

Optimasi Naive Bayes Menggunakan Algoritma Genetika Sebagai Seleksi Fitur Untuk Memprediksi Performa Siswa

by Suhendro Busono

Submission date: 13-Jul-2021 11:02AM (UTC+0700)

Submission ID: 1619001000

File name: sktp-08-07-2021_02_48_40-215524.pdf (278.55K)

Word count: 3620

Character count: 22685

Optimasi Naive Bayes Menggunakan Algoritma Genetika Sebagai Seleksi Fitur Untuk Memprediksi Performa Siswa

Suhendro Busono

Universitas Muhammadiyah Sidoarjo
e-mail: daffasubusita@gmail.com

ABSTRAK. Pada era globalisasi ini, tingkat moralitas pemuda-pemudi mengalami penurunan. Fenomena ini bisa dilihat di media massa atau media elektronik. Media massa atau media elektronik mengabarkan bahwa berita buruk sering terjadi pada pemuda-pemudi. Berita-berita buruk tersebut seperti perkelahian/tawuran, perjudian, pesta miras, durhaka pada orang tua dan sebagainya. Penyebab berita-berita tersebut bukan berasal dari sifat bawaan namun dipicu oleh kebiasaan buruk. Kekurangan perhatian orang tua, rendahnya kualitas hubungan dengan orang tua salah satu pemicu dari kebiasaan buruk. Pendidikan orang tua, pekerjaan orang tua, dukungan keluarga terhadap pendidikan dapat mempengaruhi mainset pemuda-pemudi. Durasi belajar di rumah, durasi bermain di luar rumah, dan durasi pemuda-pemudi mengakses internet dapat mempengaruhi mainset mereka. Performa siswa dengan menggunakan indikator-indikator tersebut dapat diukur dengan pendekatan sains dan teknologi terutama dibidang Data Mining. Pada penelitian sebelumnya, metode yang digunakan adalah Naive Bayes dengan 33 indikator/atribut/fitur dimana menghasilkan tingkat akurasi sebesar 91,15%. Pada penelitian kali ini penelitian mengoptimalkan keberadaan indikator/atribut/fitur dengan menggunakan algoritma genetika. Alhasil dengan optimisasi indikator/atribut/fitur didapatkan peningkatan akurasi sebesar 97,21%.

Kata Kunci : Performa siswa, Data Mining, Naive Bayes, Algoritma Genetika

ABSTRACT. In this globalisation era, the morality teenagers decrease. This phenomena can be seen on mass or electronic media. Mass or electronic media inform that the negatif case often happend on teenagers community. Negatif case such as brawl, drug, gambling, rape, disobedience to parents, and others. The cause of negatif case is not from himself or hisself but it is triggered by bad customs. The less of parent attention, the low of parent relation quality can inflict bad customs from children. Parent education, parent job, the parent support of education can influence children mainset. How long time children study, how long time children have sparetime, how long time children make friend, and how long time children access internet can influence mainset of children. The customs of children explained on sentences before, can be measured by science and technology. Data Mining that is branch of computer science can measure how much quality children or adult perform based on custom framer indicator. In the last research of student performance using Naive Bayes Methode, the number of attribute is too much (33 attribute) and the score of accuracy is 91.15 %. In this research, the researcher optimize attributes of the last research using Genetic Algorithm. Genetic Algorithm can choose relevant attribute. The choice of relevant attributes can increase score of accuracy. The score of accuracy after using Genetic Algorithm is 97.21 %.

Keywords : Student Performance, Data Mining, Naive Bayes, Genetic Algorithm

1. PENDAHULUAN

1.1 Latar Belakang

Ada dua faktor yang mempengaruhi performa siswa, faktor pertama adalah faktor internal dan faktor kedua adalah faktor eksternal. Faktor internal adalah faktor yang cenderung berasal dari bawaan seseorang atau lebih dikenal faktor genetik, sedangkan faktor eksternal adalah faktor yang berasal dari luar orang tersebut. Faktor eksternal ini biasa dimulai dari dimana anak mulai belajar untuk menyesuaikan terhadap lingkungan sosial [1]. Berdasarkan faktor eksternal tersebut, peneliti menguji baik tidaknya performa siswa melalui pendekatan sains dan teknologi. Adapun pendekatan sains dan teknologi tersebut digunakan salah satu ilmu cabang pada ilmu komputer yaitu data mining.

