

Optimasi Algoritma Naïve Bayes Menggunakan Algoritma Genetika Untuk Memprediksi Kelulusan (Studi Kasus: Mahasiswa Jurusan Matematika FMIPA Universitas Mulawarman)

Optimization of Naïve Bayes Algorithm Using Genetic Algorithm to Predict Graduation (Case Study: Mathematics Department Student, Faculty of Mathematics and Natural Sciences, Mulwarman University)

Elisa Feronica¹, Yuki Novia Nasution², Ika Purnamasari³

^{1,2}Laboratorium Matematika Komputasi FMIPA Universitas Mulawarman

³Laboratorium Statistika Ekonomi dan Bisnis FMIPA Universitas Mulawarman

E-mail: ¹elisafeonica2000@gmail.com

ABSTRACT

The Naïve Bayes algorithm is classification method that uses the principle of probability to create predictive models. Naïve Bayes is based on the assumption that all its attributes are independent which can be optimized by genetic algorithms. Genetic algorithm is an optimization technique which works by imitating the process of evaluating and changing the genetic structure of living creatures. In this study, the Naive Bayes algorithm was optimized using by genetic algorithm to predict student graduation with attributes, namely gender, regional origin, admission path and employment status. The data used is the students of the Mathematics Department, Faculty of Mathematics and Natural Sciences, Mulawarman University who graduated in March 2018 to December 2020. The results of this study indicate the accuracy value generated by Naïve Bayes of 50% increased by 16,67% after the attributes were optimized by using the genetic algorithm to 66,67% with 3 selected attributes, namely regional origin, admission path and employment status.

Keywords: Genetic algorithm, Graduation, Naive Bayes algorithm, Optimization

Pendahuluan

Kelulusan adalah keberhasilan menyelesaikan program studi di Universitas dengan rentang waktu yang dipersyaratkan. Rentang waktu dalam perkuliahan sering disebut dengan Masa Studi. Berdasarkan peraturan akademik Universitas Mulawarman, masa studi mahasiswa untuk program sarjana (S1) kategori dengan pujian, mahasiswa dapat menyelesaikan masa studi tidak melampaui $n + 1$ semester dimana $n = 8$ semester untuk mahasiswa S1, sehingga dapat dikatakan bahwa masa studi mahasiswa lulus tepat waktu tidak lebih dari 9 semester atau 4 tahun 6 bulan. Jika lebih dari itu maka dapat dikatakan bahwa mahasiswa lulus tidak tepat waktu.

Pada prakteknya, tidak banyak mahasiswa yang dapat menyelesaikan studi tepat pada waktunya. Hal ini membuat ketidakseimbangan antara jumlah mahasiswa masuk dan keluar, sehingga terjadi penumpukan mahasiswa yang dapat mengakibatkan proses belajar tidak berjalan dengan maksimal. Oleh karena itu, perlu dilakukan prediksi kelulusan mahasiswa sejak dini untuk mencegah terjadinya penumpukan mahasiswa. Bila nanti mahasiswa tersebut teridentifikasi lulus tidak tepat pada waktunya maka mahasiswa tersebut dapat diberikan bimbingan agar dapat meningkatkan prestasi belajar. Terdapat dua faktor yang mempengaruhi mahasiswa lulus tidak tepat waktu, yaitu faktor

internal dan faktor eksternal. Faktor internal adalah faktor dari dalam diri mahasiswa yang mempengaruhi prestasi belajar, sedangkan faktor eksternal adalah faktor dari luar yang mempengaruhi prestasi belajar mahasiswa (Lembang & Fendjalang, 2015). Dari faktor eksternal tersebut maka akan dilakukan prediksi dengan menggunakan *data mining*.

Data mining merupakan proses yang sangat efektif dalam menyelesaikan permasalahan yang berkaitan dengan prediksi. Salah satu metode yang termasuk dalam kategori ini ialah klasifikasi. Klasifikasi adalah proses untuk mendapatkan sejumlah model yang menggambarkan, mengenali serta membedakan kelas data atau konsep yang bertujuan untuk menggunakan model yang telah diperoleh agar dapat memprediksi kelas objek yang label kelasnya belum diketahui (Han, Kamber & Pei, 2012).

