ANALISIS ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI STATUS KESEJAHTERAAN RUMAH TANGGA KELUARGA BINAAN SOSIAL

Erfan Karyadiputra, S.Kom, M.Kom
(erfantsy@gmail.com)

ABSTRAK

Kemiskinan merupakan salah satu permasalahan yang sering dihadapi dalam upaya peningkatan kesejahteraan di hampir semua negara. Tersedianya data kemiskinan yang akurat dan berkesinambungan merupakan salah satu instrumen penting untuk mengevaluasi kebijakan pemerintah dalam mengentaskan kemiskinan dengan memfokuskan perhatian pada pendistribusian bantuan sesuai rumah tangga sasaran (RTS). Penelitian terkait klasifikasi kesejahteraan rumah tangga sering menggunakan variabel target/kelas berupa kategori miskin dan tidak miskin. Kategori tersebut jika dilihat dari aspek pendistribusian bantuan masih bersifat umum, hal tersebut karena kategori rumah tangga miskin tersebut dapat diklasifikasikan lagi kedalam status kesejahteraan rumah tangga sesuai rumah tangga sasaran (RTS) sehingga dalam pendistribusian bantuan dapat disesuaikan dengan status kesejahteraan rumah tangga sasaran (RTS). Oleh sebab itu diperlukan variabel kelas baru yang sesuai RTS yaitu sangat miskin dan miskin. Dalam penelitian ini akan dilakukan pengujian menggunakan algoritma Naive Bayes untuk mengklasifikasi status kesejahteraan rumah tangga miskin yaitu rumah tangga sangat miskin (RTSM) dan rumah tangga miskin (RTM). Hasil pengujian yang didapatkan adalah algoritma Naive Bayes menghasilkan akurasi sebesar 85.80% dan nilai AUC sebesar 0.930.

Kata Kunci : *Data Mining, Kemiskinan, Naive Bayes , Non-Monetary*

PENDAHULUAN

Kemiskinan di Indonesia saat ini telah menjadi masalah yang harus ditanggulangi secara tuntas karena sebagaimana diamanatkan dalam Pembukaan Undang – Undang Dasar 1945. Tersedianya data kemiskinan yang akurat, dan berkesinambungan merupakan salah satu instrumen penting bagi pengambil kebijakan dalam memfokuskan perhatian pada pendistribusian bantuan sesuai rumah tangga sasaran (RTS).

Penelitian terkait klasifikasi kesejahteraan rumah tangga sering menggunakan variabel target/kelas berupa kategori miskin dan tidak miskin. Kategori tersebut jika dilihat dari aspek pendistribusian bantuan masih bersifat umum, hal tersebut karena kategori rumah tangga miskin tersebut dapat diklasifikasikan lagi kedalam status kesejahteraan rumah tangga sesuai rumah tangga sasaran (RTS) sehingga dalam pendistribusian bantuan dapat disesuaikan dengan status kesejahteraan rumah tangga sasaran (RTS).

Berdasarkan uraian permasalahan diatas, diperlukan analisis proses klasifikasi status kesejahteraan rumah tangga menggunakan variabel

target/kelas yang sesuai dengan kategori rumah tangga sasaran (RTS) yaitu rumah tangga miskin (RTM) dan kelompok rumah tangga sangat miskin/fakir miskin (RTSM) berdasarkan aspek *non-monetary* menggunakan metode klasifikasi data mining yaitu *naïve bayes.*

RUMUSAN MASALAH

Berdasarkan permasalahan diatas maka rumusan masalah penelitian ini adalah klasifikasi status kesejahteran RTS (Rumah Tangga Sasaran) Keluarga Binaan Sosial mengunakan *Naive Bayes*. Sehingga "Seberapa besar akurasi metode *Naive Bayes* untuk klasifikasi status kesejahteran RTS (Rumah Tangga Sasaran) Keluarga Binaan Sosial ?"

TUJUAN PENELITIAN

Berdasarkan latar belakang dan rumusan masalah diatas, maka penelitian ini bertujuan untuk menerapkan metode dari algoritma *Naive Bayes* dalam klasifikasi status kesejahteran RTS (Rumah Tangga Sasaran) Keluarga Binaan Sosial.

