

# Analisis Kinerja Algoritma Klasifikasi *Naïve Bayes* Menggunakan *Genetic algorithm* dan *Bagging* untuk Data Publik Risiko Transaksi Kartu Kredit

Manda Agustriya<sup>a1</sup>, Munirul Ula<sup>a2</sup>, Kurniawati<sup>a3</sup>

<sup>a</sup>Teknik Informatika, Fakultas Teknik, Universitas Malikussaleh

Jln. Kampus Unimal Bukit Indah, Blang Pulo, Kec. Muara Satu.

Kabupaten Aceh Utara, Aceh, 24355

<sup>1</sup>manda.200170141@mhs.unimal.ac.id

<sup>2</sup>munirulula@unimal.ac.id

<sup>3</sup>kurniawati@unimal.ac.id

## Abstrak

Peningkatan penggunaan kartu kredit telah meningkatkan risiko penipuan dan kejahatan terkait transaksi kartu kredit. Hal ini memerlukan pengembangan metode yang efektif dalam mendeteksi dan mengklasifikasikan risiko tersebut secara akurat. Tujuan penelitian ini adalah untuk membandingkan tingkat akurasi algoritma *naïve bayes* setelah memeriksa kinerjanya dengan menggunakan metode *bagging* dan *genetic algorithm*. Meskipun *naïve bayes* dikenal karena kesederhanaan dan kecepatan pemrosesannya, penelitian ini mengeksplorasi potensi peningkatan akurasi dengan menggabungkan teknik analisis tersebut. Penelitian ini melibatkan serangkaian eksperimen yang dirancang untuk menguji efektivitas *genetic algorithm* dan teknik *bagging* dalam meningkatkan performa *naïve bayes*. *Genetic algorithm*, dengan kemampuan optimasinya, digunakan untuk menemukan kombinasi parameter terbaik yang dapat meningkatkan akurasi *naïve bayes*. Di sisi lain, teknik *bagging* diterapkan untuk mengurangi varian dan meningkatkan stabilitas prediksi. Hasil penelitian menunjukkan bahwa penggunaan *genetic algorithm* berhasil meningkatkan akurasi *naïve bayes* dari 99.44% menjadi 99.90%, menunjukkan peningkatan sebesar 0.46%. Sementara itu, teknik *bagging* tidak memberikan peningkatan yang signifikan dalam akurasi. Implikasi dari temuan ini adalah bahwa teknik analisis seperti *genetic algorithm* dapat secara efektif meningkatkan performa algoritma klasifikasi, terutama dalam konteks mendeteksi penipuan transaksi kartu kredit. Temuan ini memberikan wawasan penting bagi industri keuangan dalam mengembangkan sistem keamanan yang lebih baik dan efisien untuk melindungi transaksi kartu kredit dari risiko penipuan.

**Kata kunci:** *Naïve bayes, Genetic algorithm, Bagging, Transaksi Kartu Kredit, Klasifikasi.*

# Performance Analysis of *Naïve bayes* Classification Algorithm Using *Genetic algorithm* and *Bagging* for Public Data on Credit Card Transaction Risk

## Abstract

Increased use of credit cards has increased the risk of fraud and crime related to credit card transactions. This requires the development of effective methods for detecting and classifying these risks accurately. The aim of this research is to compare the accuracy level of the *Naïve bayes* algorithm after checking its performance using *bagging* and *genetic algorithm* methods. Although *Naïve Bayes* is known for its simplicity and processing speed, this research explores the potential for increasing accuracy by combining these analysis techniques. This research involved a series of experiments designed to test the effectiveness of *genetic algorithms* and *bagging* techniques in improving the performance of *naïve Bayes*. *Genetic algorithms*, with their optimization capabilities, are used to find the best combination of parameters that can improve the accuracy of *Naïve Bayes*. On the other hand, *bagging* techniques are applied to reduce variance and increase prediction stability. The research results show that the use of a *genetic algorithm* succeeded in increasing the accuracy of *Naïve bayes* from 99.44% to 99.90%, showing an increase of 0.46%. Meanwhile, the *bagging* technique does not provide a significant increase in accuracy. The implication of these findings is that analysis techniques such as *genetic algorithms* can effectively improve the performance of classification algorithms, especially in the context of detecting fraudulent credit card transactions. These findings provide important insights for the financial industry in developing better and more efficient security systems to protect credit card transactions from the risk of fraud.