Algoritma Naïve Bayes adalah metode klasifikasi yang menggunakan prinsip probabilitas untuk membuat model prediksi. Metode ini menghitung kemungkinan suatu kejadian dan dapat berubah jika terdapat informasi pendukung tambahan yang disediakan (Kurniawan, 2020). Naive Bayes didasarkan pada Teorema Bayes dengan asumsi bahwa semua atributnya bersifat independen (tidak saling berhubungan), yang artinya semua atribut sama pentingnya (Arhami & Nasir, 2020). Akan tetapi, asumsi tersebut tidak selalu benar sehingga dapat mempengaruhi

tingkat akurasi. Sifat independen tersebut dapat dioptimasi dengan algoritma genetika. Algoritma genetika adalah suatu teknik pencarian (*searching technique*) dan teknik optimasi yang cara kerjanya meniru proses evaluasi dan perubahan struktur genetika pada makhluk hidup (Arkeman, Seminar & Gunawan, 2012).

Penelitian mengenai algoritma Naïve Bayes dengan algoritma genetika telah dilakukan oleh beberapa peneliti, antara lain Fithri & Darmanto (2014) telah melakukan penelitian tentang Sistem Pendukung Keputusan untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes didapatkan tingkat akurasi sebesar 93%. Sementara itu, Suryani & Priyanti (2019) Optimasi Naïve Bayes dan Algoritma Genetika untuk Prediksi Beasiswa Pendidikan pada SMP Utama dengan tingkat akurasi sebesar 83,33%. Kemudian, Busono (2020) Optimasi Naïve Bayes Menggunakan Algoritma Genetika sebagai Seleksi Fitur untuk Memprediksi Performa Siswa didapatkan peningkatan akurasi sebesar 97,21%.

Tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui hasil optimasi Algoritma Naïve Bayes dengan Algoritma Genetika dan mengetahui atribut-atribut yang mempengaruhi kelulusan mahasiswa Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman.

Data Mining

Menurut Santosa dan Umam (2018) dalam karyanya menjelaskan bahwa *data mining* adalah kegiatan mengekstrak informasi atau pengetahuan penting dari suatu data berukuran besar dengan menggunakan teknik tertentu. Informasi atau pengetahuan yang dihasilkan dari *data mining* ini dapat digunakan untuk memperbaiki pengambilan keputusan. Dinamakan *data mining* atau penambangan data karena proses penemuan informasi atau pengetahuan dalam set data dilakukan seperti orang melakukan kegiatan penambangan.

Data mining bertujuan untuk menemukan pola yang sebelumnya tidak diketahui. Pada umumnya, pola ini sulit untuk ditemukan dengan eksplorasi data secara biasa, hal ini disebabkan oleh terlalu rumitnya hubungan antardata atau juga dapat disebabkan oleh jumlah data yang begitu besar. *Data mining* menggunakan analisis matematika dalam menemukan pola dari suatu data. Jika pola-pola tersebut telah ditemukan maka dapat digunakan untuk menyelesaikan berbagai macam permasalahan.

Klasifikasi

Klasifikasi adalah proses untuk mendapatkan sejumlah model yang menggambarkan, mengenali serta membedakan kelas data atau konsep yang bertujuan untuk menggunakan model yang telah

diperoleh agar dapat memprediksi kelas objek yang label kelasnya belum diketahui (Han, Kamber & Pei, 2012).

Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah metode klasifikasi yang menggunakan prinsip probabilitas untuk membuat model prediksi (Kurniawan, 2020). Berikut ini merupakan langkah-langkah pada Algoritma Naïve Bayes:

1. Data yang dikumpulkan akan diacak lalu dibagi menjadi dua, yaitu data *training* dan data *testing* dengan proporsi 80:20.
2. Membaca data *training* dan menghapus kelas pada data *testing*.
3. Menghitung nilai probabilitas setiap kelas dengan menggunakan persamaan (1).

$$p(h_j) = \frac{N_j}{N} \quad (1)$$

4. Menghitung nilai probabilitas setiap atribut terhadap setiap kelas pada data *training* (2).

$$\frac{p(A \cap B)}{p(B)} \quad (2)$$

5. Menghitung perkalian nilai probabilitas setiap atribut berdasarkan kondisi pada data *testing* dengan menggunakan nilai probabilitas data *training* terhadap masing-masing kelas.
6. Menghitung perkalian masing-masing atribut dengan nilai probabilitas setiap kelas menggunakan persamaan (3).

$$\text{prediksi kelas} = \underset{h_j}{\text{argmax}} p(h_j) \prod_{i=1}^n p(x_i|h_j) \quad (3)$$

7. Membandingkan nilai probabilitas kelas untuk menentukan kelas. Jika $p(h_1|x) < p(h_2|x)$, maka x diklasifikasikan sebagai h_2 (dan sebaliknya)
8. Mengevaluasi model klasifikasi menggunakan *confusion matrix* berdasarkan Gambar 1.

		Kelas Prediksi		
		Ya	Tidak	Jumlah
Kelas Aktual	Ya	TP	FN	P
	Tidak	FP	TN	N
	Jumlah	P'	N'	P+N

Gambar 1 Confusion Matrix

Rumus menghitung akurasi adalah sebagai berikut:

$$\text{akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Algoritma Genetika

Algoritma genetika adalah suatu teknik pencarian (*searching technique*) dan teknik optimasi yang cara kerjanya meniru proses evaluasi dan perubahan struktur genetika pada makhluk hidup (Arkeman, Seminar & Gunawan, 2012). Berikut ini merupakan langkah-langkah pada Algoritma Genetika:

1. Representasi Kromosom

- Representasi kromosom yang digunakan pada penelitian ini adalah representasi bilangan biner dimana setiap kromosom merupakan kombinasi nilai kehadiran setiap atribut.
2. Inisialisasi Populasi
Inisialisasi populasi dilakukan secara acak dengan jumlah populasi (n) = 5.
 3. Evaluasi Nilai *Fitness*
Setelah terbentuk populasi awal, evaluasi nilai *fitness* dengan menghitung nilai akurasi yang dihasilkan Naïve Bayes dari *Confusion Matrix* menggunakan atribut yang sesuai dengan kromosom masing-masing.
 4. Seleksi
Selanjutnya hasil populasi tersebut akan diseleksi. Penelitian ini menggunakan metode seleksi roda roulette (*roulette wheel selection*). Langkah-langkah seleksi dengan metode *roulette wheel* adalah sebagai berikut:
 - a. Menghitung total *fitness* pada populasi awal.
 - b. Menghitung probabilitas masing-masing kromosom untuk terpilih ke generasi berikutnya (p_k).
 - c. Menghitung probabilitas kumulatif (q_k) untuk masing-masing kromosom.
 - d. Kromosom dipilih dengan membangkitkan bilangan acak pada *range* [0,1] sebanyak ukuran populasi. Apabila $q_{n-1} < r_i < q_n$, kromosom v_n terpilih ke generasi berikutnya.
 5. *Crossover*
Lalu tahap *crossover*, pada penelitian ini metode perkawinan silang yang digunakan adalah perkawinan silang satu titik potong atau *one-cut point* dengan probabilitas *crossover* (P_c) yang akan digunakan sebesar 0,6. Probabilitas *crossover* digunakan untuk menentukan individu yang berhak dikawin silangkan atau tidak. Jika kromosom dengan nilai acak lebih kecil daripada nilai probabilitas *crossover*, maka kromosom tersebut akan dikawin silangkan dengan kromosom yang senilai dengannya.
 6. Mutasi
Tahap selanjutnya setelah melakukan proses *crossover* yaitu mutasi. Mutasi yang digunakan pada penelitian ini adalah dengan metode mutasi pengganti atau *flip mutation* dengan probabilitas mutasi (P_m) yang akan digunakan sebesar 0,1. Probabilitas mutasi menentukan jumlah kromosom yang mengalami mutasi dalam satu populasi. Proses mutasi dilakukan dengan mengganti satu gen dengan gen lainnya jika gen dengan nilai acak lebih kecil daripada nilai probabilitas mutasinya.
 7. Ulangi langkah 3 sampai dengan 6 hingga nilai *fitness* terbaik pada iterasi ke- n sama dengan nilai *fitness* terbaik pada iterasi ke- $(n + 1)$, $(n + 2)$ dan $(n + 3)$.

