

PERBANDINGAN ALGORITMA NAIVE BAYES VALIDASI 2 DAN 3 PADA KLASIFIKASI KELUARGA MISKIN DI KABUPATEN BANJAR

Fitria

Program Studi DIII Teknik Informatika, Politeknik Hasnur

fitria.acnh@gmail.com

ABSTRAK

Indonesia merupakan salah satu negara berkembang yang memiliki tingkat pertumbuhan penduduk yang sangat pesat dengan jumlah total populasi sekitar 255 juta penduduk, Indonesia adalah negara berpenduduk terpadat nomor empat di dunia. Terkait hal tersebut, penulis ingin mencoba menerapkan salah satu algoritma dalam *data mining* yaitu *Naive Bayes* dengan membandingkan validasi 2 dan 3 ke dalam data yang diperoleh dari BPS Kabupaten Banjar, karena sebelumnya tidak ada penelitian yang terkait dengan tingkat kesejahteraan masyarakat di Kota Martapura menggunakan metode tersebut, sehingga dapat diketahui tingkat akurasi, presisi, *recall* dan AUC dari algoritma yang diusulkan dan secara tepat dapat menentukan tingkat kemiskinan di daerah tersebut dengan mengumpulkan dan mengolah data, serta secara cepat dapat mencari solusi untuk mengurangi tingkat kemiskinan dan dapat mencegah bertambahnya angka kemiskinan di daerah tersebut. Penerapan validasi 2 pada Algoritma *Naive Bayes* menghasilkan akurasi sebesar 84,62%, lebih akurat sebesar 0,57% dibandingkan dengan penerapan validasi 3.

Kata Kunci : Kemiskinan, *Naive Bayes*

PENDAHULUAN

Indonesia merupakan salah satu negara berkembang yang memiliki tingkat pertumbuhan penduduk yang sangat pesat dengan jumlah total populasi sekitar 255 juta penduduk, Indonesia adalah negara berpenduduk terpadat nomor empat di dunia. Hal tersebut tentu saja mempengaruhi tingkat kesejahteraan masyarakat dari berbagai aspek kehidupan. Di Indonesia, untuk mengukur kemiskinan Badan Pusat Statistik (BPS) menggunakan konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*).

Menurut data dari Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) sampai dengan tahun 2014, tingkat kemiskinan nasional telah dapat diturunkan menjadi 11,25 persen dari 13,33 persen pada tahun 2010. Keberhasilan dalam menurunkan tingkat kemiskinan di samping diperoleh melalui peningkatan pertumbuhan ekonomi yang berkualitas dan melalui 3 (tiga) klaster program penanggulangan kemiskinan. Keberhasilan penanggulangan kemiskinan

sangat tergantung dari kemampuan mengidentifikasi sasaran penerima manfaat Program Penanggulangan Kemiskinan. Untuk itu, peran Data dan Indikator menjadi faktor yang sangat penting.

Terkait hal tersebut, penulis ingin mencoba menerapkan salah satu algoritma dalam data mining yaitu *Naive Bayes* dengan membandingkan validasi 2 dan 3 ke dalam data yang diperoleh dari BPS Kabupaten Banjar, karena sebelumnya tidak ada penelitian yang terkait dengan tingkat kesejahteraan masyarakat di Kota Martapura menggunakan metode tersebut, sehingga dapat diketahui tingkat akurasi, presisi, *recall* dan AUC dari algoritma yang diusulkan dan secara tepat dapat menentukan tingkat kemiskinan di daerah tersebut dengan mengumpulkan dan mengolah data, serta secara cepat dapat mencari solusi untuk mengurangi tingkat kemiskinan dan dapat mencegah bertambahnya angka kemiskinan di daerah tersebut

LANDASAN TEORI

Kemiskinan diartikan sebagai suatu keadaan dimana seseorang tidak sanggup memelihara dirinya sendiri sesuai dengan taraf kehidupan kelompok dan juga tidak mampu memanfaatkan tenaga mental maupun fisiknya dalam kelompok tersebut. Menurut sejarah, keadaan kaya dan miskin secara berdampingan tidak merupakan masalah sosial sampai saatnya perdagangan berkembang dengan pesat dan timbulnya nilai-nilai sosial yang baru.

Pada masyarakat modern yang rumit, kemiskinan menjadi suatu masalah sosial karena sikap yang membenci kemiskinan tadi. Seseorang bukan merasa miskin karena kurang makan, pakaian atau perumahan tetapi karena harta miliknya tidak cukup untuk memenuhi taraf kehidupan yang ada.