TARGET LUARAN DAN MANFAAT PENELITIAN

Target luaran dalam pembuatan laporan penelitian ini adalah untuk memfasilitasi tercapainya tingkat akurasi dari data PPLS dengan menggunakan metode *naïve bayes* sehingga didapatkan hasil yang akurat dalam pengelompokkan data.

Manfaat dari penelitian ini adalah:

1. Manfaat Praktis

   Hasil penelitian ini diharapkan dapat digunakan untuk membantu mengidentifikasi kesejahteraan rumah tangga sasaran keluarga binaan sosial.
2. Manfaat Teoritis

   Hasil penelitian ini diharapkan dapat memberikan sumbangan untuk pengembangan teori yang berkaitan metode klasifikasi data mining untuk klasifikasi kesejahteraan rumah tangga khususnya metode *Naive Bayes.*
3. Manfaat Kebijakan

   Hasil penelitian ini diharapkan dapat memberikan masukan pada instansi terkait dalam rangka mengambil kebijakan untuk mengevaluasi kebijakan dalam upaya mengentaskan kemiskinan rumah tangga sasaran.

METODE PENELITIAN

Beberapa metode dalam penelitian ini diantaranya adalah

1. Pengumpulan Data
2. Pengolahan Data
3. Eksperimen Dan Pengujian

1. Pengumpulan Data

Dataset yang digunakan merupakan data sekunder hasil pendataan PPLS 2011 Provinsi Kalimantan Selatan, Kecamatan Rantau Badauh Barito Kuala. *Dataset* tersebut memiliki 1 variabel sebagai kelas yaitu status rumah tangga sangat miskin (RTSM) dan status rumah tangga miskin (RTM) dan 16 variabel sebagai atribut. Sebagian besar variabel atribut bertipe data *nominal* kecuali atribut umur, jumlah keluarga dan jumlah individu. Variabel dan kategori yang digunakan pada penelitian ini adalah :

VAR			KATEGORI
Y	Status Kesejahteraan	1	Kelompok 1 (Paling Miskin)
		2	Kelompok 2 (Miskin)

