JURNAL TEKNIK KOMPUTER AMIK BSI

PENERAPAN NAÏVE BAYES BERBASIS GENETIC ALGORITHM UNTUK PENENTUAN KLASIFIKASI DONOR DARAH

Hilda Amalia[1]

***Abstract*— Blood Donors is an important activity undertaken by every human being. Noble blood donor activity and can be good for itself and others. Many things can lead a person need the help of others in this respect is the blood such as accidents, surgery and others. Meeting the needs of blood must be properly managed; it aims to facilitate people who need blood. Data processing donor and perform good management of the data of blood donors is important. Find behaviour patterns donors so as to obtain blood stocks that meet. It is important for the assessment of the likelihood of someone donating blood back resulting in a classification of blood donors. In this research will be to improve the accuracy of naïve Bayes using a genetic algorithm. From this research, the accuracy value generated by naïve Bayes method is 74.07%, and accuracy are produced by methods that increase the accuracy of genetic algorithm with naïve Bayes as many as 76.48%**

I*ntisari*— Donor Darah merupakan suatu kegiatan yang penting dilakukan oleh setiap manusia. Donor darah kegiatan yang mulia dan dapat berdampak baik bagi sipendonor itu sendiri dan orang lain. Banyak hal yang dapat mengakibatkan seseorang memerlukan pertolongan orang lain dalam hal ini adalah darah seperti kecelakaan, operasi dan lain-lain. Pemenuhan kebutuhan darah harus dikelola dengan baik, hal ini bertujuan untuk mempermudah masyarakat yang memerlukan darah. Pengolahan data pendonor dan melakukan pengelolaan yang baik terhadap data-data donor darah menjadi penting. Menemukan pola perilaku para pendonor sehingga dapat diperoleh stok darah yang memenuhi. Untuk itu penting dilakukan penilaian mengenai kemungkinan seseorang mendonorkan darahnya kembali sehingga menghasilkan klasifikasi donor darah. Dalam penelitian ini akan dilakukan peningkatan akurasi naïve bayes dengan menggunakan genetic algorithm. Dari penelitian ini diperoleh bahwa nilai akurasi yang dihasilkan oleh metode naïve bayes yaitu 74,07%, dan akurasi yang dihasilkan oleh metode peningkatan akurasi yaitu genetic algoritm dengan naïve bayes yaitu sebanyak 76,48%

Keyword: donor darah, data mining, naïve bayes

[1] *AMIK BSI Jakarta, Jl. RS Fatmawat No. 24, Jakarta. Telp (021)7500282; e-mail: hilda.ham@bsi.ac.id.*

Kata Kunci: Sistem Irigasi, Mikrokontroler, Wavecom

I. PENDAHULUAN

Donor darah merupakan suatu aktivitas yang bertujuan untuk menyelamatkan hidup manusia. donor darah merupakan kegiatan medis dasar yang diperlukan untuk hampir semua kegiatan medis[1] (Rani & Ganesh, 2014). Setiap dua detiknya seseorang diluar sana membutuhkan darah, tidak ada pengganti untuk darah manusia, setiap harinya darah dibutuhkan oleh rumah sakit dan fasilitas darurat kesehatan lainnya untuk pasien-pasiennya dengan berbagai penyakit yang mereka hadapi. Donor darah dibutuhkan untuk menyelamatkan hidup manusia dari kecelakan[2](Ashoori & Taheeri, 2013). permintaan darah terus meningkat. Semua statistik ini jelas menunjukkan pentingnya permintaan darah dan menimbulkan pertanyaan pembakaran donor darah[3](Rahman dkk, 2011).Pemenuhan kebutuhan akan stok darah yang cukup dapat dilakukan melalui pengolahan dan pengumpulan darah yang baik. Untuk dapat mencukupi stok darah maka diperlukan sukarelawan yang mau mendonorkan darahnya. Untuk itu perlu dilakukan penelitian mengenai perilaku pendonor darah sehingga mereka yang telah mendonorkan darahnya akan kembali mendonorkan lagi darahnya.