Pada penelitian performa siswa sebelumnya telah dilakukan penelitian dengan menggunakan metode Naive Bayes oleh Departemen Sistem Informasi Universitas Minho Portugal [2]. Alasan Penelitian menggunakan metode Naive Bayes karena Metode Naive Bayes dapat memberikan kemudahan dalam menghitung dan dapat menentukan kemungkinan-kemungkinan yang akan terjadi. Alasan lainnya adalah metode Naive Bayes mampu menghitung data dalam jumlah yang cukup besar. Pada Penelitian performa siswa kali ini, peneliti menambahkan algoritma genetika untuk meningkatkan nilai akurasi dengan memilih atribut-attribut yang dominan, mengingat pada penelitian performa siswa sebelumnya memiliki banyak atribut.

Naive Bayes adalah salah satu teknik klasifikasi pada data mining dengan menggunakan pendekatan probabilitas dan statistik. Teori Naive Bayes ditemukan oleh Thomas Bayes dimana pada teori ini bisa diprediksi kemungkinan besar yang akan terjadi dengan atribut-attribut yang telah ada dan Naive Bayes mampu menangani data dengan jumlah yang besar [3].

Algoritma Genetik adalah algoritma yang digunakan untuk mendapatkan nilai yang optimum. Prinsip dari algoritma genetik adalah mendapatkan individu yang terbaik dimana untuk mendapatkan individu yang terbaik harus mengalami proses seleksi dengan membangkitkan populasi awal kemudian proses pindah silang (*Cross Over*) dan terakhir adalah proses mutasi [4].

1.2 Rumusan Masalah

1. Pada Penelitian terdahulu jumlah atribut yang digunakan terlalu besar yaitu sebanyak 33 atribut sehingga diperlukan adanya pemilihan atribut yang dominan berdasarkan Algoritma Genetika.
2. Pada penelitian terdahulu jumlah atribut yang digunakan terlalu besar sehingga berdampak pada nilai akurasi / prosentase kebenaran yang dihasilkan. Untuk mendapatkan nilai akurasi / prosentase kebenaran yang lebih tinggi, maka jumlah atribut yang digunakan harus dioptimasi dengan menggunakan Algoritma Genetika.

1.3 Tujuan Penelitian

1. Memilih atribut yang dominan berdasarkan Algoritma Genetika dalam memprediksi performa siswa
2. Menguji nilai akurasi / prosentase kebenaran metode Naive Bayes sebelum dan setelah penambahan Algoritma Genetika.

1.4 Manfaat Penelitian

1. Pendidik diharapkan mengetahui indikator-indikator apa saja yang mempengaruhi performa siswa.
2. Dengan adanya penilaian performa siswa, pendidik dapat memprediksi sampai sejauh mana performa siswa berdasarkan indikator-indikator yang ada.
3. Dengan adanya penelitian ini, pemerintah khususnya kementerian sosial diharapkan dapat mengetahui dan merespon upaya untuk memperbaiki performa siswa kearah yang lebih baik
4. Para orang tua diharapkan dapat mengetahui dan merespon indikator-indikator yang mempengaruhi performa siswa.