8. Kromosom dengan nilai *fitness* terbaik pada akhir generasi akan dipilih untuk klasifikasi.

Kelulusan

Predikat kelulusan mahasiswa telah diatur pada SN-Dikti, yaitu mahasiswa program diploma dan program sarjana dinyatakan lulus apabila telah menempuh beban belajar minimum sebanyak 144 sks dan memiliki capaian pembelajaran lulusan yang ditargetkan oleh program studi dengan indeks prestasi kumulatif (IPK) lebih besar atau sama dengan 2,00 (dua koma nol) (Kemenristekdikti, 2018).

**Hasil dan Pembahasan
Algoritma Naïve Bayes**

1. Menghitung Probabilitas Prior

Langkah pertama yang dilakukan adalah menghitung probabilitas *prior* dari setiap kelas pada data *training* sebagai berikut:

$$P(\text{Tepat Waktu}) = \frac{\text{mahasiswa tepat waktu}}{\text{jumlah seluruh mahasiswa}}$$

$$= \frac{50}{96}$$

$$= 0,52083$$

$$P(\text{Tidak Tepat Waktu}) = \frac{\text{mahasiswa tidak tepat waktu}}{\text{jumlah seluruh mahasiswa}}$$

$$= \frac{46}{96}$$

$$= 0,47917$$

2. Menghitung Probabilitas Likelihood pada Data Training

Selanjutnya menghitung probabilitas *likelihood* pada setiap atribut terhadap masing-masing kelas dari data *training*. Berikut ini hasil perhitungan probabilitas *likelihood* pada setiap atribut terhadap masing-masing kelas:

Tabel 1. Atribut Jenis Kelamin

	Jumlah		Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu	Tepat Waktu	Tidak Tepat Waktu
Laki-Laki	8	19	0,16	0,413043
Perempuan	42	27	0,84	0,586957
Jumlah	50	46	1	1

Tabel 2. Atribut Asal Daerah

	Jumlah		Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu	Tepat Waktu	Tidak Tepat Waktu
Samarinda	23	21	0,46	0,456522
Luar Samarinda	27	25	0,54	0,543478
Jumlah	50	46	1	1

Tabel 3. Atribut Jalur Masuk

	Jumlah		Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu	Tepat Waktu	Tidak Tepat Waktu
SBMPTN	16	8	0,32	0,173913
SNMPTN	23	24	0,46	0,521739
SMMPTN	11	14	0,22	0,304348
Jumlah	50	46	1	1

Tabel 4. Atribut Status Pekerjaan

	Jumlah		Probabilitas	
	Tepat Waktu	Tidak Tepat Waktu	Tepat Waktu	Tidak Tepat Waktu
Bekerja	4	1	0,08	0,02173
Tidak Bekerja	46	45	0,92	0,97826
Jumlah	50	46	1	1

3. Menghitung Probabilitas Likelihood pada Data Testing

Berdasarkan probabilitas *likelihood* pada data *training* akan dihitung probabilitas *likelihood* pada data *testing* dengan ketentuan nilai masing-masing kriteria. Berikut ini contoh perhitungan probabilitas *likelihood* pada data *testing*:

Tabel 5. Data Testing

Data Ke-	Jenis Kelamin	Asal Daerah	Jalur Masuk	Status Pekerjaan
1	L	Luar Samarinda	SNMPTN	Tidak Bekerja

$$\begin{aligned}
 P(\text{Atribut}|\text{Tepat Waktu}) &= P(\text{Laki-laki}|\text{Tepat Waktu}) \times \\
 &P(\text{Luar Samarinda}|\text{Tepat Waktu}) \times \\
 &P(\text{SNMPTN}|\text{Tepat Waktu}) \times \\
 &P(\text{Tidak Bekerja}|\text{Tepat waktu}) \\
 &= 0,16 \times 0,54 \times 0,46 \times 0,9 \\
 &= 0,03656
 \end{aligned}$$

$$\begin{aligned}
 P(\text{Atribut}|\text{Tidak Tepat Waktu}) &= P(\text{Laki-laki}|\text{Tidak Tepat Waktu}) \times \\
 &P(\text{Luar Samarinda}|\text{Tidak Tepat Waktu}) \times \\
 &P(\text{SNMPTN}|\text{Tidak Tepat Waktu}) \times \\
 &P(\text{Tidak Bekerja}|\text{Tidak Tepat waktu}) \\
 &= 0,413043 \times 0,543478 \times 0,521739 \times \\
 &0,978261 \\
 &= 0,11457
 \end{aligned}$$

4. Menghitung Probabilitas Posterior

Kemudian nilai dari masing-masing kriteria tersebut dikalikan dengan probabilitas kelas masing-masing dengan menggunakan data *training* dan didapat hasil sebagai berikut :

$$\begin{aligned}
 P(\text{Tepat Waktu}|\text{Atribut}) &= P(\text{Atribut}|\text{Tepat Waktu}) \times \\
 &P(\text{Tepat Waktu}) \\
 &= 0,03656 \times 0,52083 \\
 &= 0,01904
 \end{aligned}$$

$$\begin{aligned}
 P(\text{Tidak Tepat Waktu}|\text{Atribut}) &= P(\text{Atribut}|\text{Tidak Tepat Waktu}) \times \\
 &P(\text{Tidak Tepat Waktu}) \\
 &= 0,11457 \times 0,47917 \\
 &= 0,0549
 \end{aligned}$$

Setelah itu, dilakukan perbandingan antara kedua kelas. Dari perhitungan di atas didapatkan nilai probabilitas *posterior* Tepat Waktu sebesar 0,01904 dan nilai probabilitas *posterior* Tidak Tepat Waktu sebesar 0,0549. Berdasarkan nilai tersebut dapat diambil kesimpulan bahwa probabilitas *posterior* Tidak Tepat Waktu lebih besar dari nilai probabilitas *posterior* Tepat Waktu sehingga data ke-1 diprediksi masuk kedalam kelas Tidak Tepat Waktu. Perbandingan kedua kelas ini dilakukan sebanyak jumlah data *testing*.

5. Menghitung Nilai Akurasi Algoritma Naïve Bayes

Pada penelitian ini, akan dihitung akurasi algoritma Naïve Bayes dengan menggunakan *confusion matrix*. Hasil pengujian yang didapatkan pada metode Naïve Bayes dapat dilihat pada tabel berikut.

Tabel 6. Perhitungan Akurasi dengan Confusion Matrix

	Ya	Tidak	Jumlah
Ya	7	5	12
Tidak	7	5	12
Jumlah	14	10	

Pada tabel *confusion matrix*, Tabel 6, dapat dihitung nilai akurasi sebagai berikut:

$$\begin{aligned}
 \text{akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} \\
 \text{akurasi} &= \frac{7+5}{7+5+7+5} = \frac{12}{24} = 0,5
 \end{aligned}$$

Hasil pengukuran data akurasi yang diperoleh dari data *training* mencapai 50%. Jumlah prediksi mahasiswa lulus tepat waktu yang diklasifikasikan lulus tepat waktu yaitu 7 data dan jumlah prediksi mahasiswa lulus tidak tepat waktu yang diklasifikasikan sebagai lulus tepat waktu yaitu 5. Jumlah prediksi mahasiswa lulus tepat waktu yang diklasifikasikan lulus tidak tepat waktu yaitu 7 data dan jumlah prediksi mahasiswa lulus tidak tepat waktu yang diklasifikasikan sebagai lulus tidak tepat waktu yaitu 5.