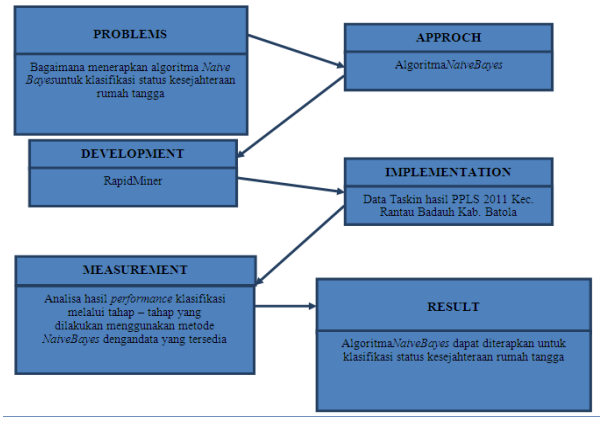
X_1	Jenis Kelamin Kepala Rumah Tangga	1	Laki-laki
		2	Perempuan
X_2	Umur Kepala Rumah Tangga	-	
X_3	Jumlah Keluarga	1	Satu Keluarga
		2	Dua Keluarga
		3	Lebih dari Dua Keluarga
X_4	Jumlah Anggota Keluarga	1	1 Orang
		2	2 Orang
		3	3 Orang
		4	4 Orang
		5	Lebih dari 4 Orang
X_5	Pendidikan Kepala Rumah Tangga	0	Tidak Punya Ijazah
		1	SD/Sederajat
		2	SMP/Sederajat
		3	SMA/Sederajat
		4	Perguruan Tinggi
X_6	Lapangan Usaha Kepala Rumah Tangga	1	Pertanian (Padi & Palawija)
		2	Hortikultura
		3	Perkebunan
		4	Perikanan Tangkap
		5	Perikanan Budidaya
		6	Peternakan
		7	Kehutanan & Pertanian Lain
		8	Pertambangan / Penggalian
		9	Industri Pengolahan
		10	Listrik & Gas
		11	Bangunan / Konstruksi
		12	Pedagang
		13	Hotel & Rumah Makan
		14	Transportasi & Pergudangan
		15	Informasi & Komunikasi
		16	Keuangan & Asuransi
		17	Jasa Pendidikan/ Jasa Kesehatan/ JasaKemasyarakatan, Pemerintahan dan Perorangan
		18	Lainnya
X_7	Status Kependudukan dalam Pekerjaan Kepala Rumah Tangga	1	Berusaha Sendiri
		2	Berusaha dibantu Buruh tidak tetap / tidak dibayar
		3	Berusaha dibantu Buruh tetap/dibayar
		4	Buruh/ Karyawan/ Pegawai Swasta
		5	Pekerja Bebas
		6	Pekerja Keluarga / Tidak dibayar
X_8	Status Penguasaan Bangunan Tempat Tinggal	1	Milik Sendiri
		2	Kontrak/ Sewa
		3	Lainnya
X_9	Jenis Atap Terluas	1	Beton
		2	Genteng
		3	Sirap
		4	Seng
		5	Asbes
		6	Ijuk/ Rumbia
		7	Lainnya
X_{10}	Kualitas Atap	1	Bagus/ Kualitas Tinggi
		2	Biasa/ Kualitas Sedang
		3	Jelek/ Kualitas Rendah
		1	Tembok
X_{11}	Jenis Dinding Terluas	2	Kayu
		3	Bambu
		4	Lainnya
X_{12}	Kualitas Dinding	1	Bagus/ Kualitas Tinggi
		2	Biasa/ Kualitas Sedang
		3	Jelek/ Kualitas Rendah
X_{13}	Jenis Lantai	1	Bukan Tanah/ Bambu/ Kayu
		2	Tanah
		3	Bambu
		4	Kayu
X_{14}	Sumber Air Minum	1	Air Kemasan
		2	Air Ledeng
		3	Air Terlindung
		4	Air Tidak Terlindung
X_{15}	Sumber Penerangan Utama	1	Listrik PLN
		2	Listrik Non-PLN
		3	Tidak ada Listrik
X_{16}	Bahan Bakar Utama Memasak	1	Listrik/ Gas/ Elpiji
		2	Minyak Tanah
		3	Kayu
		4	Lainnya

2. Pengolahan Data

Setelah pengumpulan data maka data tersebut kemudian diolah agar dapat diproses dalam *data mining* karena dalam *data mining*, attribut yang kurang lengkap tidak dapat ditangani secara sistematis oleh algoritma *data mining.*

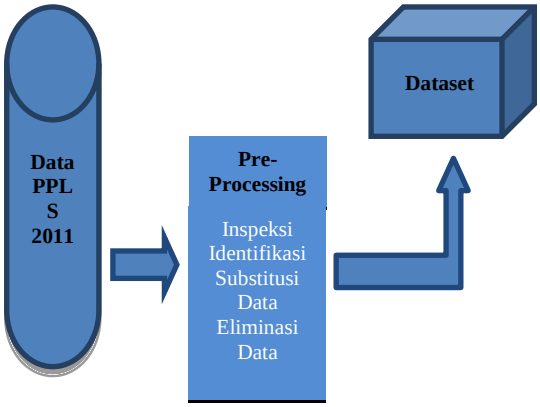
3. Eksperimen Dan Pengujian

Beberapa tahapan dalam eksperimen, yaitu (A) Melakukan pengujian klasifikasi algoritma *Naïve bayes* menggunakan data original yang masih terdapat data kosong (B) Mengisi data kosong menggunakan *replace missing value* (C) Memisahkan dataset menjadi data *training* dan data *testing* (D) Mengevaluasi ketepatan klasifikasi *Naïve Bayes* berupa evaluasi dan validasi hasil.



ANALISA HASIL DAN PEMBAHASAN

Sebelum melakukan eksperimen dan pengujian model/metode maka data yang dikumpulkan terlebih dahulu diolah agar dapat diproses dalam *data mining*. Data yang terkumpul kemudian dilakukan *pre-processing* dengan cara melakukan identifikasi dan inspeksi sejumlah atribut-atribut yang tidak relevan. Kemudian setelah diketahui sejumlah atribut yang tidak relevan maka tahap selanjutnya adalah mengeliminasi sejumlah atribut tersebut dan melakukan subtitusi data sehingga data dapat diproses dalam *data mining*.



