

# Peningkatan Algoritma Naïve Bayes Menggunakan Algoritma Genetika Pada Klasifikasi Bakteri

Evy Priyanti

Universitas Bina Sarana Informatika  
Email : [evy.evp@bsi.ac.id](mailto:evy.evp@bsi.ac.id)

## Abstraksi

Pada penelitian sebelumnya hanya menggunakan algoritma naïve bayes dan menghasilkan nilai akurasi sebesar 80.93% dan saat ini nilai akurasi akan ditingkatkan dengan menggunakan algoritma genetika. Pola pembelajaran pada algoritma genetika dapat menginformasikan adanya pendatang baru atau klasifikasi baru dengan waktu yang lebih cepat. Kesulitan dalam konfigurasi yang ada pada naïve bayes dapat terbantu oleh algoritma genetika ini selain itu dapat menyiapkan pemodelan yang memadai untuk menggambarkan sistem. Bakteri terdiri dari tiga klasifikasi bakteri: berdasarkan cara memperoleh makanan, berdasarkan pewarnaan gram, dan berdasarkan bentuknya. Pada data bakteri ini yang terdiri dari parameter numerik yang berisi sequence name, Mcg, gvh, Lip, chg, aac, alm1, alm2 dan class distributron sejumlah 336 atribut, pada class distribution terdapat 8 kelas protein yang diklasifikasikan yaitu sitoplasma (cp), membran dalam tanpa urutan sinyal (im), perisplasm (pp), dalam membran dengan uncleavable sinyal urutan (IMU), luar membran (om), luar membran lipoprotein (OML), dalam membran lipoprotein (IML), membran dalam dengan cleavable urutan sinyal (IMS). Hasil dari penelitian ini menghasilkan nilai akurasi menjadi 81.19%. Dengan demikian terbukti bahwa algoritma naïve bayes dapat ditingkatkan performanya dengan menggunakan algoritma genetika pada klasifikasi bakteri.

**Kata Kunci** : Naïve Bayes, Algoritma Genetika, bakteri, akurasi

## Abstract

*In previous studies, only using the nave Bayes algorithm and resulted in an accuracy value of 80.93% and currently the accuracy value will be increased by using a genetic algorithm. Learning patterns in genetic algorithms can inform new entrants or new classifications with a faster time. Difficulties in the configuration that exist in nave Bayes can be helped by this genetic algorithm in addition to being able to provide adequate modeling to describe the system. Bacteria consists of three classifications of bacteria: based on how they obtain food, based on gram staining, and based on their shape. In this bacterial data, which consists of numerical parameters containing the sequence name, Mcg, gvh, Lip, chg, aac, alm1, alm2 and a class distributron of 336 attributes, in the class distribution there are 8 protein classes classified, namely cytoplasm (cp), membrane inside without signal sequence (im), perisplasm (pp), inside membrane with uncleavable signal sequence (IMU), outer membrane (om), outer membrane lipoprotein (OML), inside membrane lipoprotein (IML), inner membrane with cleavable signal sequence ( STI). The results of this study resulted in an accuracy value of 81.19%. Thus, it is proven that the performance of the nave Bayes algorithm can be improved by using a genetic algorithm for bacterial classification.*

**Keyword** : Naïve Bayes, Genetic Algorithm, bacteria, accuracy

## 1. Pendahuluan

Jenis bakteri dapat dibedakan menjadi beberapa jenis diantaranya adalah klasifikasi bakteri berdasarkan cara mereka memperoleh makanan. Ada bakteri yang disebut dengan Autotrof atau bakteri

fotosintetik yang merupakan bakteri yang dapat membuat makanannya sendiri yang berasal dari bahan anorganik berdasarkan proses fotosintesis. Fotosintesis sendiri membutuhkan hydrogen sulfida yang berasal dari air dan pigmen hijau atau ungu seperti pada klorofil pada proses fotosintesis

tumbuhan. Fotosintesis tersebut membuat hydrogen sulfida dan electron menjadi energi dan belerang.