2. METODE PENELITIAN

2.1 Penelitian Terkait

Pada penelitian pertama, metode yang digunakan adalah metode Naive Bayes dengan meneliti student performance. Tingkat akurasi yang didapatkan pada penelitian yang pertama adalah 60 %. Software Tool yang digunakan pada penelitian ini adalah WEKA[5]. Pada penelitian kedua peneliti menggunakan Algoritma Naive Bayes dengan menggunakan software R.Miner. Tingkat akurasi yang didapatkan pada penelitian tersebut sebesar 91.9 % untuk program matematika, dan 89.7 % untuk program bahasa[2]. Pada penelitian yang ketiga, penelitian dilakukan untuk mereduksi atribut-attribut yang ada pada student performance. Untuk mereduksi atribut-attribut yang ada, pada penelitian ini menggunakan fitur selection. Terdapat enam teori fitur selection di penelitian ini yaitu: Correlation-based-Attribute evaluation, Chi-Square Attribute evaluation, Gain-Ratio Attribute evaluation, Information-Gain Attribute evaluation, Relief Attribute evaluation, Symmetrical Uncertainty Attribute evaluation. Hasil akurasi tertinggi yang didapatkan pada penelitian ini adalah Correlation-based Attribute evaluation sebesar 72.9% dan Information-Gain Attribute evaluation sebesar 72.9 % [6]. Pada penelitian yang keempat, peneliti mengoptimasi data sentimen review jasa maskapai penerbangan menggunakan algoritma genetika dengan nilai akurasi sebesar 89.50% yang sebelumnya hanya sebesar 60 %.[7]. Pada penelitian yang kelima, peneliti melakukan penelitian prediksi hasil pemilu legeslatif DKI Jakarta menggunakan naive bayes dengan algoritma genetika sebagai fitur seleksi. Hasil yang didapatkan pada penelitian tersebut memiliki nilai akurasi sebesar 97.84 % yang sebelumnya hanya sebesar 92.28 % tanpa adanya penambahan algoritma genetika[8].

2.2 Instrument Penelitian

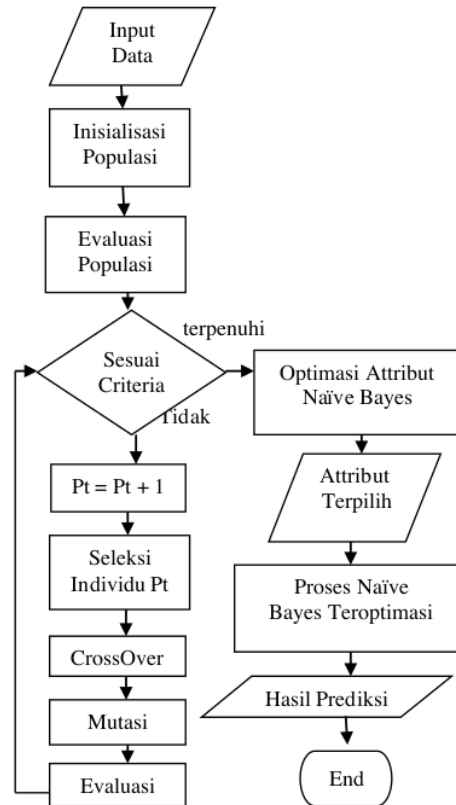
Pada penelitian ini, data yang digunakan dari data UCI (*UC Irvine Machine Learning Repository*). Data penelitian ini bisa di download di archive.ics.uci.edu/ml/datasets.html di kolom *student performance* tahun 2014.

Penelitian ini menggunakan data student-math.csv, dimana data tersebut adalah data untuk siswa yang penilaiannya berdasarkan program matematika. Data penelitian ini terdapat 33 indikator / 33 atribut dan 1 label / target penilaian. 33 atribut dan 1 label

2.3 Metode Yang Diusulkan

Pada penelitian yang dilakukan oleh Departmen Sistem Informasi Universitas Minho Portugal, peneliti menggunakan pendekatan data mining dengan menggunakan metode klasifikasi yaitu Naive Bayes. Pada penelitian tersebut peneliti menggunakan 30 atribut, namun pada penelitian kali ini, peneliti ingin mereduksi atribut-attribut yang kurang relevan. Pada penelitian kali ini, peneliti mengusulkan metode algoritma

genetika untuk mereduksi attribute-attribut yang kurang dominan karena metode algoritma genetika memiliki kemampuan dalam menseleksi atribut yang kurang dominan. Gambar 1 menunjukkan diagram alir Algoritma Naive Bayes setelah penambahan Algoritma Genetika.