Dataset akan diproses dengan metode *Naïve Bayes* dengan beberapa tahapan sehingga perlu menentukan besarnya nilai fakta umum dari data yaitu P(Y = sangat miskin) dan P(Y = miskin). Banyaknya data label status kesejahteraan bernilai sangat miskin adalah 16 dari 30 sampel data yang digunakan sehingga dapat ditentukan nilai P(status kesejahteraan = sangat miskin) sedangkan banyaknya data pada label status kesejahteraan bernilai miskin adalah 14 dari 30 sampel data sehingga dapat ditentukan nilai P(status kesejahteraan = miskin).

Perhitungan fakta umum dari sampel 30 data		Hasil dalam decimal
P(Y = sangat miskin)	$\frac{16}{30}$	0,53
P(Y = miskin)	$\frac{14}{30}$	0,46

Hasil perhitungan fakta umum yaitu P(Y=sangat miskin) adalah 0,53 sedangkan P(Y=miskin) adalah 0,46. Selanjutnya menghitung berdasarkan rumus *NaiveBayes* atau HMAP data *training* dengan hipotesa berdasarkan nilai probabilitas kondisi prior yang diketahui.

Diketahui :	Y	X
Asumsi Variabel Data PPLS 2011	Y = Status Kesejahteraan KRT 1 = sangat miskin/RTSM 2 = miskin/RTM	X1 = Jenis Kelamin X2 = Umur KRT X3 = Pendidikan KRT X4 = Lapangan Usaha KRT X5 = SKP KRT X6 = SPB KRT X7 = Jenis Atap X8 = Kualitas Atap X9 = Jenis Dinding X10 = Kualitas Dinding X11 = Jenis Lantai X12 = Sumber Air Minum X13 = Bahan Bakar Memasak X14 =Sumber Penerangan X15 = Jumlah Keluarga X16 = Jumlah Individu

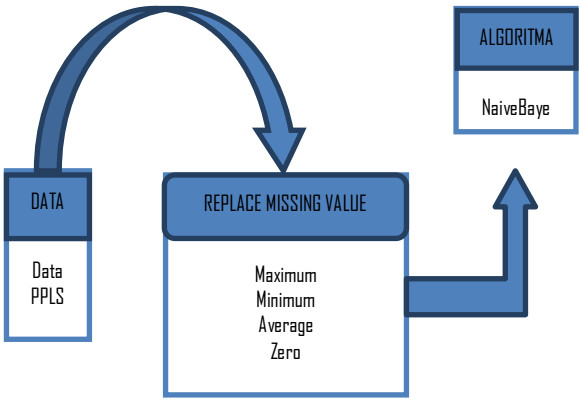
Hasil Fakta Umum	P (Y = Sangat Miskin)	P (Y = Miskin)
	0,53	0,46
P (X1 = Jenis Kelamin, X2 = Umur KRT, X3 = Pendidikan KRT, X4 = Lapangan Usaha KRT, X5 = Status Kependudukan Pekerjaan, X6 = Status Penguasaan Bangunan, X7 = Jenis Atap, X8 = Kualitas Atap, X9 = Jenis Dinding, X10 = Kualitas Dinding, X11 = Jenis Lantai, X12 = Sumber Air Minum, X13 = Bahan Bakar Memasak, X14 =Sumber Penerangan, X15 = Jumlah Keluarga, X16 = Jumlah Individu | Y = Status Kesejahteraan)		
Rumus HMAP (NaiveBayes)	$P(S \mid X) = \underset{x \in X}{\operatorname{argmax}} \frac{P(Y \mid X)\,P(X)}{P(X)}$ $= \underset{x \in X}{\operatorname{argmax}}\, P(Y \mid X)\,P(X)$	

Hasil yang didapatkan diketahui perbandingan hasil keputusan *Naïve Bayes* dengan data original yaitu 29 data teridentifikasi akurat dan hanya satu yang tidak akurat.