**Keywords:** *Naïve bayes, Genetic algorithm, Bagging, Credit Card Transactions, Classification.*

## I. PENDAHULUAN

Di era digital yang semakin berkembang, penggunaan kartu kredit sebagai metode pembayaran mengalami pertumbuhan yang signifikan. Namun, Seiring dengan meningkatnya penggunaan kartu kredit dan frekuensi transaksi, tindakan kriminal berupa kecurangan finansial, termasuk penipuan kartu kredit, juga meningkat [1]. Oleh karena itu, perlu adanya pengembangan metode yang efektif untuk secara akurat mendeteksi dan mengklasifikasikan risiko yang terkait dengan transaksi kartu kredit.

Penipuan kartu kredit merugikan konsumen dan industri keuangan dengan kerugian finansial langsung, biaya penanganan kasus penipuan, dan kerusakan reputasi perusahaan, sehingga pentingnya perlindungan data kartu kredit tidak bisa dianggap sepele karena penyalahgunaan data oleh pihak tidak berwenang dapat mengakibatkan dampak serius. Dengan meningkatnya kompleksitas dan volume data transaksi, teknik data mining menjadi salah satu solusi yang menjanjikan dalam meningkatkan keamanan dan efektivitas sistem pembayaran. Teknik klasifikasi dalam data mining dapat diterapkan untuk menilai risiko kredit [2].

Data mining adalah proses mengubah kumpulan informasi yang besar menjadi pengetahuan yang berarti [3]. Implementasi data mining sangat penting untuk menganalisis dan mengeksplorasi data yang ada, guna menemukan informasi baru yang lebih berharga dan bermanfaat bagi perusahaan [4]. Ada beberapa algoritma yang digunakan dalam klasifikasi *data mining*, dan salah satunya adalah *naive bayes*.

*Naive bayes* merupakan salah satu algoritma klasifikasi yang sangat terkenal. Algoritma *data mining* ini dikenal karena kemudahannya dan kemampuannya untuk diproses dengan cepat. Algoritma ini memiliki struktur yang simpel namun menunjukkan tingkat tinggi keberhasilan [5].

Kelebihan algoritma *naive bayes* termasuk kemampuan untuk disesuaikan dengan jumlah prediksi dan titik data yang berbeda, kemampuan untuk memprediksi nilai probabilitas (atau peluang), dan kemampuan untuk menangani baik data kontinu maupun diskrit [6].

Saat ini, penggunaan algoritma *naive bayes* dalam analisis transaksi kartu kredit telah menjadi praktik umum dalam industri keuangan. Namun, terus ada upaya untuk meningkatkan kinerja dan akurasi algoritma ini melalui penggunaan metode optimisasi seperti algoritma genetika dan teknik *ensemble* seperti *bagging*. Dengan demikian, penelitian ini berada dalam konteks yang relevan dan penting dalam pengembangan teknik analisis data untuk transaksi kartu kredit.

Algoritma di ranah pengetahuan dalam komputasi, algoritma genetika merupakan metode pencarian yang digunakan untuk menemukan penyelesaian optimal dan permasalahan pencarian. Algoritma genetika melibatkan tahapan *crossover*, evaluasi kecocokan (*fitness*), dan proses evolusi, populasi, serta kromosom [7]. Sebagai bagian dari penelitian ini, algoritma genetika akan digunakan untuk menemukan kombinasi parameter terbaik untuk meningkatkan kinerja *naive bayes*.

Selain itu, Salah satu metode pembelajaran kelompok yang menggunakan hanya satu tipe *base-model* adalah *Bagging*. Metode ini menggunakan pembelajaran secara paralel dan independen pada tiap *base-model*, dan kemudian digabungkan untuk memperoleh hasil paling optimal [8]. Sebagai bagian dari skripsi ini, *bagging* akan digunakan untuk menggabungkan beberapa model *naive bayes* yang dioptimalkan menggunakan algoritma genetika.

Melalui analisis ini diharapkan kinerja algoritma klasifikasi *naive bayes* dapat ditingkatkan secara signifikan. Hasil dari analisis ini dapat memberikan wawasan baru mengenai penggunaan dan analisis algoritma klasifikasi *naive bayes*, serta memberikan kontribusi terhadap pengembangan dalam teknik analisis data.