Menurut Teguh Dartanto, profil kemiskinan di Indonesia masih merupakan fenomena pedesaan yang sebagian besar penduduk miskin masih di pedesaan sehingga penanggulangan kemiskinan pedesaan merupakan salah satu upaya strategis dalam mewujudkan kesejahteraan sosial terutama pemenuhan kebutuhan dasar masyarakat miskin dipedesaan.

Kriteria Kemiskinan

Saat ini, Penanganan Fakir Miskin di Indonesia telah diatur dalam perundang-undangan tersendiri, yaitu Undang-Undang Nomor 13 Tahun 2011 tentang Penanganan Fakir Miskin. Lainnya yang berkaitan adalah Undang-Undang Nomor 6 Tahun 2014 tentang Desa. Pada pasal 78 ayat (1) dinyatakan Pembangunan Desa bertujuan meningkatkan kesejahteraan masyarakat Desa dan peningkatan kualitas hidup manusia serta penanggulangan kemiskinan melalui pemenuhan kebutuhan dasar, pembangunan sarana dan prasarana Desa, pengembangan potensi ekonomi lokal, serta pemanfaatan sumberdaya alam dan lingkungan secara berkelanjutan.

Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Algoritma Naive Bayes

Algoritma *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut di asumsikan memiliki atribut saling bebas (independen).

Teorema Bayes memiliki bentuk umum sebagai berikut :

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

Dalam hal ini :

X	= Data dengan class yang belum diketahui
H	= Hipotesis data X merupakan suatu class spesifik
$P(H X)$	= Probabilitas hipotesis H berdasar kondisi X (posteriori probability)
$P(H)$	= Probabilitas hipotesis H (priorprobability)
$P(X H)$	= Probabilitas X berdasar kondisi pada hipotesis H
$P(X)$	= Probabilitas dari X [17]

METODE PENELITIAN

Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data yang berasal dari hasil Sensus Sosial Ekonomi Nasional (Susenas) BPS Kabupaten Banjar Provinsi Kalimantan Selatan pada tahun 2014. Total data anggota rumah tangga di Kota Martapura sebesar 2.076 data. Namun dalam penelitian ini digunakan 2 jenis data, yaitu data original adalah yang belum

melewati tahap *preprocessing* sebesar 2.076 data dan data modifikasi adalah data yang telah melewati tahap *preprocessing* sebesar 1.222 data dengan objek yang diteliti adalah status kesejahteraan keluarga yang termasuk dalam kategori (1) rumah tangga sangat miskin dan (2) rumah tangga miskin. Variabel dan kategori yang akan digunakan dalam penelitian ini dijelaskan pada tabel berikut ini :

Tabel 1. Penjelasan Variabel dan Kategori

Variabel	Keterangan	Skala	Kategori
Y	Status Rumah Tangga Sasaran (RTS)	Nominal	1 : Sangat Miskin (RTSM)
			2 : Miskin (RTM)
X1	Jenis Kelamin KRT	Nominal	1 : Laki-Laki
			2 : Perempuan
X2	Umur KRT	Numerik	-
X3	Pendidikan KRT	Nominal	1 : Tingkat Dasar
			2 : Tingkat Menengah
			3 : Tingkat Atas
X4	Lapangan Usaha KRT	Nominal	1 : Pertanian (Padi & Palawija)
			2 : Hortikultura
			3 : Perkebunan
			4 : Perikanan Tangkap
			5 : Perikanan Budidaya
			6 : Peternakan
			7 : Kehutanan & Pertanian Lain
			8 : Pertambangan / Penggalan
			9 : Industri Pengolahan
			10 : Listrik dan Gas
			11 : Bangunan / Konstruksi
			12 : Perdagangan
			13 : Hotel dan Rumah Makan
			14 : Transportasi dan Pergudangan
			15 : Informasi dan Komunikasi
			16 : Keuangan dan Asuransi
			17 : Jasa Pendidikan / Jasa Kesehatan / Jasa Kemasyarakatan, Pemerintahan dan Perorangan
			18 : Lainnya
X5	Status Kependudukan	Nominal	1 : Berusaha Sendiri