Pada tahap pertama eksperimen dan pengujian model dilakukan dengan data original yang sebagian masih memiliki data kosong. Kemudian pengujian dilakukan menggunakan algoritma *Naïve Bayes* dengan validasi model klasifikasi dilakukan terhadap data *testing* dengan teknik 10-*folds cross validation.*

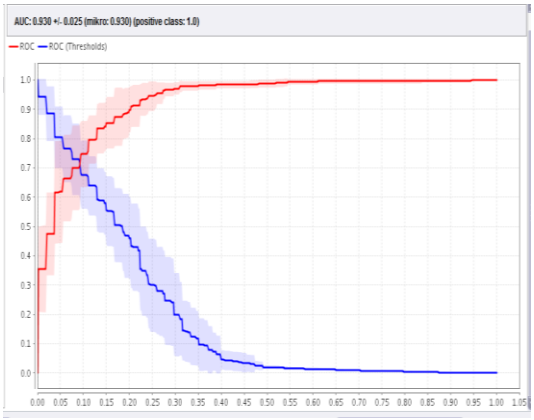
accuracy = 89.79% +/- 2.34% (mikro: 89.79%)			
	TrueMiskin	TrueSangat Miskin	Classprecision
PredMiskin	478	42	91.92%
PredSangat Miskin	63	445	87.60%
Classrecall	88.35%	91.38%	

Dari hasil pengujian tersebut didapatkan akurasi yang sudah cukup baik sekitar 89,79% dengan nilai AUC sebesar 0.959. Namun hasil tersebut belum maksimal karena pada data original yang digunakan terdapat data yang kosong. Oleh sebab itu dilakukan *pre-processing* untuk memastikan data yang akan diolah dalam *data mining* adalah data yang baik dan lengkap sehingga menghasilkan model keputusan yang baik pula. Salah satu teknik yang digunakan metode *replace missing values.*



accuracy = 85.80% +/- 3.44% (mikro: 85.80%)			
	PredMiskin	PredSangat Miskin	Classprecision
PredMiskin	449	54	89.26%
PredSangat Miskin	92	433	82.48%
Classrecall	82.99%	88.91%	

Dari hasil pengujian membuktikan bahwa algoritma *Naïve Bayes* sudah dapat diterapkan untuk mengidentifikasi status kesejateraan rumah tangga miskin dengan akurasi 85.80%. Meskipun mengalami penurunan akurasi setelah dilakukan *replace missing value*, namun hasil tersebut masih lebih baik karena telah menggunakan data yang lengkap daripada pada sebelum dilakukan *pre-processing* sekalipun *Naïve Bayes* merupakan algoritma yang mampu menangani data yang hilang. Kemudian dari hasil pengujian juga dianalisa berdasarkan kurva ROC



Menurut Gorunescu, tingkat keakurasian AUC dapat diklasifikasikan menjadi lima kelompok yaitu :

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

Nilai AUC yang didapatkan dari pengujian berdasarkan kurva ROC menggunakan metode *Naïve bayes* sebesar 0.930 sehingga dari hasil tersebut berdasarkan kriteria diatas menunjukan klasifikasi yang dihasilkan termasuk kedalam kelompok *excellent*

classification atau klasifikasi yang sangat baik.

Pada tahap evaluasi dan validasi model klasifikasi dilakukan pengujian yang dilakukan terhadap data *testing* dengan metode tersebut dilakukan dengan teknik *cross validation* dengan pengujian data mulai 2,3,4,5,6,7,8,9 dan 10 sehingga dapat di evaluasi hasilnya dengan mengukur seberapa keakuratan akurasi yang dihasilkan dari beberapa percobaan tersebut menggunakan metode *Naïve Bayes*. Namun sebelum melakukan percobaan tersebut, terlebih dahulu akan dilakukan percobaan dengan mengganti jenis *parameter* pada *X-validation* yaitu *sampling type* yang terdiri atas tiga jenis, diantaranya *linear sampling*, *shuffled sampling*, dan *stratified sampling* sehingga didapatkan *sampling type* yang terbaik dan sesuai untuk digunakan pada data yang diuji.