Di negara berkembang, kurangnya sumber daya, kurangnya manajemen profesional, mitos dan kesalahpahaman yang timbul dari perbedaan budaya dan sosial membentuk penghalang untuk donor darah[3](Rahman dkk, 2011). masalah utama dalam pengumpulan darah adalah ketidakmampuan untuk mendapatkan darah yang cukup untuk memenuhi kebutuhan pasien atau kesulitan untuk menyeimbangkan antara permintaan dan penawaran dalam pemenuhan kebutuhan darah. Dengan persiapan yang baik, dengan mengelompokkan donor potensial di sedemikian rupa sehingga niat donor untuk menyumbangkan darah di masa depan dapat ditentukan[4](Boonyanusith & Jittamai, 2012)

Penelitian dengan menggunakan dataset donor darah yang memanfaatkan teknik data mining telah dilakukan oleh beberapa peneliti sebelumnya seperti Rani dan Ganesh pada tahun 2014 melakukan komparasi teknik data mining[1], Ashoori dan Taheri tahun 2014 menggunakan teknik clustering untuk menganilisa perilaku pendonor darah[2]. Santhanam dan Sundaram tahun 2010 melakukan klasifikasi

donor darah dengan menggunakan metode CART algoritma[5]. Dalam penelitian ini akan dilakukan pengolahan data donor darah dengan menggunakan teknik data mining yaitu naïve bayes yang ditingkatakan akurasinya dengan menggunakan metode optimasi genetic algorithm

II. KAJIAN LITERATUR

Data mining menjadi metode yang setiap saat menjadi lebih luas pemakaian dalam segala aspek kehidupan manusia, karena data mining dapat dipergunakan dalam memberdayakan perusahaan untuk mengungkap pola yang menguntungkan dan tren dari database yang ada. Perusahaan dan lembaga telah menghabiskan banyak uang untuk mengumpulkan data tetapi tidak mengambil keuntungan dari informasi yang berharga dari kumpulan data yang disimpan. Data mining hadir sebagai ilmu yang mampu menindaklanjuti kebutuhan perusahan atau sebuah lembaga mengenai pemberdayaan kumpulan data yang telah tersimpan selama ini.

Menurut Han dan Kamber[6] (2007) Data mining secara sederhana dapat didefinisikan untuk meng-ektrasi atau menambang data dari kumpulan data. Data mining secara lebih tepat dapat diartikan sebagai penambangan pengetahuan. Sehingga dengan metode data mining kumpulan data yang dimiliki oleh perusahaan atau lembaga dapat dikelola ditambang digali sehingga dapat menghasilkan pengetahuan yang berharga darinya.

Menurut Guronescu [7](2011) data mining mempunyai dua tugas utama yaitu prediksi dan deskripsi. Hal-hal yang dapat dilakukan data mining dalam melakukan tugas prediksinya adalah klasifikasi, regrasi dan melakukan pendekteksian, tujuan utama dari tugas prediksi ini adalah menghasilakan variabel yang dapat memprediksi. Sedangkan hal-hal yang dapat dilakukan data mining dalam tugasnya sebagai alat deksripsi yaitu clustering, penemuan aturan asosisi dan penemuan pola terurut, tujuan dari melakukan tugas deskripsi adalah mengidentifkasi pola sehingga mudah dipahami oleh pengguna.

Naïve bayes untuk menambang data mengenai donor darah. Naive bayes merupakan salah satu algoritma klasifikasi dalam data mining yang menggunakan teori probabilitas untuk menemukan kemungkinan yang paling tepat[8](Bramer, 2007). Klasifikasi Bayes juga dikenal dengan Naïve Bayes, memiliki kemampuan sebanding dengan dengan pohon keputusan dan neural network[6](Han & Kamber,2007). Klasifikasi Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas[9](Kusrini, 2009). Naïve Bayes dapat menggunakan penduga kernel kepadatan, yang meningkatkan kinerja jika asumsi normalitas sangat tidak benar, tetapi juga dapat menangani atribut numeric menggunakan diskritisasi diawasi[10] (Witten & Frank, 2011). Teknik Naïve Bayes (NB) adalah salah satu bentuk sederhana dari Bayesian yang jaringan untuk klasifikasi. Sebuah jaringan Bayes dapat dilihat sebagai diarahkan sebagai tabel dengan distribusi probabilitas gabungan lebih dari satu set diskrit dan variabel stokastik (Pearl 1988)[11] (Liao, 2007).