Selain itu ada juga bakteri kemautotrof yaitu jenis bakteri yang tidak memiliki pigmen klorofil sehingga tidak dapat melakukan proses fotosintesis atau pemanfaatan sinar matahari untuk makan, sebagai gantinya bakteri ini menggunakan reaksi gelap yaitu mensintesis hydrogen sulfida dan oksigen menjadi karbonat, karbon dioksida dan sulfur. Bakteri juga dibedakan berdasarkan pewarnaan gram diantaranya adalah bakteri gram positif yaitu bakteri yang memiliki peptidoglikan atau dinding sel yang tebal dengan struktur yang rumit, namun berlapis-lapis. Lapisan peptidoglikan ini yang akan membuat bakteri gram positif dapat mempertahankan warnanya pada pengujian kristal violet. Bakteri ini dapat bertahan bila diganggu fisik akan tetapi rentan jika diberi antibiotik penisilin. Yang termasuk bakteri gram positif adalah streptococcus yang menyebabkan penyakit radang paru dan radang otak. Selain itu ada juga bakteri gram negative yaitu bakteri yang memiliki peptidoglikan yang tipis dan tidak dapat mempertahankan pewarnaan ungu pada pengujian karena dindingnya yang sangat tipis, namun bakteri ini lebih kuat terhadap antibiotik penisilin. (Kompas, 2020)

Bakteri gram negatif menyebabkan infeksi termasuk pneumonia, infeksi aliran darah, infeksi luka atau tempat operasi, dan meningitis dalam pengaturan perawatan kesehatan. Bakteri gram negatif resisten terhadap beberapa obat dan semakin resisten terhadap sebagian besar antibiotik yang tersedia. Bakteri ini memiliki kemampuan bawaan untuk menemukan cara baru untuk menjadi resisten dan dapat meneruskan materi genetik yang memungkinkan bakteri lain menjadi resisten terhadap obat juga. Rekomendasi agresif CDC, jika diterapkan, dapat mencegah penyebaran gram negatif (Prevention, 2021).

Identifikasi pada setiap pasien yang terkena bakteri juga berbeda-beda tergantung beberapa hal diantaranya identifikasi pilihan pengobatan untuk infeksi gram negatif, berdasarkan lokasi dan jenis infeksi, mekanisme bakteri gram negatif menjadi resisten terhadap antibiotik, modalitas untuk meningkatkan koordinasi perawatan di antara anggota tim interprofesional untuk meningkatkan hasil

bagi pasien yang terkena bakteri gram negative (Oliveira & Reygaert., 2021).

Pasien akan dianalisa bakteri apa yang hinggap pada tubuh pasien dengan jumlah data pada bakteri gram negative ini berjumlah 336 atribut yang terdiri dari sequence name, Mcg, gvh, Lip, chg, aac,alm1,alm2 dan class distributron, akan diteliti menggunakan algoritma naïve bayes dan algoritma genetika.

## 2. Metode Penelitian

Pada penelitian kali ini peneliti akan menguji validasi dari bakteri dengan algoritma Naïve bayes dan algoritma genetic untuk mengklasifikasi jenis bakteri. Metode Naive Bayes adalah seperangkat algoritma pembelajaran terawasi berdasarkan penerapan teorema Bayes dengan asumsi "naif" dari independensi bersyarat antara setiap pasangan fitur yang diberikan nilai variabel kelas. Teorema Bayes menyatakan hubungan berikut, mengingat variable  $y$  kelas dan vektor fitur dependen:  $x_1, \dots, x_n$  (Zhang, 2004).

Pengklasifikasi Naive Bayes adalah kumpulan algoritma klasifikasi berdasarkan Teorema Bayes. Ini bukan algoritma tunggal tetapi keluarga algoritma di mana semuanya memiliki prinsip yang sama, yaitu setiap pasangan fitur yang diklasifikasikan independen satu sama lain. (Geeksforgeeks, 2020).