Sampling Type	Linear Sampling	Shuffled Sampling	Stratified Sampling
Accuracy	83.18	85.41	85.80
AUC	0.931	0.937	0.930

Hasilnya penggunaan *X-Validation* dengan *sampling type stratified* memiliki tingkat akurasi yang sedikit lebih baik daripada *sampling type* yang lain meskipun dalam hal kehandalan klasifikasi masih sedikit lebih unggul penggunaan *sampling typestratified*.

Kemudian percobaan selanjutnya akan dilakukan pengujian algoritma *Naïve Bayes* dengan teknik *folds cross validation* dengan pengujian data mulai 2,3,4,5,6,7,8,9 dan 10.

Val	2	3	4	5	6	7	8	9	10
Acc	85.70	85.80	86.28	85.50	85.70	85.80	86.19	85.60	85.80
Pre	82.96	82.44	83.31	82.12	82.47	82.96	83.21	82.63	82.75
Rec	87.89	89.33	88.92	89.12	88.91	88.31	89.13	88.71	88.90
AUC	0.937	0.935	0.940	0.927	0.938	0.935	0.935	0.938	0.930

Dari hasil pengujian didapatkan pengulangan pengujian sebanyak 4 kali dengan hasil pengukuran berupa nilai rata-rata 4 kali pengujian memiliki tingkat akurasi tertinggi sebesar 86.28% dengan nilai AUC sebesar 0.940. Namun hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa penggunaan 10-*fold cross-validation* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat. Hal tersebut karena 10-*fold cross-validation* akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian sehingga dalam evaluasi hasil perbandingan menggunakan pengukuran 10 kali didapatkan akurasi sebesar 85.80% dan pengujian *performance* dengan menggunakan *Confusion Matrix* didapatkan nilai AUC sebesar 0.930. maka hasilnya algoritma *Naïve Bayes* tergolong kedalam *Excellent Classification* dalam mengklasifikasi datsa status kesejahteraan rumah tangga.

PENUTUP

1.KESIMPULAN

Dari hasil penelitian yang dilakukan dari tahap awal hingga pengujian, dan hasil perbandingan dapat disimpulkan bahwa model yang terbentuk dengan menggunakan algoritma *Naïve Bayes* menghasilkan akurasi yang cukup baik yaitu sebesar 85.80 %. Berdasarkan kehandalan dalam klasifikasi berupa nilai AUC yang didapat dari algoritma *Naïve Bayes* adalah 0.930 sehingga tergolong sebagai *Excellent Classification* sehingga dapat disimpulkan algoritma *Naïve Bayes* dapat diterapkan untuk melakukan klasifikasi status kesejahteraan rumah tangga.

2. Saran

Pada penelitian ini hanya menggunakan 2 variabel target/kelas dari RTS (Rumah Tangga Saran) yaitu rumah tangga sangat miskin dan rumah tangga miskin sehingga untuk penelitian selanjutnya diharapkan dapat mencoba menggunakan 3 variabel target sesuai RTS (Rumah Tangga Saran) dengan menambahkan variabel target/kelas rumah tangga hampir miskin.

DAFTAR PUSTAKA

[1] UUD, Undang-Undang Dasar Negara Republik Indonesia 1945 [Amandemen], surabaya: pustaka agung harapan, 2002.

[2] M. Jhingan, ekonomi pembangunan dan perencanaan, jakarta: Raja grafindo persada, 2004.

[3] Bappenas, "Indikator Kemiskinan Indonesia," 31 5 2013. [Online]. Available: http://ewalmart.blogspot.com. [Diakses 20 3 2015].

[4] T. Dartanto, "Profil Kemiskinan Indonesia," 28 maret 2014. [Online]. Available: http://www.bbc.co.uk/indonesia/majalah/2014/03/140327_bisnis_kemiskinan_profil.shtml. [Diakses 9 september 2014].

[5] Perpres, Perpres No. 15 Tahun 2010 tentang Percepatan Penanggulangan Kemiskinan, 2010.

[6] D. A. p. Pratama, Klasifikasi Kesejahteraan Rumah Tangga di Jawa Timur dengan Pendekatan Multivariate Adaptive Regression Spline Bootstrap Aggregating (MARS Bagging), Surabaya: Institute Teknologi Sepuluh September, 2011.