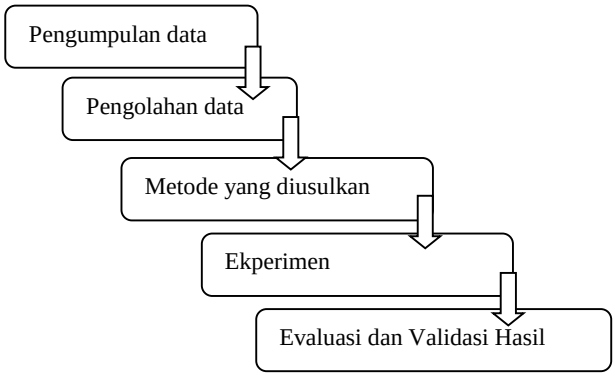
Algoritm Genetika

Algoritma Genetika biasa digunakan untuk klasifikasi dan juga masalah optimisasi. Dalam data mining, metode ini juga digunakan untuk melakukan evaluasi terhadap nilai fitness pada sebuah algoritma. Beberapa hal yang harus dilakukan dalam algoritma genetika[12](Whitcombe , 2006) adalah

1 Menentukan Populasi Awal.
Diperlukan Populasi awal sebelum optimasi dilakukan. Populasi awal dilakukan seperti melakukan pendeklarasian dalam sebuah program
2 Evaluasi Nilai Fitness
Nilai Fitness dievaluasi dengan tujuan agar setiap kromosom memiliki nilai baik atau tidak. Kemudian semua nilai fitness ditentukan probablitasnya masing – masing. Dari hasil probabilitas tertinggi, dihasilkan bahwa kromosom 1 mempunyai nilai fitness paling tinggi. Maka kromosom 1 juga mempunyai kesempatan paling besar dalam proses seleksi selanjutnya dengan *Roullete Wheel*.
3 Seleksi Kromosom Induk
Proses dimana dipilih kromosom yang akan dijadikan kromosom induk dalam populasi yang akan dihitung. Proses seleksi kromosom yang digunakan dengan *Roullete Wheel.*
4 Melakukan Crossover (Perkawinan Silang)
Dalam *crossover* juga melanjutkan ke langkah selanjutnya menggunakan bilangan acak R antara 0 sampai 1. Setelah melakukan pemilihan *parent* (Induk), proses selanjutnya adalah menentukan posisi *crossover*. Setelah didapatkan posisi c*rossover* maka kromosom *parent* (Induk) akan dipotong mulai gen posisi crossover kemudian potongan gen tersebut saling ditukarkan antar *parent* (Induk).
5 Mutasi Kromosom
Jumlah kromosom yang mengalami mutasi dalam satu populasi ditentukan oleh persentase *p mutation*. Proses mutasi dilakukan dengan cara mengganti satu gen yang terpilih secara acak dengan suatu nilai baru yang didapat secara acak. Kromosom tersebut kemudian diuji bila belum sesuai tujuan, maka populasi ini belum memiliki kromosom yang ingin dicapai. Kromosom-kromosom pada populasi ini akan mengalami proses yang sama seperti generasi sebelumnya yaitu proses evaluasi, seleksi, crossover dan mutasi yang kemudian akan menghasilkan Kromosom-Kromosom baru untuk generasi yang selanjutnya. Proses ini akan berulang sampai sejumlah generasi yang telah ditetapkan sebelumnya

III. HASIL PENELITIAN

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari web pada laman uci repository. Berikut tahapan penelitian yang dilakukan:



Gambar 2
Tahapan Penelitian

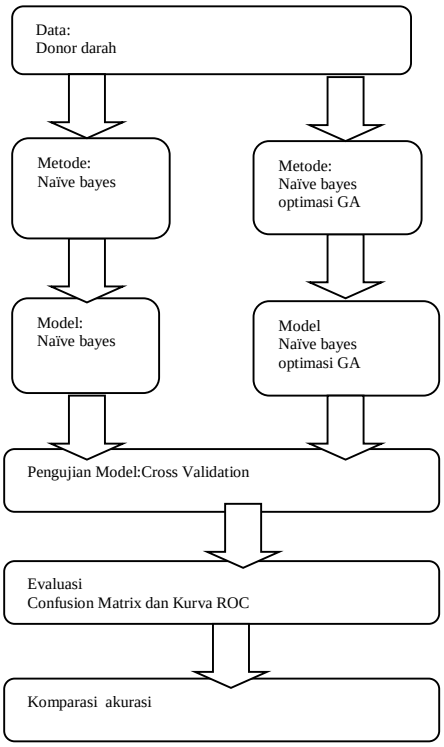
a. pengumpulan data
Data penelitian menggunakan data yang tersedia disebuh situs penyedia dataset diperuntukan untuk keperluan penelitian yaitu pada https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center. Dataset yang digunakan adalah data donor darah terdiri dari enam atribut yaitu frekuensi(bulan) sejak terkahir kali mendonorkan, frekuensi(total jumlah mendonorkan darah), total volume darah yang didonorkan, waktu(bulan) sejak pertama kali mendonorkan darahnya dan satu class label yaitu apakah akan mendonorkan darahnya pada maret 2007. Total data yang terdapat pada dataset tranfusi darah adalah 749 record.

Tabel 1
Dataset donor darah

Recency (Months)	Frequency (times)	Monter y (c.c blood)	Times (months)	Wheher they will donatted in march 2007?
2	50	12500	98	1
0	13	3250	28	1
1	16	4000	35	1
2	20	5000	45	1
1	24	6000	77	0
4	4	1000	4	0
2	7	1750	14	1
1	12	3000	35	0
2	9	2250	22	1
5	46	11500	98	1
4	23	5750	58	0
0	3	750	4	0
2	10	2500	28	1
5	46	11500	98	1

b. Pengolahan data awal
Untuk mendapatkan hasil yang mewakili dari atribut maka diperlukan pengolahan data awal. Ada beberapa teknik yang dilakukan dalam pengolahan data awal dilakukan aitu *data validation, data integration dan transformation, data reduction and dicrization* [13](Vercellis, 2009). Data validation digunakan untuk menghilangkan noise pada data. Noise dapat berupa data tidak lengkap(*missing value*). Data integration dan transformation digunakan untuk menyatukan dan merubah susunan tapi bukan merubah isi dari data. Hal ini dilakukan dengan tujuan menghilangkan atribut yang tidak diperlukan dalam penelitian yang sedang dilakukan. Data reduction and dircrization digunakan untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif. Dataset tranfusi darah yang diambil dari uci repository ubah format fileny dari format .txt(notepad) menjadi format .xls (excel)

c. Metode Yang Diusulkan
Pada penelitian ini digunakan dataset donor darah yang telah dilakukan pengolahan data awal sebelumnya. Dataset donor darah tersebut diolah dengan metode data mining yaitu naïve bayes, tahapan selanjutanya dari pengolahan dataset menggunakan naïve bayes diperoleh model atau hasil nilai akurasinya. Pada model tersebut dilakukan pengujian model dengan metode cross validation, dan dilakukan evaluasi terhadap model tersebut dengan menggunakan confussion matrix dan kurva ROC. Setelah itu diperoleh hasil akurasi dengan menggunakan metode naïve bayes dibandingkan dengan nilai akurasi yang diperoleh dari perhitungan naïve bayes yang telah dioptimasi dengan menggunakan genetic algorithm(GA). Tahapan dalam penggunakaan metode yang diusulkan seperti berikut:



Sumber: peneliti

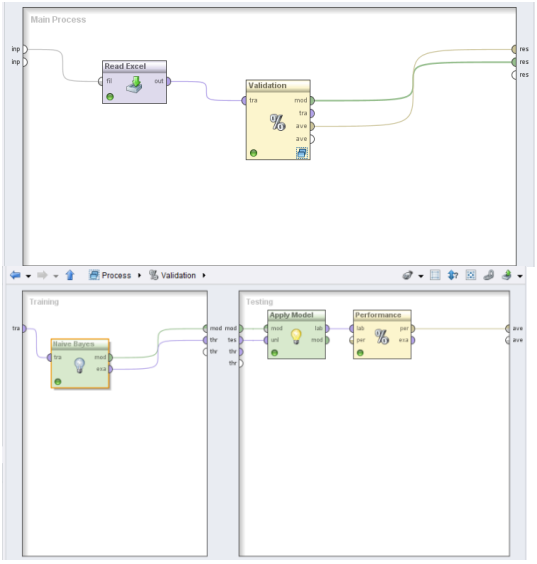
Gambar 3
Model yang diusulkan

1 Ekperimen dan Pengujian Metode

Pada tahapan ini dilakukan ekperimen melalui pengolahan data donor darah. Pengolahan data donor darah dilakukan dengan menggunakan tools aplikasi RapidMiner. RapidMiner merupakan sebuah software yang memiliki kemampuan mengolah dataset dengan berbagai metode data mining yang ada.

Dalam penelitian ini akan dilakukan beberapa percobaan dataset donor darah dengan hanya menggunakan metode data mining naïve bayes dan penggunaan metode optimasi genetic algorithm (GA) untuk meningkatkan akurasi naïve bayes.

Berikut langkah dalam pengolahan dataset donor darah software RapidMiner dengan menggunakan metode naïve bayes, modul yang digunakan adalah modul read excel yang didalamnya terdapat dataset donor darah dalam bentuk excel yang dihubungkan dengan modul validation, didalam modul validatioan terdapat modul naïve bayes yang dihubungakan dengan modul apply model dan modul performance, berikut ilustrasi penggunaan RapidMiner:



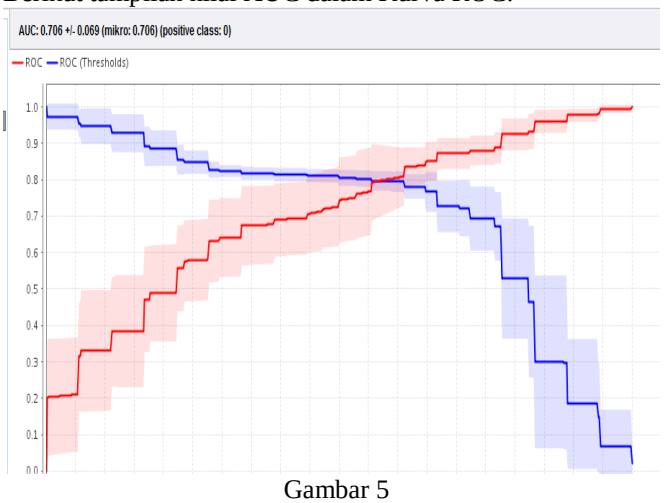
Gambar 4
Tampilan pengolahan data menggunakan naïve bayes

Dari hasil pengolahan dataset donor darah menggunakan naïve bayes diperoleh nilai akurasi yaitu sebesar 74,08% dan nilai AUC sebesar 0,709. Berikut tampilan hasil dari software RapidMiner:

Tabel 1 Nilai Akurasi Naïve Bayes

Accuracy: 74, 87%			
	True 1	True 2	Class precision
Pred. 1	34	44	43.59%
Pred. 2	144	526	78.81%
Class recall	19.10%	92.29%	

Berikut tampilan nilai AUC dalam Kurva ROC:



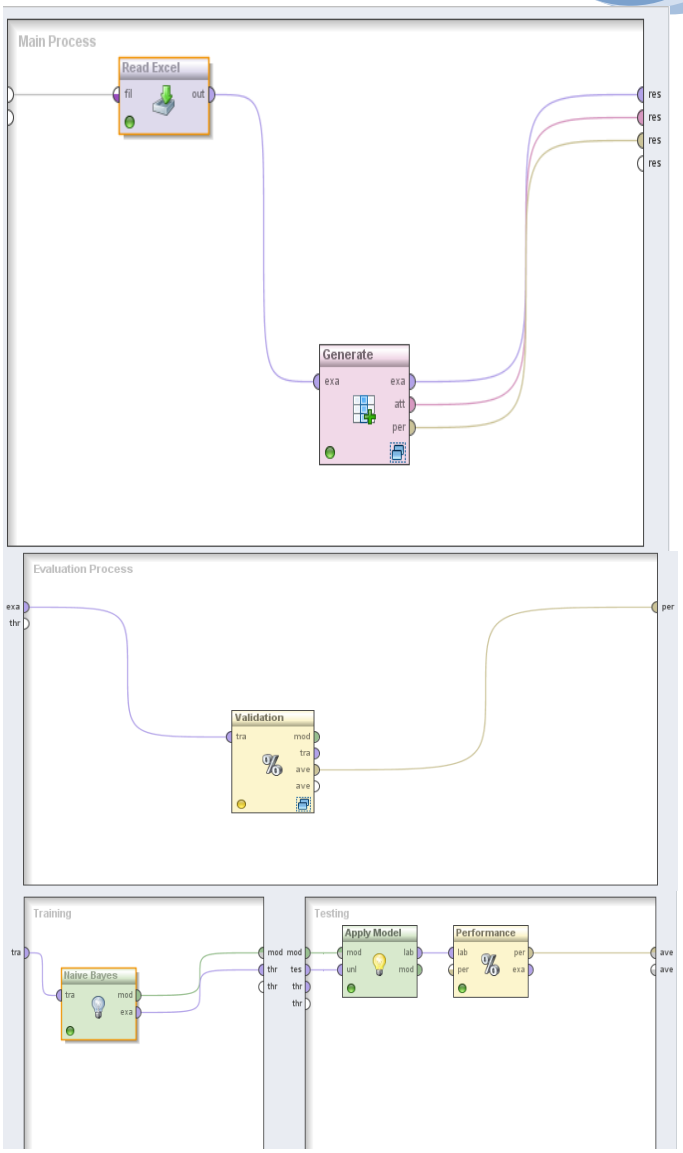
Gambar 5
Kurva ROC pengolahan Naïve Bayes

Menurut Guronescu[6](2011:319) Klasifikasi akurasi merupakan suatu alat pengukuran mengenai seberapa baik suatu klasfikasi melakukan klasifikasi objectnya. Confusion matrix merupakan suatu alat pengukuran untuk melakukan klasifkasi berdasarkan object yang tepat dan object yang tidak tepat.

Kurva ROC atau Reveicer Operating Characterictic Curve, digunakan untuk menilai hasil dari prediksi(peramalan) yang telah dilakuakan. ROC adalah suatu teknik untuk memvisualisasi, organisasi dan klasifikasi terpilih berdasarkan kinerjanya. Secara lengkap hasil perhitungan naïve bayes disajikan dalam performa vektor dibawah ini:

PerformanceVector:
accuracy: 74.87% +/- 2.09% (mikro: 74.87%)
ConfusionMatrix:
True: 1 0
1: 34 44
0: 144 526
precision: 78.51% +/- 1.05% (mikro: 78.51%) (positive class: 0)
ConfusionMatrix:
True: 1 0
1: 34 44
0: 144 526
recall: 92.28% +/- 2.38% (mikro: 92.28%) (positive class: 0)
ConfusionMatrix:
True: 1 0
1: 34 44
0: 144 526
AUC (optimistic): 0.709 +/- 0.068 (mikro: 0.709) (positive class: 0)
AUC: 0.706 +/- 0.069 (mikro: 0.706) (positive class: 0)
AUC (pessimistic): 0.703 +/- 0.070 (mikro: 0.703) (positive class: 0)