Gambar 1. Diagram Alir Algoritma Naïve Bayes + Algoritma Genetika

2.4 Tahapan Penelitian

Prosedur diagram alir pada optimasi Naive Bayes menggunakan algoritma Genetika adalah sebagai berikut:

1. Pembentukan Individu
2. Inisialisasi Populasi
Inisialisasi dilakukan dengan cara memberikan nilai awal pada masing-masing individu sesuai dengan jumlah populasi yang diinginkan.
3. Evaluasi Populasi
Evaluasi dilakukan untuk mendapat nilai individu yang terbaik. Kriteria untuk mendapatkan nilai individu terbaik berdasarkan nilai yang telah ditentukan. Nilai fitness adalah nilai untuk mendapatkan nilai terbaik pada suatu populasi, namun jika tidak maka dilakukan proses seleksi individu.
4. Seleksi Individu
Proses seleksi dilakukan dengan cara membuat individu yang mempunyai fungsi objektif kecil mempunyai kemungkinan untuk dipilih. Untuk mendapatkan individu terpilih tersebut digunakan fungsi fitness. Rumus dari fungsi fitness adalah $(1/(1+\text{fungsi objektif}))$.
5. *CrossOver* (Pindah Silang)
Setelah proses seleksi dikerjakan, maka proses selanjutnya adalah *CrossOver*. Metode yang digunakan untuk penelitian ini adalah menggunakan *one cut point*. Metode *one cut point* yaitu memilih acak satu

posisi dalam kromosom induk yang kemudian saling menukar gen. Kromosom yang dijadikan induk dipilih secara acak dan jumlah kromosom yang mengalami *crossover* dipengaruhi oleh parameter *crossover rate*.

6. Mutasi

Jumlah kromosom yang mengalami mutasi dalam satu populasi ditentukan oleh parameter *mutation rate*. Proses mutasi dilakukan dengan cara mengganti satu gen yang terpilih secara acak dengan suatu nilai baru yang didapatkan secara acak.

2.5 Pengujian dan Evaluasi

Untuk mengetahui tingkat akurasi naïve bayes sebelum penambahan algoritma genetika dan setelah penambahan algoritma genetika, maka digunakan *Confusion Matrix*. Pengujian pada penelitian ini menggunakan 10 fold validation.

Metode *Confusion Matrix* adalah metode yang menggunakan tabel matriks, dimana terdapat 5 persamaan model *confusion matrix* yaitu nilai akurasi, *sensitivity* dan *specificity*, *positif predictive value*, *negative predictive value*.

3. ANALISA HASIL PENELITIAN DAN PEMBAHASAN

Adapun perhitungan Metode Naive Bayes Untuk Data *Student Performance* adalah sebagai berikut :

Tahap ke_1 :

Pengambilan Data Student Performance

Sebelum melangkah ke perhitungan Naive Bayes, atribut-atribut harus diketahui terlebih dahulu. Tabel 1 adalah diskripsi untuk masing-masing atribut.

Tabel 1. Atribut Student Performance

No	Nama Atribut	Instances / Nilai	Keterangan Atribut
1	Sekolah	GP/MS	GP : Gabriel Pereira MS : Mousinho da Silveira
2	Jenis Kelamin	Pria /Wanita	
3	Umur	15 s/d 22	15 s/d 22 tahun
...
33	Nilai kelas akhir	0 s/d 20	

Setelah diketahui masing-masing atribut, maka selanjutnya data *student performance* digunakan untuk perhitungan Naive Bayes. Tabel 2 adalah data *student performance* dengan jumlah data sebesar 395.

Tabel 2. Dataset Student Performance

No	Sekolah	Jenis Kelamin	Umur	Alamat	Jumlah Keluarga	Nilai Kelas akhir	Label
1	GP	F	18	U	GT3	6	fail
2	GP	F	17	U	GT3	6	fail
....
395	MS	M	19	U	LE3	9	fail

Tahap ke_2 :

Hitung probabilitas data yang lulus dan gagal untuk masing-masing atribut pada data *student performance*, dimana rumus probabilitas kelulusan dan kegagalan pada masing-masing atribut

Tahap ke_3 :

Kalikan semua probabilitas yang lulus dan yang gagal pada masing-masing atribut

Rumus probabilitas lulus untuk semua atribut :

$$\text{Prob}(\text{pass}) = \text{Prob}(\text{pass}_{n_{x1}}) \times \text{Prob}(\text{pass}_{n_{x+1}}) \times \text{Prob}(\text{pass}_{n_{x+n}})$$

Tahap Ke_4

Bandingkan probabilitas yang lulus dengan yang gagal, jika nilai probabilitas lulus lebih besar dari probabilitas gagal, maka hasil adalah lulus. Sebaliknya jika nilai probabilitas gagal lebih besar dari

probabilitas lulus, maka hasil adalah gagal. Jika melihat pada tabel 2, nilai Prob(lulus) lebih besar daripada nilai Prob(gagal), maka hasil prediksi adalah **LULUS**