[7] H. M. G. Prakosa, Klasifikasi Kesejahteraan Rumah Tangga Di Provinsi Jawa Timur Dengan Pendekatan Bootstrap Aggregatting Classification And Regression Trees (CART Bagging), Surabaya: Institute Teknologi Sepuluh September, 2011.

[8] I. Widyandono, Klasifikasi Kesejahteraan Rumah Tangga Di Provinsi Jawa Timur Dengan Pendekatan Cart Arcing [Thesis], Surabaya: Institute Teknologi Sepuluh September, 2010.

[9] Y. Mirawanti, Pebandingan Metode Regresi Logistik Ordinal Dengan Jaringan Syaraf Tiruan Fungsi Radial Basis [thesis], surabaya: INSTITUT TEKNOLOGI SEPULUH NOVEMBER SURABAYA, 2012.

[10] A. Wijaya, Analisis Kemiskinan Di Provinsi Lampung Dengan Pendekatan Analisis Kemiskinan Di Provinsi Lampung Dengan Pendekatan Spatial Autoregressive Model (Linear Contiguity Method) [Thesis], Surabaya: Institute Teknologi Sepuluh September, 2011.

[11] K. Suganda, Metode Analisis Diskriminan untuk klasifikasi

kemiskinan di kabupaten Jawa tengah [thesis], Semarang: Universitas Kristen Satya Wacana, 2009.

[12] Y. Huang dan L. Li, Naive bayes classification algorithm based on small sample set, IEEE Cloud Computing and Intelligence Systems, 2011.

[13] K. A. Susanto, "rekomendasi pilihan sekolah menggunakan metode naïve bayes classification pada penerimaan peserta didik baru," Malang, Universitas Barwijaya Malang, 2013.

[14] Y. A. Bakhtiar, "implementasi metode naïve bayes untuk klasifikasi kenaikan grade karyawan pada fuzzyfikasi data kinerja karyawan," Malang, Universitas Brawijaya Malang.

[15] Kemensos, petunjuk pelaksanaan kelompok usaha bersama (KUBE), jakarta: direktorat penanggulangan kemiskinan pedesaan, 2014.

[16] R. Undang-Undang, Undang-Undang Republik Indonesia Nomor 6 Tentang Ketentuan-Ketentuan Pokok Kesejahteraan Sosial, 1974.

[17] Swasono, Kebersamaan dan Asas Kekeluargaan, Jakarta: UNJ Press, 2004.

[18] Rohidi, Ekspresi Seni Orang Miskin, Bandung: Nuansa, 2000.

[19] BPS, Indikator Kemiskinan; Konsep dan Penghitungan, Jakarta: Badan Pusat Statistik, 2012.

[20] BPS, Indikator Kemiskinan; Konsep dan Penghitungan, Jakarta: Badan Pusat Statistik, 2004.

[21] A. Barrientos, Social Policy and Development Programme, E-paper United Nations Research Institute for Social Development, 2010.

[22] M. Tjokrowinoto, Pembangunan: Dilema dan Tantangan, Jakarta: Pustaka Rajawali, 1996,.

[23] Chambers, Rural Development, Puting The Last First, 1983, p. 113 – 114 .

[24] Bappenas, Indikator Kemiskinan Indonesia, Jakarta: Badan Perencanaan dan Pembangunan Nasional, 2006.

[25] V. Moertini, Data Mining Sebagai Solusi Bisnis, Integral Vol. 7 No. 1, 2002.

[26] R. Bellazzi dan B. Zupanb, Predictive Data Mining In Clinical Medicine: Current Issues And And Guidelines, International Journal Of Medical Informatics, 2008.

[27] J. Han dan M. Kamber, Data Mining : Concepts and Techniques, Second penyunt., M. R. Jim Gray, Penyunt., San Francisco: Morgan Kaufmann Publishers, 2007.

[28] W. Ian H. dan E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2 penyunt., San

Francisco: Morgan Kaufmann Publishers is an imprint of Elsevier, 2005.

[29] F. Gorunescu, Data Mining: Concept, Models and Techniques, Romania: Springer, 2010.