Kumpulan data bakteri terdiri dari tiga klasifikasi bakteri: berdasarkan cara memperoleh makanan, berdasarkan pewarnaan gram, dan berdasarkan bentuknya. Parameter numerik yang berisi dataset adalah sequence name, Mcg, gvh, Lip, chg, aac,alm1,alm2 dan class distributron. Dengan data ini penulis akan memprediksi klasifikasi bakteri berdasarkan parameter ini. Data terdiri dari nilai numerik berkelanjutan yang menggambarkan dimensi fitur masing-masing. Bahkan jika fitur-fitur ini bergantung satu sama lain atau pada keberadaan fitur-fitur lain, pengklasifikasi Naive Bayes akan mempertimbangkan semua properti ini untuk berkontribusi secara independen terhadap kemungkinan bahwa bakteri itu milik spesies tertentu. (Consulting, 2019)

Teorema Bayes sangat penting untuk statistik inferensial dan banyak pembelajaran mesin tingkat lanjut. Model. Penalaran Bayesian adalah pendekatan logis untuk memperbarui probabilitas hipotesis dalam terang bukti baru, dan

karena itu memainkan peran penting dalam sains (Berrar, 2018).

Algoritma genetika, program komputer yang mensimulasikan evolusi alam, semakin diterapkan di banyak disiplin ilmu, dan telah banyak digunakan untuk memecahkan berbagai masalah optimasi dari pencarian arsitektur jaringan nyata untuk permainan strategis, dan untuk memodelkan fenomena adaptasi dan pembelajaran. Dalam penelitian ini peneliti menyajikan algoritma genetika, kualitasnya, batasannya, dan tantangan, serta beberapa perspektif tentang bakteri. Algoritma genetika mampu menjelajahi ruang yang besar dan kompleks solusi yang mungkin, untuk dengan cepat menemukan elemen yang menjanjikan, dan menyediakan alat pemodelan yang memadai untuk menggambarkan sistem evolusi, dari dan untuk ekonomi. Namun biaya komputasi yang tinggi, kesulitan konfigurasi parameter kultus, dan representasi penting dari solusi.

Konsep metode kontrol parameter yang kuat, dan pendekatan baru dalam strategi representasi, menjadi kunci untuk mengatasi keterbatasan dalam klasifikasi bakteri ini. Tinjauan kompilasi ini bertujuan untuk menginformasikan praktisi dan pendaftar baru di lapangan sama dalam penelitian algoritma genetika (ymeric Vi'e, 2021).

Pada penelitian sebelumnya hanya menggunakan algoritma naïve bayes dan menghasilkan nilai akurasi sebesar 80.93% dan saat ini nilai akurasi akan ditingkatkan dengan menggunakan algoritma genetika. Pola pembelajaran pada algoritma genetika dapat menginformasikan adanya pendaftar baru atau klasifikasi baru dengan waktu yang lebih cepat. Kesulitan dalam konfigurasi yang ada pada naïve bayes dapat terbantu oleh algoritma genetika ini selain itu dapat menyiapkan pemodelan yang memadai untuk menggambarkan sistem.

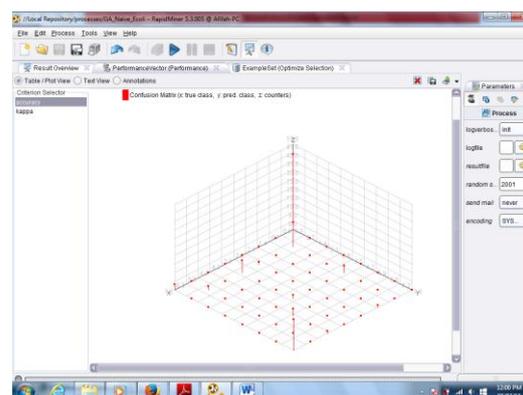
### 3. Hasil dan Pembahasan

Algoritma Naïve Bayes akan kelas dari atribut yang ada pada bakteri kemudian dari kelas yang sudah terbentuk akan dibantu dipersiapkan pemodelan dalam klasifikasi dengan menggunakan algoritma genetika dan dari hasil penelitian yang sudah diuji dapat terbukti bahwa dengan algoritma naïve bayes dan algoritma genetika dapat meningkatkan nilai akurasi hingga sebesar 81.19%.

