP} / \text{Predicted Miskin} \\ &: 13 / 13 = 92 \% \\ \text{Recall} &: \text{TP} / \text{Actual Miskin} \\ &: 12 / 14 = 86 \% \end{aligned}$$



3.3 Hasil Klasifikasi pada Sistem

Form di bawah ini menunjukkan Hasil akhir dari sistem, yang merupakan hasil klasifikasi pada setiap objek dengan menggunakan metode yang diusulkan, pada halaman ini juga nantinya dapat dilakukan proses pencetakan dari hasil klasifikasi tersebut sebagai bahan laporan kepada pimpinan/ yang membutuhkan hasil klasifikasi tersebut.

ID	Umur	Status	Pendidikan	Tanggungan	Pekerjaan	Penghasilan	Klasifikasi
1	Tua	Kawin	SD	2	Petani	Rendah	MISKIN
2	Muda	Kawin	SD	3	Buruh Lepas	Rendah	MISKIN
3	Muda	Kawin	SD	2	Buruh Lepas	Rendah	MISKIN
4	Tua	Kawin	SD	3	Buruh Lepas	Sedang	TIDAK MISKIN
5	Muda	Kawin	SD	3	Petani	Sedang	TIDAK MISKIN
6	Tua	Kawin	SD	4	Buruh Lepas	Rendah	MISKIN
7	Tua	Kawin	SD	2	Petani	Rendah	MISKIN

Gambar 2. Halaman Hasil Klasifikasi

3.4 Pengujian *Black-Box*

Tabel 4. Hasil Pengujian *Black-Box*

Input event	Fungsi	Hasil sistem	Hasil uji
Klik menu kelas	Untuk menampilkan form input kelas	Ditampilkan form input kelas	Sesuai
Klik menu parameter	Untuk menampilkan form input parameter / atribut	Ditampilkan form input parameter / atribut	Sesuai
Klik menu sub parameter	Untuk menampilkan form input sub parameter	Ditampilkan form input sub parameter	Sesuai
Klik menu training	Untuk menampilkan form input data latih / training	Ditampilkan form input data latih / training	Sesuai
Klik menu objek	Untuk menampilkan form input objek	Ditampilkan form input objek	Sesuai
Klik menu hasil Klasifikasi	Untuk menampilkan hasil Klasifikasi keseluruhan objek	Ditampilkan hasil Klasifikasi keseluruhan objek	Sesuai
Beri masukan pada form kelas dan klik tombol simpan/save	Untuk menyimpan data kelas yang diketikkan pada database	Data kelas disimpan kedalam database	Sesuai
Klik tombol tambah pada data objek	Untuk menampilkan form input data fakta / uji pada objek yang dimaksudkan	Ditampilkan form input data latih	Sesuai
Diberi masukan pada form input data latih dan klik tombol simpan/save	Untuk menyimpan data fakta/uji untuk objek tersebut	Data fakta/uji disimpan kedalam database	Sesuai

4. Kesimpulan dan Saran

4.1 Kesimpulan

1. Sistem klasifikasi masyarakat miskin di wilayah pemerintahan Kab. Gorontalo dapat direayasa, hal tersebut dapat dibuktikan melalui *interface* yang disajikan dan sistem yang telah direayasa sudah dilakukan pengujian sistem dengan menggunakan *whitebox* untuk memeriksa alur logika yang digunakan pada sistem dan juga telah dilakukan pengujian *blackbox* untuk memeriksa kesesuaian fungsi pada setiap *interface* yang ada.
2. Berdasarkan hasil pengujian *confussion matrix* dengan teknik split validasi, penggunaan metode klasifikasi *naive bayes* terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori *Good*. Sementara nilai *Precision* sebesar 92% dan *Recall* sebesar 86%. Berdasarkan hal tersebut dapat dinyatakan bahwa sistem klasifikasi yang dibangun dapat digunakan sebagai bahan masukan bagi pengambil keputusan

4.2 Saran

1. Mengingat nilai akurasi masih berada pada angka 73%, maka masih sangat dimungkinkan untuk dapat dilakukan penelitian selanjutnya untuk meningkatkan nilai akurasi dengan menambahkan fitur seleksi atau penggunaan Algoritma komputer yang lain.
2. Pada penelitian lain diharapkan dapat digunakan dataset dalam jumlah yang lebih besar atau dengan sejumlah variabel lainnya guna meningkatkan performa dari metode yang digunakan.

5. Terima Kasih

Terwujudnya publikasi karya ilmiah ini tidak lepas dari bantuan berbagai pihak, yang telah memberikan dana penelitian melalui Hibah Penelitian Kompetitif di lingkungan Universitas Ichsan Gorontalo, oleh karena itu penulis mengucapkan terima kasih yang sebesar-besarnya kepada : Ketua Yayasan PIPT Ichsan Gorontalo, Rektor Universitas Ichsan Gorontalo, Ketua Lembaga Penelitian Unisan Gorontalo dan Sekretaris Lemlit Unisan Gorontalo. Semoga hasil publikasi karya ilmiah ini, dapat bermanfaat bagi peneliti dan masyarakat.

Daftar Pustaka

- [1] Sumanta, Jaka. 2005. *Fenomena lingkaran kemiskinan di Indonesia : Analisis ekonometri regional data panel propinsi tahun 1999-2002*. MPKP UI.
- [2] Suryawati. 2004. *Teori Ekonomi Mikro*. UPP. AMP YKPN. Yogyakarta
- [3] Ayub, Mewati. (2007). "Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer", *Jurnal Sistem Informasi* Vol. 2 No. 1 Maret 2007 : 21-30
- [4] Mustafa, Muhammad Syukri. Simpen, I Wayan. (2014). *Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar)*, ISSN: 2354-5771
- [5] Kursini, Luthfi, E. T., 2009, *Algoritma Data Mining*, Andi Offset, Yogyakarta.
- [6] Jananto, Arief. 2013. *Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa*.
- [7] Supriyanto, Catur. Purnama Parida. 2013. *deteksi penyakit diabetes type ii dengan naive bayes berbasis particle swarm optimization*. *Jurnal Teknologi Informasi*, Volume 9 Nomor 2, Oktober 2013
- [8] Kusumadewi, Sri dkk. 2006. *Fuzzy Multi Attribute Decision Making (FMADM)*. Yogyakarta. Graha Ilmu.
- [9] Bambang Hariyanto, (2004), *Sistem Manajemen Basis Data*, Informatika, Bandung
- [10] Jeffrey, L. Whitten, et al. 2004. *Metode Desain dan Analisis Sistem*. Edisi I. Diterjemahkan oleh tim penerjemah ANDI. Yogyakarta: Penerbit Andi Madcoms.
- [11] Jogiyanto, HM., 2005, *Analisis dan Desain Sistem Informasi : Pendekatan Terstruktur Teori dan Praktek Aplikasi Bisnis*, Yogyakarta : Andi