Algoritma Genetika

1. Representasi Kromosom

Untuk Setiap atribut didefinisikan menjadi

$A_1 - A_4$, dimana

A_1 : Jenis Kelamin

A_2 : Asal Daerah

A_3 : Jalur Masuk

A_4 : Status Pekerjaan Mahasiswa

2. Inisialisasi Populasi

Inisialisasi dilakukan dengan kombinasi nilai kehadiran atribut secara acak. Ukuran populasi yang akan dibangkitkan berjumlah (n) = 5

C_1 : [1, 1, 1, 0]

C_2 : [0, 1, 1, 0]

C_3 : [1, 0, 0, 1]

C_4 : [1, 0, 0, 0]

C_5 : [0, 0, 1, 0]

3. Evaluasi Nilai Fitness

Evaluasi nilai fitness dilakukan dengan menghitung nilai akurasi yang dihasilkan Naïve Bayes dengan *confusion matrix* berdasarkan atribut yang sesuai dengan masing-masing kromosom. Hasil perhitungan akurasi dari 5 populasi adalah sebagai berikut:

$Fitness C_1 = 0,458333$

$Fitness C_2 = 0,625$

$Fitness C_3 = 0,5$

$Fitness C_4 = 0,458333$

$Fitness C_5 = 0,625$

Dari nilai *fitness* diatas, diketahui total *fitness* adalah 2,666666.

4. Seleksi

Proses seleksi diperlukan untuk mengevaluasi calon kromosom induk. Metode seleksi yang akan digunakan yaitu seleksi *Roulette Wheel*. Tahapan awal yang dilakukan, yaitu menghitung nilai probabilitas masing-masing kromosom dengan membagi nilai *fitness* dengan total nilai *fitness*.

$$P(C_1) = \frac{\text{Fitness } C_1}{\text{Total Fitness}} = \frac{0,458333}{2,666666} = 0,171875$$

$$P(C_2) = \frac{\text{Fitness } C_2}{\text{Total Fitness}} = \frac{0,625}{2,666666} = 0,234375$$

$$P(C_3) = \frac{\text{Fitness } C_3}{\text{Total Fitness}} = \frac{0,5}{2,666666} = 0,1875$$

$$P(C_4) = \frac{\text{Fitness } C_4}{\text{Total Fitness}} = \frac{0,458333}{2,666666} = 0,171875$$

$$P(C_5) = \frac{\text{Fitness } C_5}{\text{Total Fitness}} = \frac{0,625}{2,666666} = 0,234375$$

Langkah selanjutnya adalah menghitung probabilitas kumulatif dengan menjumlahkan nilai probabilitas sebelumnya dengan nilai probabilitas sekarang, proses perhitungannya adalah sebagai berikut:

$$P_k(C_1) = 0,171875$$

$$P_k(C_2) = 0,171875 + 0,234375 = 0,40625$$

$$P_k(C_3) = 0,40625 + 0,1875 = 0,59375$$

$$P_k(C_4) = 0,59375 + 0,171875 = 0,765625$$

$$P_k(C_5) = 0,765625 + 0,234375 = 1$$

Bilangan acak antara 0 sampai 1 dibangkitkan sebanyak jumlah kromosom.

$$R_1 = 0,247062$$

$$R_2 = 0,919782$$

$$R_3 = 0,952602$$

$$R_4 = 0,182819$$

$$R_5 = 0,618726$$

Selanjutnya memilih kromosom ke-n sebagai orang tua dengan syarat $P_k(C_{n-1}) < R_i < P_k(C_n)$.

Hasil Seleksi:

$$C_1[1, 1, 1, 0] = C_1'[0, 1, 1, 0]$$

$$C_2[0, 1, 1, 0] = C_2'[0, 0, 1, 0]$$

$$C_3[0, 0, 0, 1] = C_3'[0, 0, 1, 0]$$

$$C_4[1, 0, 0, 0] = C_4'[0, 1, 1, 0]$$

$$C_5[0, 0, 1, 0] = C_5'[1, 0, 0, 0]$$

5. Crossover

Crossover dilakukan dengan menggunakan metode penyilangan satu titik (*one-cut point*), dimana dilakukan pemilihan posisi secara acak dari kromosom induk untuk saling menukar gen. Proses *crossover* dipengaruhi oleh bilangan acak yang dibangkitkan dan probabilitas *crossover*. Probabilitas *crossover* (P_c) yang digunakan adalah 0,6. Selanjutnya bilangan acak antara 0 sampai 1 dibangkitkan sebanyak jumlah kromosom, jika bilangan acak tersebut lebih kecil atau sama dengan probabilitas *crossover*, maka kromosom tersebut akan diganti dengan hasil *crossover* dua kromosom induk.