	true ip	true im	true iml	true imj	true im	true oml	true op
pred ip	44	0	0	0	0	0	0
pred im	0	14	0	1	7	0	2
pred iml	0	0	0	0	0	0	0
pred imj	0	0	0	0	0	0	0
pred im	0	4	0	0	2	0	0
pred om	0	0	0	0	4	0	0
pred oml	0	0	0	0	0	0	0
pred op	0	0	1	0	0	1	18
class recall	100.00%	77.78%	0.00%	0.00%	77.22%	80.00%	85.71%

Gambar 1. Nilai Akurasi NB dan GA  
Sumber : Hasil Penelitian

Pada gambar 1 terlihat hasil yang didapat pada pengujian antara naïve bayes dengan algoritma genetika dapat meningkatkan nilai akurasi sebesar 81.19% dibandingkan hanya menggunakan algoritma naïve bayes saja yaitu nilai akurasi sebesar 80.93%.



Gambar 2. Confusion Matrix  
Sumber : Hasil Penelitian

Gambar 2 pada hasil penelitian grafik confusion matrik yg menunjukkan pengukuran performa untuk masalah klasifikasi machine learning dimana keluaran berupa class distribution dari bakteri mendapatkan nilai yang actual dan akurasi yang tinggi.

The screenshot shows a software window titled 'PerformanceVector' with a table of numerical data. The table is organized into two sections, each with a 'ConfusionMatrix' header. The data includes values for attributes like 'op', 'im', 'sm', 'ml', 'ml2', 'ml3', 'ml4', 'ml5', 'ml6', 'ml7', 'ml8', 'ml9', 'ml10', 'ml11', 'ml12', 'ml13', 'ml14', 'ml15', 'ml16', 'ml17', 'ml18', 'ml19', 'ml20', 'ml21', 'ml22', 'ml23', 'ml24', 'ml25', 'ml26', 'ml27', 'ml28', 'ml29', 'ml30', 'ml31', 'ml32', 'ml33', 'ml34', 'ml35', 'ml36', 'ml37', 'ml38', 'ml39', 'ml40', 'ml41', 'ml42', 'ml43', 'ml44', 'ml45', 'ml46', 'ml47', 'ml48', 'ml49', 'ml50', 'ml51', 'ml52', 'ml53', 'ml54', 'ml55', 'ml56', 'ml57', 'ml58', 'ml59', 'ml60', 'ml61', 'ml62', 'ml63', 'ml64', 'ml65', 'ml66', 'ml67', 'ml68', 'ml69', 'ml70', 'ml71', 'ml72', 'ml73', 'ml74', 'ml75', 'ml76', 'ml77', 'ml78', 'ml79', 'ml80', 'ml81', 'ml82', 'ml83', 'ml84', 'ml85', 'ml86', 'ml87', 'ml88', 'ml89', 'ml90', 'ml91', 'ml92', 'ml93', 'ml94', 'ml95', 'ml96', 'ml97', 'ml98', 'ml99', 'ml100'. The values are mostly 0, with some non-zero values in the first section.

Gambar 3. Performance Vector  
Sumber: Hasil Penelitian

Gambar 3 menunjukkan hasil dari performance vector dimana setiap atribut terlihat jelas nilainya.