3.1 Pengujian Metode Naive Bayes Untuk Data *Student Performance*

Pada penelitian ini menggunakan 395 data dengan mengambil program matematika. Penelitian ini menggunakan 10 fold cross validation untuk menguji tingkat akurasi. 10 fold validation memiliki maksud bahwa 395 data dipecah menjadi 10 kelompok dimana kelompok pertama bertindak sebagai data testing dan 9 kelompok yang lainnya sebagai data training. Untuk proses selanjutnya kelompok kedua bertindak sebagai data testing dan 9 kelompok yang sebagai data training, begitu selanjutnya hingga kelompok kesepuluh sebagai data testing dan 9 kelompok yang lain sebagai data training. Tabel 3 ini adalah probabilitas untuk masing-masing atribut berdasarkan hasil pengujian dengan menggunakan 10 fold cross validation

Tabel 3. Probabilitas Masing-Masing Atribut

Atribut	Nilai Atribut	Probabilitas Gagal	Probabilitas Lulus
Sekolah	GP	0.8692	0.8906
School	MS	0.1308	0.1094
.....
Kondisi Kesehatan	4	0.1846	0.1585

3.2 Perhitungan Algoritma Genetika Untuk Data *Student Performance*

Adapun perhitungan Algoritma Genetika untuk data *Student Performance* sebagai berikut :

1. Inisialisasi Gen
Setiap atribut diberi inisial gen yang berbeda-beda tergantung dari tipe data
2. Inisialisasi Kromosom
Inisialisasi kromosom berdasarkan pada inisialisasi gen pada tiap-tiap atribut.
3. Pengambilan Populasi
Setelah inisialisasi kromosom terbentuk, maka proses selanjutnya adalah pengambilan populasi
4. Menentukan nilai fitness pada masing-masing kromosom
Rumus Nilai Fitness = Nilai Kromosom Ke_x / Total Jumlah bit
5. Menentukan Probabilitas masing-masing kromosom
Nilai Probabilitas digunakan untuk mendapatkan nilai tertinggi
Rumus Total Fitness :
$$\text{Total Fitness} = \sum_{x=1}^n \text{Nilai Fitness Kromosom } Cx$$
6. Seleksi Individu
Tipe Seleksi pada penelitian ini menggunakan Roulette Wheel
7. Cross Over / Kawin Silang
Untuk melakukan proses kawin silang, maka diperlukan probabilitas CrossOver.
8. Mutasi
Untuk melakukan proses mutasi, maka diperlukan probabilitas mutasi.
9. Mencari Nilai Fitness Terbaik

Pada perhitungan ini menggunakan satu iterasi / satu generasi, sehingga untuk mendapatkan kromosom terbaik berasal dari proses mutasi yang pertama.

Tabel 4. Atribut Terpilih Setelah Proses GA

No	Nama Atribut	Hasil Seleksi Atribut
1	Sekolah	Ada Atribut
2	Jenis Kelamin	Ada Atribut
3	Umur	Atribut Hilang
4	Alamat	Ada Atribut
5	Jumlah Keluarga	Atribut Hilang
...
33	Nilai kelas kedua	Ada Atribut

Dari perhitungan GA diatas didapatkan adanya pengurangan fitur / atribut yang jumlah awalnya sebanyak 30 menjadi 28. Pengurangan atribut tersebut adalah *age*(Umur) dan *famsize*(Jumlah keluarga).

3.3 Pengujian Metode Naive Bayes Dengan Algoritma Genetika

Pada penelitian ini terdapat 6 parameter yang mempengaruhi proses pengujian Algoritma Genetika :

1. *Population Size* (Ukuran Populasi)

Proses pertama yang dilakukan oleh Algoritma Genetika adalah menginisialisasi individu. Pada penelitian ini, proses inisialisasi individu menggunakan bilangan biner. Setelah pembentukan individu terbentuk, langkah selanjutnya adalah menentukan jumlah individu dalam satu populasi. Pada penelitian ini ukuran populasi dimulai dari 20 individu sampai dengan 100 individu dengan kelipatan 20 (20,40,60,80,100).