	dalam Pekerjaan Kepala Rumah Tangga		2 : Buruh / Karyawan / Pegawai Swasta 3 : Pekerja Keluarga / Tidak dibayar
X6	Status Penguasaan Bangunan Tempat Tinggal	Nominal	1 : Milik Sendiri 2 : Kontrak / Sewa 3 : Lainnya
X7	Jenis Atap Terluas	Nominal	1: Genteng 2 : Sirap 3 : Ijuk / Rumbia 4 : Lainnya
X8	Kualitas Atap	Nominal	1 : Bagus / Kualitas Sedang 2 : Jelek / Kualitas Rendah
X9	Jenis Dinding Terluas	Nominal	1 : Tembok 2 : Kayu 3 : Lainnya
X10	Kualitas Dinding	Nominal	1 : Biasa / Kualitas Sedang 2 : Jelek / Kualitas Rendah
X11	Jenis Lantai	Nominal	1 : Bukan Tanah / Bambu 2 : Tanah 3 : Bambu
X12	Sumber Air Minum	Nominal	1 : Air Terlindung 2 : Air Tidak Terlindung
X13	Bahan Bakar Utama Memasak		1 : Listrik / Gas / Elpiji 2 : Lainnya
X14	Sumber Penerangan	Nominal	1 : Listrik PLN 2 : Listrik Non-PLN 3 : Tidak Ada Listrik
X15	Jumlah Keluarga	Numerik	-
X16	Jumlah Individu	Numerik	-

Dataset pada penelitian ini memiliki 1 (satu) variabel sebagai kelas yaitu rumah tangga sangat miskin (RTSM) dan rumah tangga miskin (RTM). Di dalam dataset ini terdapat 16 variabel yang berfungsi sebagai atribut yang memiliki tipe data nominal, namun ada beberapa variabel yang memiliki tipe data berbeda yaitu atribut umur, jumlah keluarga dan jumlah individu yang memiliki tipe data numerik.

Metode Pengolahan Data Awal

Setelah dikumpulkan, data tersebut diproses dengan cara dihilangkan/ dihapuskan, diidentifikasi, dan digantikan.

Penelitian ini menghasilkan angka-angka yang dapat diukur dan nantinya akan dinilai untuk menentukan apakah metode ini cocok dan menghasilkan tingkat akurasi yang tinggi untuk klasifikasi status keluarga di Kota Martapura. Apabila dari hasil penelitian diperoleh tingkat akurasi yang tinggi, maka angka akurasi tersebut dapat dinyatakan cocok untuk melakukan

klasifikasi dan akan menjadi salah satu faktor menyatakan kesimpulan.

Metode yang diusulkan

Pada penelitian ini metode yang diusulkan adalah *Naive Bayes* dengan membandingkan penerapan validasi 2 dan 3 pada algoritma tersebut untuk mengklasifikasi status keluarga di Kota Martapura sehingga nantinya akan diperoleh tingkat akurasi penggunaan metode tersebut.

Eksperimen dan Pengujian Metode

Eksperimen dan pengujian data pada metode yang telah diusulkan akan dilakukan menggunakan *RapidMiner*. Eksperimen dan pengujian tersebut akan dilakukan dengan dua tahap, yang pertama data diuji menggunakan data original dengan menerapkan algoritma *Naive Bayes 2-validation* dan tahap kedua data tersebut diuji kembali dengan menerapkan algoritma *Naive Bayes 3-validation* yang hasilnya nanti akan dibandingkan.

Evaluasi dan Validasi Hasil

Evaluasi dan validasi hasil akan diperoleh setelah proses eksperimen dan pengujian dilaksanakan. Pada tahap ini tingkat akurasi dari hasil eksperimen dan pengujian pada data menggunakan *RapidMiner* dengan menerapkan metode *Naive Bayes* validasi 2 dan 3 yang hasilnya akan dibandingkan sehingga nantinya akan diketahui akurasi terbaik dan metode mana yang lebih akurat dalam melakukan klasifikasi status keluarga miskin khususnya di Kota Martapura.

HASIL DAN PEMBAHASAN

Pengujian 2-Validation

Pengujian dimulai dari tahap pertama yang dilakukan dengan menggunakan *2-validation* dengan menerapkan metode *Naive Bayes*.

Pengujian dengan menerapkan metode *Naive Bayes* validasi 2 menghasilkan data sesuai tabel 2 di bawah ini :

Tabel 2. Hasil Pengujian Naive Bayes Validasi 2

Naive Bayes 2 Validation Data Original			
accuracy: 82.76% +/- 0.29% (mikro: 82.76%)			
	true 2.0	true 1.0	class precision
pred. 2.0	1579	225	87.53%
pred. 1.0	133	139	51.10%
class recall	92.23%	38.19%	
AUC: 0.835 +/- 0.018 (mikro: 0.835) (positive class: 1.0)			

Pengujian data dengan menerapkan metode *Naive Bayes* validasi 2 menghasilkan akurasi sebesar 82,76% dengan nilai *precision* sebesar 87,53%, *recall* sebesar 92,23% dan *AUC* sebesar 0,835.