The screenshot shows a software window titled 'Kappa 0.734' with a table of numerical data. The table is organized into two sections, each with a 'ConfusionMatrix' header. The data includes values for attributes like 'op', 'im', 'sm', 'ml', 'ml2', 'ml3', 'ml4', 'ml5', 'ml6', 'ml7', 'ml8', 'ml9', 'ml10', 'ml11', 'ml12', 'ml13', 'ml14', 'ml15', 'ml16', 'ml17', 'ml18', 'ml19', 'ml20', 'ml21', 'ml22', 'ml23', 'ml24', 'ml25', 'ml26', 'ml27', 'ml28', 'ml29', 'ml30', 'ml31', 'ml32', 'ml33', 'ml34', 'ml35', 'ml36', 'ml37', 'ml38', 'ml39', 'ml40', 'ml41', 'ml42', 'ml43', 'ml44', 'ml45', 'ml46', 'ml47', 'ml48', 'ml49', 'ml50', 'ml51', 'ml52', 'ml53', 'ml54', 'ml55', 'ml56', 'ml57', 'ml58', 'ml59', 'ml60', 'ml61', 'ml62', 'ml63', 'ml64', 'ml65', 'ml66', 'ml67', 'ml68', 'ml69', 'ml70', 'ml71', 'ml72', 'ml73', 'ml74', 'ml75', 'ml76', 'ml77', 'ml78', 'ml79', 'ml80', 'ml81', 'ml82', 'ml83', 'ml84', 'ml85', 'ml86', 'ml87', 'ml88', 'ml89', 'ml90', 'ml91', 'ml92', 'ml93', 'ml94', 'ml95', 'ml96', 'ml97', 'ml98', 'ml99', 'ml100'. The values are mostly 0, with some non-zero values in the first section.

Gambar 4. Nilai Kappa  
Sumber : Hasil Penelitian

Konsistensi pengukuran dengan tabel kappa menunjukkan nilai 0,734 yang berarti nilai kesepakatan antara klasifikasi adalah kuat dan hampir mendekati nilai sangat kuat karena sangat dekat dengan nilai 0,81 pada penilaian kappa.

#### 4. Kesimpulan

Penelitian ini dilakukan untuk mendapatkan hasil penelitian yang terbaik dengan peningkatan nilai akurasi dikarenakan penambahan fitur yaitu penerapan algoritma genetika pada untuk meningkatkan konfigurasi yang ada pada naive bayes selain itu juga membantu dalam pemodelan system dalam mempermudah klasifikasi bakteri. Hasil yang didapatkan pada penelitian ini adalah nilai akurasi sebesar 81.19%.

Akan tetapi karena keterbatasan penelitian ini perlu dilakukan penelitian

lanjutan yang berkaitan dengan klasifikasi untuk mendapatkan hasil yang lebih baik. Adapun saran-saran yang perlu diberikan yaitu memungkinkan untuk dilakukan penelitian yang sejenis dengan metode seleksi dalam pengklasifikasian misalnya feature selection.

#### Referensi

- Berrar, D. (2018). *Bayes Theorem and Naive Bayes Classifier*. Elsevier pp 403-412.
- Consulting, A. (2019, November 08). *Floydhub*. Retrieved from Floydhub: <https://blog.floydhub.com/naive-bayes-for-machine-learning/>
- Geeksforgeeks. (2020, May 15). *Geeksforgeeks*. Retrieved from Geeksforgeeks: <https://www.geeksforgeeks.org/naive-bayes-classifiers/>
- Kompas. (2020, Oktober 05). *Skola*. Retrieved from <https://www.kompas.com/skola/read/2020/10/05/194554769/klasifikasi-bakteri?page=all>
- Oliveira, J., & Reygaert., W. C. (2021, March 29). *Gram Negative Bacteria*. Retrieved from ncbi: <https://www.ncbi.nlm.nih.gov/books/NBK538213/>
- Prevention, C. f. (2021). *Centers for Disease Control and Prevention*. Retrieved from <https://www.cdc.gov/hai/organisms/gram-negative-bacteria.html>
- Ymeric Vi'e, A. M. (2021). *Qualities, challenges and future of genetic algorithms: a literature review*. Stanford University.
- Zhang, H. (2004). The Optimality of Naive Bayes. *Proc. FLAIRS*.