2. *Number of Generation*

Number of Generation adalah menentukan berapa banyak jumlah iterasi pada proses Algoritma Genetika. Satu iterasi atau satu generasi diawali dari menentukan ukuran individu dalam satu populasi sampai dengan proses mutasi. Pada penelitian ini *number of generation* atau jumlah iterasi mulai dari 10 sampai dengan 100 iterasi dengan kelipatan 20 (10,20,30,40,50,60,70,80,90,100).

3. *Selection Scheme* / Skema Seleksi Individu

Selection Scheme adalah proses algoritma genetika pada tahap seleksi individu. Pada tahapan proses seleksi terdapat beberapa macam teknik yaitu *uniform*, *roulette wheel*, *stochastic universal sampling*, *boltzman*, *tournament*. Pada penelitian ini skema seleksi yang digunakan adalah *roulette wheel*. *Roulette Wheel* adalah menjumlahkan secara akumulatif nilai probabilitas sebelumnya dengan nilai probabilitas sekarang.

4. *Cross Over Type*

Pada penelitian ini, tipe *cross over* yang diuji menggunakan *one point*. One point adalah **1** yaitu memilih acak satu posisi dalam kromosom induk yang kemudian saling menukar gen.

5. Probabilitas *Cross Over*

Probabilitas *cross over* digunakan untuk menentukan apakah individu tersebut berhak dikawin silangkan atau tidak. Jika kromosom yang nilai R nya lebih kecil daripada nilai probabilitas crossover, maka kromosom tersebut berhak untuk dikawin silangkan dengan kromosom yang senilai dengannya dan sebaliknya. Pada penelitian ini probabilitas *cross over* sebesar 50 %, yang artinya bahwa jika kromosom yang nilai R nya lebih kecil dari pada 0.5, maka kromosom tersebut berhak untuk dikawin silangkan dengan kromosom yang senilai dengannya.

6. Probabilitas Mutasi

Jumlah kromosom yang mengalami mutasi dalam satu populasi ditentukan oleh probabilitas mutasi. Proses mutasi dilakukan dengan mengganti satu gen yang terpilih secara acak dengan suatu nilai baru didapatkan secara acak juga. Pada penelitian ini probabilitas mutasi yang digunakan adalah 10 %.

Pada penelitian ini, 33 atribut yang telah diuji dengan metode naive bayes dioptimasi menggunakan algoritma genetika. Sebelum proses algoritma dijalankan maka terlebih dahulu ditentukan 6 parameter masukan yaitu *population size*, *number of generation*, *selection scheme*, *crossover type*, probabilitas *crossover* dan probabilitas mutasi. Pada *population size* ditentukan range dari 20 hingga 100, pada *number of generation* ditentukan range dari 10 hingga 100. *Selection scheme* yang digunakan adalah *rouelette wheel*, tipe *cross over* adalah *one point*. Probabilitas *crossover* yang digunakan adalah 50%, dan probabilitas mutasi sebesar 10 %.

3.4 Hasil dan Evaluasi Metode Naive Bayes

Pada penelitian ini hasil pengujian yang didapatkan pada metode naive bayes sebelum penambahan algoritma genetika dapat dilihat pada tabel 5 confusion matriks dibawah ini.

Tabel 5. Confusion Matriks Hasil Proses Naive Bayes

	True Fail	True Pass
Prediction Fail	113	18
Prediction Pass	17	247

Pada tabel confusion matriks diatas dapat diketahui 5 persamaan persamaan model confusion matriks yaitu nilai akurasi, sensitivity dan specificity, positif predictive value, negative predictive value.