Pengujian 3-Validation

Pengujian data dengan menerapkan metode *Naive Bayes* validasi 3 menghasilkan data sesuai tabel 3 di bawah ini :

Tabel 3. Hasil Pengujian Naive Bayes Validasi 3

Naive Bayes 3 Validation Data Original			
accuracy: 82.51% +/- 0.82% (mikro: 82.51%)			
	true 2.0	true 1.0	class precision
pred. 2.0	1578	229	87.33%
pred. 1.0	134	135	50.19%
class recall	92.17%	37.09%	
AUC: 0.862 +/- 0.002 (mikro: 0.862) (positive class: 1.0)			

Pengujian data dengan menerapkan metode *Naive Bayes* validasi 3 menghasilkan akurasi sebesar 82,51% dengan nilai *precision* sebesar 87,33%, *recall* sebesar 92,17% dan *AUC* sebesar 0,862.

Evaluasi dan Validasi Hasil

Pengukuran penelitian dilakukan dengan membandingkan hasil pengujian data dengan menerapkan algoritma *Naive Bayes* validasi 2 dan 3 sehingga dapat diketahui hasil akurasi dan nilai *AUC* dan akhirnya dapat diketahui nilai akurasi terbaik dan model dengan metode mana yang paling akurat dalam melakukan klasifikasi status kesejahteraan rumah tangga. Pada tahap pengujian terhadap data dengan metode tersebut dilakukan dengan teknik *x validation* dengan pengujian data menggunakan *2-validation* sehingga dapat di evaluasi hasilnya dengan membandingkan dan mengukur seberapa keakuratan akurasi yang dihasilkan dari beberapa percobaan tersebut.

Hasil validasi pengujian algoritma *Naive Bayes* terhadap masing-masing data tersebut kemudian dievaluasi dan

dibandingkan sehingga memperoleh hasil seperti pada tabel 4 berikut :

Tabel 4. Perbandingan Hasil Pengujian

Pengujian	A	B	Perbandingan
Validasi	2	3	
Accuracy	84,62%	84,05%	0,57%
Precision	88,06%	87,64%	0,42%
Recall	94,30%	94,10%	0,20%
AUC	0,5847 222	0,5854 167	-0,07%

Pada tabel 4 terlihat bahwa secara umum pengujian *Naive Bayes* menggunakan *2-validation* pada data lebih akurat, karena memiliki nilai yang lebih tinggi dibandingkan daripada pengujian menggunakan *3-validation*.

KESIMPULAN DAN SARAN

Kesimpulan

Dari perbandingan hasil penelitian yang telah dilakukan dari tahap awal hingga hasil pengujian diperoleh, dapat disimpulkan bahwa penerapan validasi 2 pada Algoritma *Naive Bayes* menghasilkan akurasi sebesar 84,62%, lebih akurat sebesar 0,57% dibandingkan dengan penerapan validasi 3.

Saran

Penelitian merupakan uji coba untuk mengetahui dan menilai tingkat akurasi penerapan metode tertentu, dalam hal ini *Naive Bayes* pada klasifikasi keluarga miskin di Kota Martapura. Agar penelitian ini dapat memberikan kontribusi yang lebih besar, diperlukan pengembangan metode dengan cara

menerapkan *Naive Bayes* dalam suatu alat bantu (berupa *software*) yang dalam penelitian ini menggunakan *RapidMiner*, sehingga dapat langsung diterapkan untuk penyelesaian masalah dalam mengidentifikasi keluarga miskin di Kota Martapura dengan cepat, tepat dan mudah.

DAFTAR PUSTAKA

- TNP2K, "Data dan Indikator Tim Nasional Percepatan Penanggulangan Kemiskinan," [Online]. Available: <http://www.tnp2k.go.id/id/data-indikator/mengenai-data-indikator/>. [Accessed 20 September 2016].
- T. Dartanto, 28 Maret 2014. [Online]. Available: http://www.bbc.com/indonesia/majalah/2014/03/140327_bisnis_kemiskinan_profil.shtml.
- S. A, "Analisis Efektifitas Usaha Bersama Sebagai Instrumen Program Penanganan Fakir Miskin," *Sosio Inf*, vol. 2, pp. 54-55, 2016.
- Kusrini and E. T. Luthfi, *Algoritma Data Mining*, Yogyakarta: Andi, 2009.
- D. Iskandar and Y. K. Suprpto, "Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan antara Algoritma C4.5 Dan Naive Bayes Clasifier," *JAVA Journal of Electrical and Electronics Engineering*, vol. 11, p. 14, 2013.
- D. H. A, "Penerapan Particle Swarm Optimization untuk Seleksi Atribut pada Klasifikasi Dokumen Berita Menggunakan Naive Bayes Classifier," 2014.
- E. Junianto, "Penerapan Particle Swarm Optimization untuk Seleksi Fitur pada Klasifikasi Dokumen Berita Menggunakan Naive Bayes Classifier," Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Jakarta, 2014.