$$R_1 = 0,126787$$

$$R_2 = 0,947277$$

$$R_3 = 0,934965$$

$$R_4 = 0,357764$$

$$R_5 = 0,212860$$

Kromosom yang akan diganti dengan hasil *crossover* adalah C_1' , C_4' dan C_5' . Penyilangan akan dilakukan sepasang-sepasang. Karena induk berjumlah 3 (ganjil), maka salah satunya harus dibuang (misal C_5'). Pada penyilangan C_1' dengan C_4' dibangkitkan bilangan acak antara 1 sampai panjang kromosom (=4), didapatkan bilangan acak 1 yang artinya posisi *crossover* ada pada bit ke-1.

$$C_1' \times C_4'$$

$$C_1'[0, 1, 1, 0] = [0, 1, 1, 0]$$

$$C_4'[0, 1, 1, 0] = [0, 1, 1, 0]$$

Populasi baru setelah dilakukan proses *crossover*:

$$C_1''[0, 1, 1, 0]$$

$$C_2'[0, 0, 1, 0]$$

$$C_3'[0, 0, 1, 0]$$

$$C_4''[0, 1, 1, 0]$$

$$C_5'[1, 0, 0, 0]$$

6. Mutasi

Jumlah kromosom yang mengalami mutasi dalam satu populasi ditentukan oleh probabilitas mutasi (P_m). Probabilitas mutasi yang digunakan yaitu 0,1. Proses mutasi dilakukan dengan cara membangkitkan bilangan acak sebanyak ukuran populasi \times panjang kromosom ($5 \times 4 = 20$). Kemudian gen-gen yang memiliki bilangan acak lebih kecil dari probabilitas mutasi diganti dengan gen baru. Apabila gen terpilih bernilai 1 maka nilai gen pada kromosom selanjutnya diganti dengan 0 dan demikian sebaliknya. Selanjutnya bilangan acak antara 0 sampai 1 dibangkitkan sebanyak jumlah mutasi.

$$R_1 = [0,042284 \quad 0,034786 \quad 0,721812 \quad 0,394879]$$

$$R_2 = [0,389706 \quad 0,475952 \quad 0,010241 \quad 0,967564]$$

$$R_3 = [0,301233 \quad 0,870073 \quad 0,819020 \quad 0,556991]$$

$$R_4 = [0,716622 \quad 0,216582 \quad 0,754797 \quad 0,915283]$$

$$R_5 = [0,366128 \quad 0,153907 \quad 0,550798 \quad 0,279247]$$

diperoleh posisi gen yang akan dilakukan mutasi adalah sebagai berikut:

$$C_1'' = [0, 1, 1, 0]$$

$$C_2' = [0, 0, 1, 0]$$

$$C_3' = [0, 0, 1, 0]$$

$$C_4'' = [0, 1, 1, 0]$$

$$C_5' = [1, 0, 0, 0]$$

Populasi baru setelah dilakukan proses mutasi:

$$C_1^*[1, 0, 1, 0]$$

$$C_2^*[0, 0, 0, 0]$$

$$C_3'[0, 0, 1, 0]$$

$$C_4''[0, 1, 1, 0]$$

$$C_5'[1, 0, 0, 0]$$

dihitung kembali nilai *fitness* pada nilai akurasi yang dihasilkan Naïve Bayes dengan menggunakan *confusion matrix* berdasarkan kriteria yang sesuai dengan masing-masing kromosom. Hasil perhitungan akurasi dari 5 kromosom adalah sebagai berikut:

$Fitness C_1 = 0,458333$

$Fitness C_2 = 0,5$

$Fitness C_3 = 0,625$

$Fitness C_4 = 0,625$

$Fitness C_5 = 0,458333$

dari nilai *fitness* diatas, didapatkan nilai *fitness* terbaik, yaitu 0,625 pada kromosom ke-3 dan kromosom ke-4.