1. Nilai Akurasi

$$akurasi = \frac{tp + tn}{tp + tn + fp + fn}$$

$$akurasi = \frac{247 + 18}{247 + 18 + 17 + 113} = 67 \%$$

2. Nilai Sensitivity

$$Sensitivity = \frac{tp}{tp + fn}$$

$$Sensitivity = \frac{247}{247 + 113} = 69 \%$$

3. Nilai Specificity

$$Specificity = \frac{tn}{tn + fp}$$

$$Specificity = \frac{18}{18 + 17} = 51 \%$$

4. Positif Predictive Value

$$PPV = \frac{tp}{tp + fp}$$

$$PPV = \frac{247}{247 + 17} = 94 \%$$

5. Negative Predictive Value

$$NPV = \frac{tn}{tn + fn}$$

$$NPV = \frac{18}{18 + 113} = 14 \%$$

3.5 Hasil dan Evaluasi Naive Bayes Dengan Algoritma Genetika

Pada penelitian ini, probabilitas mutasi ditentukan sebesar 10 % dan probabilitas crossover sebesar 50 % untuk semua pengujian. Tipe crossover menggunakan one point dan skema seleksi menggunakan roulette wheel untuk semua pengujian. Pada penelitian ini nilai yang dirubah secara berkala adalah pada population size (jumlah individu pada satu populasi) dan number of generation (jumlah iterasi). Pada penelitian ini dilakukan 50 kali pengujian dengan mengubah nilai population size / jumlah individu pada satu populasi dan nilai number of generation / banyaknya iterasi pada proses GA. Tabel 6 adalah tabel pengujian pada data student performance setelah penambahan algoritma genetika dengan jumlah populasi 20 dan jumlah iterasi 10. Tabel 6 adalah tabel pengujian pada data student performance setelah penambahan algoritma genetika dengan jumlah populasi 40 dan jumlah iterasi 50.

Tabel 6. Pengujian GA Jumlah Populasi = 40 dan Jumlah Iterasi = 50

Jumlah Iterasi	Jumlah Populasi	Prob Cross Over	Prob Mutasi	Jumlah Atribut	Atribut	Akurasi
50	40	0.5	-1.0	10	Sekolah,alamat rumah,alasan memilih sekolah,pembayaran ekstra,sekolah perawat yang dihadiri, keinginan pendidikan yang lebih tinggi,konsumsi alkohol per minggu,kondisi kesehatan, skor periode kedua,skor periode terakhir	95.45 % +/- 3.32 %

Pada beberapa pengujian metode naive bayes ditambahkan dengan GA, nilai akurasi tertinggi dihasilkan pada jumlah populasi sebesar 60 dengan jumlah iterasi sebesar 20 hingga 100 yaitu 97.21 % dengan range toleransi sebesar +/- 2.87 %. Jumlah atribut yang dihasilkan untuk nilai akurasi tertinggi sebanyak 11 atribut. 11 atribut tersebut adalah Asal sekolah, alamat rumah, jumlah keluarga, pekerjaan ibu, alasan memilih sekolah, hubungan yang baik dengan keluarga, intensitas pergi dengan teman, konsumsi alkohol per hari, kondisi kesehatan, skor periode kedua, skor periode terakhir.

3.6 Perbandingan Metode Naive Bayes Dengan Metode Naive Bayes + GA

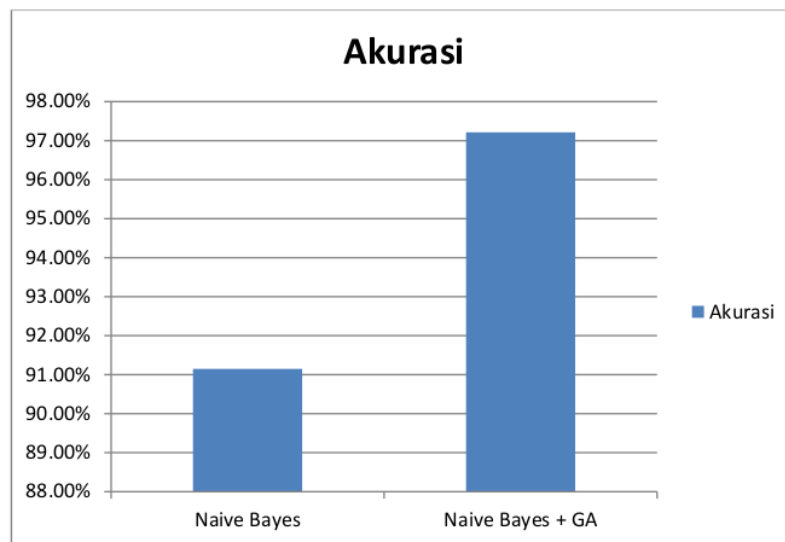
Berdasarkan hasil pengujian metode Naive Bayes sebelum dioptimasi dengan metode Naive Bayes setelah dioptimasi dengan Algoritma Genetika, maka dihasilkan akurasi pada tabel 7.