[30] Larose, Data Mining Methods And Models, Canada: John Wiley & Sons, Inc, 2006.

[31] M. Dunham, Data Mining Introuctory and Advanced Topics, New Jersey: Prentice Hall, 2003.

[32] R. IT, "data mining data preprocessing," 8 3 2010. [Online]. Available: https://rencanait.wordpress.com. [Diakses 1 2 2015].

[33] D. Aprilla, D. A. Baskoro, L. Ambarwati dan I. W. S. Wicaksana, "Belajar Data Mining dengan RapidMiner," Jakarta, academia.edu, 2013.

[34] C. Sammut dan G. Web, Encyclopedia of machine learning, New York: Springer, 2011.

[35] Zhang dan Wang, Application of Bayesian Method to Spam SMS Filtering, International Conference on Information Engineering and Computer Science, 2009.

[36] C. Vercellis, Business Intelligence:Data Mining and Optimization for Decision Making, 2009.

[37] Lewis dan R. J, An Introduction to Classification And Regression Trees (CART) Analysis, Presented at the 2000, 2000.

[38] R. Eubank, Spline Smoothing and Nonparametric Regression,, New York: Marcel Deker, 1998.

[39] X. Wu dan V. Kumar, The Top Ten Algorithms in Data Mining, Taylor ; Francis Group, LLC, 2009.

[40] Suharto dan N. Birdsal, Membangun Masyarakat Memberdayakan Masyarakat, Bandung: Refika Aditama, 2005.

[41] D. S. D. Putra, Analisis Dan Komparasi Metode Naive Bayes Dan Logistic Regression Dengan Seleksi Variabel Berbasis Genetic Algorithm Untuk Prediksi Software Defect, Jakarta: Sekolah Tinggi Manajemen Informatika Dan Komputer Eresha , 2012.

[42] A. D. R. Prabowo, Prediksi Nasabah Yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization, Semarang: Universitas Dian Nuswantoro.

[43] F. Handayanna, Penerapan Particle Swarm Optimization Untuk Seleksi Atribut Pada Metode Support Vector Machine Untuk Prediksi Penyakit Diabetes, Jakarta: Sekolah Tinggi Manajemen Informatika Dan Komputer Nusa Mandiri , 2012.

[44] J. Zurada dan K. N. Kunene, "Comparisons of the Performance of Computational Intelligence Methods for Loan Granting Decisions," dalam *Proceedings of the 44th Hawaii International Conference on System Sciences 2011*, Hawaii, 2011.

[45] M. Mahmoud dan A. A. Najla Algadi, "Expert System for Banking Credit Decision," dalam *International Conference on Computer Science and Information Technology 2008*, 2008.

[46] W. Yogi Yusuf, "Perbandingan Performance Algoritma Decision Tree C5.0, CART, Dan CHAID: Kasus Prediksi Status Resiko Kredit Di Bank X," dalam *Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2007)*, Yogyakarta, 2007.

[47] L. Daniel T, Discovering knowledge in data: An Introduction to Data Mining, Wiley Interscience, 2005.

[48] S. Sani dan S. Dedy, Pengantar Data Mining: Menggali Pengetahuan Dari Bongkahan Data, Yogyakarta: Penerbit Andi, 2010.

[49] W. Lipo dan F. Xiuju, Data Mining with Computational Intelligence, Verlag Berlin Heidelberg: Springer, 2005.

[50] L. Lin dan N. Huang, "Credit Risk Assessment Using BP Neural Network with Dempster-Shafer Theory," *International Conference on Artificial Intelligence and Computational Intelligence*, pp. 397-401, November 2009.

[51] C. Zhu, Y. Zhan dan S. Jia, "Credit Risk Identification of Bank Client Basing on Supporting Vector Machines," IEEE, 2010.

[52] J. A. Hall, Accounting Information Systems (Sistem Informasi Akuntasi), 4nd penyunt., South-Western, 2004.

[53] D. Olson dan Y. S. , Intoduction to Business Data Mining, McGraw-Hill, Penyunt., New York, 2007.

[54] O. Maimon, Data Mining And Knowledge Discovery Handbook, London: Springer, 2010.