Hasil populasi pada proses mutasi akan menjadi populasi baru yang digunakan untuk proses iterasi berikutnya. Kemudian nilai *fitness* terbaik pada setiap iterasi akan disimpan untuk dibandingkan dengan nilai *fitness* terbaik pada iterasi selanjutnya.

Iterasi berhenti pada iterasi ke-10 karena nilai *fitness* pada iterasi ke-7 tidak berubah hingga iterasi ke-10. Pada iterasi ke-10 didapatkan nilai *fitness* terbaik sebesar 0,67 dengan kromosom terpilih [0, 1, 1, 1]. Artinya, atribut terpilih yang menghasilkan nilai akurasi terbaik adalah asal daerah, jalur masuk dan status pekerjaan mahasiswa.

Kesimpulan

Berdasarkan hasil dan pembahasan dari penelitian yang telah dilakukan dapat disimpulkan bahwa:

1. Prediksi kelulusan pada mahasiswa Jurusan Matematika FMIPA Universitas Mulawarman dengan menggunakan algoritma Naïve Bayes menghasilkan akurasi yaitu sebesar 50% didapatkan dengan kondisi perbandingan data *training* sebesar 80% dan data *testing* sebesar 20%. Lalu atribut dioptimasi menggunakan algoritma genetika sehingga akurasi meningkat sebesar 16,67% menjadi 66,67%.
2. Prediksi kelulusan pada mahasiswa Jurusan Matematika FMIPA Universitas Mulawarman dengan algoritma Naïve Bayes menggunakan atribut jenis kelamin, asal daerah, jalur masuk dan status pekerjaan mahasiswa. Setelah dioptimasi menggunakan algoritma genetika didapatkan atribut yang mempengaruhi kelulusan pada mahasiswa Jurusan Matematika FMIPA Universitas Mulawarman, yaitu asal daerah, jalur masuk dan status pekerjaan mahasiswa.

Daftar Pustaka

- Arhami, M & Nasir, M. (2020). *Data Mining: Algoritma dan Implementasi*. Yogyakarta: Andi.
- Arkeman, Y, Seminar, K. B & Gunawan, H. (2012). *Algoritma Genetika: Teori dan Aplikasinya untuk Bisnis dan Industri*. Bogor: IPB Press.
- Busono, S. (2020). *Optimasi Naïve Bayes Menggunakan Algoritma Genetika sebagai*

Seleksi Fitur untuk Memprediksi Performa Siswa. Jurnal Ilmiah Teknologi Informasi Asia, 14(1), 31-40.

Fithri, D. L & Darmanto E. (2014). *Sistem Pendukung Keputusan untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes*. Prosiding SNATIF.

Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. ISRL 12. Springerlink.

Han, J, Kamber, M & Pei, J. (2012). *Data Mining: Concepts and Techniques. Third Edition*. USA: Elsevier.

Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi. (2018). *Standar Nasional Pendidikan Tinggi (SN-DIKTI) Berdasarkan Permenristekdikti RI Nomor 44 tahun 2015 dan Perubahan Permenristekdikti RI Nomor 50 tahun 2018*.

Kurniawan, D. (2020). *Pengenalan Machine Learning dengan Python*. Jakarta: PT Elex Media Komputindo.

Lembang, F. K & Fendjalang, M. (2015). *Klasifikasi Variabel Penentu Kelulusan Mahasiswa FMIPA Unpatti Menggunakan Metode CHAID*. Jurnal Statistika, 15(1), 1-6.

Suryani, N & Priyanti, E. (2019). *Optimasi Naïve Bayes dan Algoritma Genetika untuk Prediksi Penerimaan Beasiswa Pendidikan pada SMP Utama*. Jurnal Teknik Komputer AMIK BSI, 5(2), 189-196.

Suyanto. (2019). *Data Mining untuk Klasifikasi dan Klasterisasi Data. Edisi Revisi*. Bandung: Informatika.