Ekperimen selanjutnya adalah dengan melakukan peningkatan nilai optimasi metode naïve bayes pengolahan dataset donor darah dengan menggunakan metode genetic algorthm(GA). Penggunaan Rapidminer dalam tahapan ini hampir sama dengan ekperimen sebelumnnya hanya saja didalam modul validation di letakan modul optimasinya yaitu genetic algorithm (GA). Berikut tampilan pengolahan dataset donor darah dengan optimasi GA:

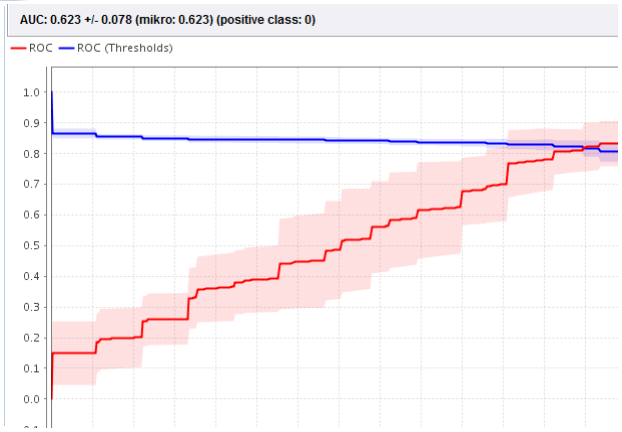


Gambar 5
Pengolahan Data Optimasi Genetic Algorithm Dengan Menggunakan Naïve Bayes

Dari hasil pengolahan diatas diperoleh nilai akurasi dan nilai AUC untuk peningkatan optimasi genetic algorithm pada metode naïve bayes. Diperoleh nilai akurasi sebesar 76,48% dan nilai AUC 0,623. Hasil dari pengolahann rapidminer adalah disajikan dalam gambar dan tabel dibawah ini:

Tabel 2 nilai akurasi Naïve bayes dan Genetic algorithm

Accuracy: 76.48%		
	True 1	True 2
Pred. 1	20	18
Pred. 2	158	552
Class recall	11.24%	96.48%



Gambar 6
Kurva ROC naïve bayes dan genetic algorithm (GA)

Secara lengkap untuk hasil pengolahan rapidminer menggunakan optimasi genetic algorithm dan naïve bayes dapat dilihat dari nilai pada performa vector:

PerformanceVector:
accuracy: 76.48% +/- 2.20% (mikro: 76.47%)
ConfusionMatrix:
True: 1 0
1: 20 18
0: 158 552
precision: 77.77% +/- 1.51% (mikro: 77.75%) (positive class: 0)
ConfusionMatrix:
True: 1 0
1: 20 18
0: 158 552
recall: 96.84% +/- 1.72% (mikro: 96.84%) (positive class: 0)
ConfusionMatrix:
True: 1 0
1: 20 18
0: 158 552
AUC (optimistic): 0.626 +/- 0.077 (mikro: 0.626) (positive class: 0)
AUC: 0.623 +/- 0.078 (mikro: 0.623) (positive class: 0)
AUC (pessimistic): 0.619 +/- 0.079 (mikro: 0.619) (positive class: 0)

Dengan menggunakan optimasi genetic algoritm (GA) diperoleh atribut-atribut yang sangat berpengaruh dan atribut yang tidak berpengaruh terhadap dataset pengolahan data mining naïve bayes. Atribut yang berpengaruh diberi nilai 1 dan atribut yang tidak terpengaruh diberi nilai 0. Berikut tabel AttributeWeight(Generate):

Tabel 3 Nilai atribut

Nama atribut	Weight/bobot
Recency (months)	0
Frequency (times)	0
Montery (cc blood)	1
Times	1