Tabel 7. Akurasi Naive Bayes dan Naive Bayes + GA

Metode	Akurasi
Naive Bayes	91,15 % +/- 5,26 %
Naive Bayes + GA	97,21 % +/- 2,87 %

Akurasi metode Naive Bayes sebelum dioptimasi sebesar 91,15 % dengan range toleransi sebesar +/- 5,26 % dan Naive Bayes setelah dioptimasi dengan algoritma genetika menghasilkan nilai akurasi sebesar 97,21 % dengan range toleransi sebesar +/- 2,87 %.. Adanya peningkatan nilai akurasi setelah dioptimasi dengan Algoritma Genetika dapat membuktikan bahwa Algoritma Genetika mampu meningkatkan nilai akurasi dengan memilih atribut-attribut yang dominan.

Bila perbandingan metode Naive Bayes dan Naive Bayes + GA direpresentasikan dalam bentuk grafik, maka gambar 2 adalah representasi dari perbandingan kedua metode tersebut



Gambar 2. Grafik Perbandingan Metode NB dan NB + GA

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

1. Pada penelitian ini Algoritma Genetika difungsikan untuk mengoptimasi atribut-attribut yang ada dengan memilih atribut-attribut yang dominan.
2. Pada penelitian ini nilai akurasi / prosentase kebenaran tertinggi pada Algoritma Genetika dihasilkan pada jumlah populasi sebesar 60 kromosom / individu dengan jumlah iterasi / perulangan program sebesar 20 hingga 100 iterasi. Nilai akurasi / prosentase kebenaran tertinggi pada Algoritma Genetika menunjukkan kinerja sebesar 97.21 % dengan range toleransi sebesar +/- 2.87 % dan jumlah atribut dihasilkan sebanyak 11 atribut dalam memprediksi performa siswa.

4.2 Saran

1. Adanya beberapa metode klasifikasi pada data mining, mendorong agar penelitian ini dapat menggunakan metode klasifikasi yang lain dengan teknik optimasi yang sama yaitu algoritma genetika.
2. Adanya teknik-teknik optimasi selain daripada algoritma genetika, mendorong penelitian ini dikembangkan pada teknik optimasi yang lainnya.

Penelitian performa siswa ini sangat cocok jika diterapkan secara real oleh pemerintah Indonesia dengan penambahan dan perubahan pada atribut atau indikator-indikator yang mempengaruhi performa siswa.

DAFTAR PUSTAKA

- [1] Yusuf LN. S, *Psikologi Perkembangan Anak dan Remaja*, Bandung: PT Remaja Rosdakarya, 2004.
- [2] Cortez Paulo, *Using Data Mining to Predict Secondary School Student Performance*, 2014.
- [3] Han & Kamber, *Data Mining Concept and Techniques*.
- [4] Randy L. Haupt, *Practical Genetic Algorithms*, 2004.
- [5] Sofia, *Predicting Student Performance by Using Data Mining Methods for Classification*, 2013.
- [6] Bhaskaran & Ramaswarni, *A Study on Feature Selection Techniques in Educational Data Mining*, 2009.
- [7] Wati, Rista, Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisa Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes, 2016
- [8] Tri, Diana, Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naive Bayes Dengan Algoritma Genetika Sebagai Fitur Seleksi, 2009
- [9] H. Almuallim and T. G. Dietterich. *Learning boolean concepts in the presence of many irrelevant features*, Artificial Intelligence, vol. 69, no. 1-2, pp. 279-305, 1994.

- [10] D. Koller and M. Sahami, *Toward optimal feature selection*, In Proceedings of the Thirteenth International Conference on Machine Learning, pp. 284–292, 1996.

Optimasi Naive Bayes Menggunakan Algoritma Genetika Sebagai Seleksi Fitur Untuk Memprediksi Performa Siswa

ORIGINALITY REPORT

7%

SIMILARITY INDEX

9%

INTERNET SOURCES

3%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1

id.scribd.com

Internet Source

3%

2

core.ac.uk

Internet Source

2%

3

www.arpnjournals.org

Internet Source

2%

Exclude quotes On

Exclude bibliography On

Exclude matches < 2%