IV. KESIMPULAN

Dari hasil pengolahan dataset donor darah menggunakan meotde peningkatkan optimasi genetic algorithm(GA) terhadapa naïve bayes diperoleh nilai akurasi sebesar 76,48 % yang berarti metode optimasi ini berhasil meningkatkan optimasi naïve bayes yang sebelumnya bernilai 74, 07%. Hasil lain yang diperoleh dalam pennelitian ini yaitu diperoleh bebrapa atribut yang tidak berpengaruh terhadap dataset donor darah hal ini dapat dilihat dari hasil yang diperoleh metode optimasi genetic algorithm yang disajikan dalam tabel AttributWeight(Generate), dalam tabel tersebut disajikan nilai 1 dan 0, nilai 1 diberikan untuk atribut monetary(jumlah darah yang disumbangankan) dan time(waktu dalam bulan dihitung dari terkahir kali mendonorkan darah), hal ini berarti dua atribut ini berpengaruh penting terhadap penelitian ini, dan atribut yang diberi nilai 0 yaitu frekuensi(month) dan frekuensi(time), hal ini berarti kedua atribut ini tidak berpengaruh penting terhadap penelitian ini.

REFERENSI

[1] Rani, Asha S. Ganesh. Hari S, A survey on blood transfusion based on data mining techniques, International Journal of Scientific & Engineering Research. Volume 5. Issue 6. 2014. 1175 ISSN 2229-5518

[2] Ashoori. Maryam, Taheri. Zahra, Using Clustering Methods for Identifying Blood Donors Behavior, 5th Iranian Conference on Electrical and Electronic Engeenering(2013), Islamic Azad University Gonabad Branch

[3] Rahman, M.S., Akter, K.H., Hossain, SH., Basak, A., and Ahmed, S.I. Smart Blood Query: A Novel Mobile Phone Based Privacy-aware Blood Donor Recruitment and Management System for Developing Regions. IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA). 2011. 22-25 March 2011, PP: 544-548.

[4] Boonyanusith. Wijai, Jittamai. Phongchai, Blood Donor Classification Using Neural Network and Decision Tree Techniques, Proceedings of the World Congress on Engineering and Computer Science 2012 Vol I WCECS. 2012. October 24-26. 2012. San Francisco. USA

[5] Sundaram, Syham. T, Santhaman. A COMPARISON OF BLOOD DONOR CLASSIFICATION DATA MINING

MODELS. Journal of Theoretical and Applied Information Technology 31st August 2011. Vol. 30 No.2. 2011, ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195.

[6] Han,J & Kamber. 2007. Data Mining Concepts, Models and Techniques. Second Edition. Morgan Kaufmann Publisher. Elsevier.

[7] Gorunescu. Florin, Data Mining Concepts, Model and Techiques vol 12, ISBN 978-3-642-19720-8, Springer, Berlin, 2011.

[8] Bramer, Max. Principles of Data Mining. Undergraduate Topics in Computer Science ISSN 1863-7310. Springer. London. 2007.

[9] Kusrini, & Luthfi, E. T. 2009. Algoritma Data Mining. Yogyakarta: AndiPublishing,

[10] Witten, H. I., Eibe, F., & Hall, A. M. 2011. Data Mining Machine Learning Tools and Techiques. Burlington: Morgan Kaufmann Publisher..

[11] Liao, Warren. Triataphyllau. Evangelos. 2007. Recent Advanced in Data Mining of Enterprise Data: Algorithm and Application. Science on Computer and Operation Research Vol.6. World Scientific Publishing Co. Pte. Ltd.Singapore.

[12] Whitcombe, J.M., Cropp, R.A., Braddock, R.D., Agranovski, I.E.. 2006. ”The use of sensitivity analysis and genetic algorithms for the management of catalystemissions from oil refi neries” Math. Comput. Model. 4 4, 430 e 438

PROFIL PENULIS

Hilda Amalia adalah dosen pada program studi manajemen informatika pada instuti AMIK BSI Jakarta, penulis lulus dari pendidikan pasca sarjana STMIK Nusa Mandiri pada tahun 2012, dan aktif menulis penelitian dalam bidang data mining. Mengajar dan membimbing mahasiswa pada AMIK BSI Jakarta