Dengan mengintegrasikan ketiga metode ini, diharapkan dapat meningkatkan akurasi dan efektivitas sistem deteksi penipuan kartu kredit. *Naive Bayes* memberikan dasar yang kuat untuk klasifikasi, algoritma genetika membantu dalam mengoptimalkan model, dan *Bagging* menggabungkan kekuatan beberapa model untuk hasil yang lebih stabil dan akurat. Kombinasi ini menawarkan pendekatan komprehensif dan canggih untuk mengatasi masalah deteksi penipuan dan klasifikasi risiko yang semakin kompleks di era digital saat ini.

Dari penjelasan sebelumnya, tujuan penelitian ini adalah untuk mengevaluasi tingkat akurasi algoritma *naive bayes* dalam mengklasifikasikan data transaksi kartu kredit setelah dilakukan analisis kinerja dengan metode *bagging* dan *genetic algorithm*, dengan memperhatikan keseimbangan kelas-kelas yang terlibat.

Penelitian ini dapat memberikan manfaat dalam pemahaman lebih mendalam tentang klasifikasi *naive bayes* dan cara mengoptimalkannya menggunakan teknik *genetic algorithm* dan *bagging*, dan juga memberikan manfaat kepada lembaga keuangan dan penyedia layanan pembayaran dengan mengurangi potensi risiko dalam transaksi kartu kredit dengan lebih baik.

## II. METODELOGI PENELITIAN

### 2.1 Data

Pada penelitian ini menggunakan data yang berasal dari sumber publik yang dapat diunduh dari <https://www.kaggle.com> dengan judul Credit Card Fraud Detection Dataset 2023. Dataset ini dapat diunduh langsung dari situs web <https://www.kaggle.com/datasets/nelgiryewithana/credit-card-fraud-detection-dataset-2023>. Total data yang terdapat dalam dataset ini mencapai 568.630 rekaman dengan 31 atribut. Dataset ini mencakup transaksi kartu kredit yang dilakukan oleh pemegang kartu Eropa selama tahun 2023. Informasi mengenai spesifikasi data tersedia dalam tabel 1.

Table 1. Dataset Credit Card Fraud

Data	Kelas	Atribut	Record
Credit Card	2	31	568.630

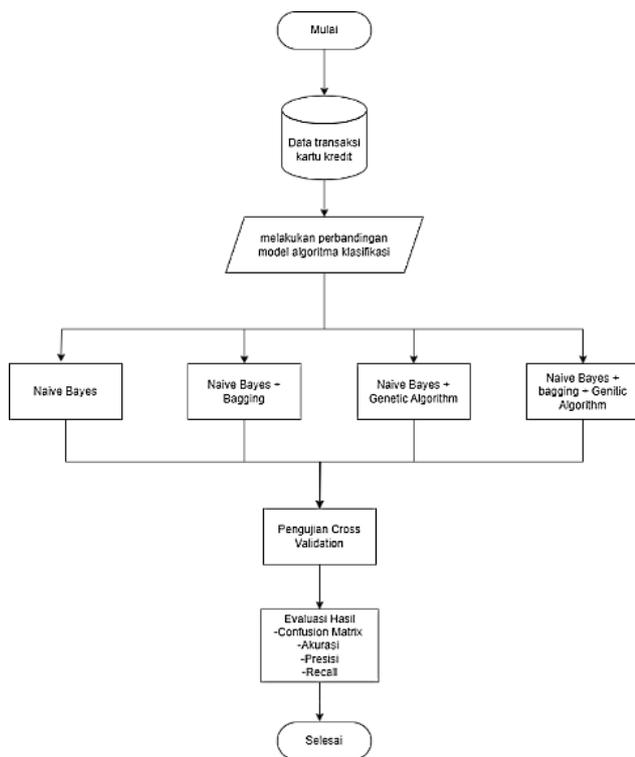
Atribut-atribut data kartu kredit, yang terdiri dari sejumlah variabel *input* dan *output*. Variabel *input* terdiri dari id (bertipe data integer), V1 hingga V28 (bertipe data *float64*), dan Amount (bertipe data *float64*). Setiap variabel

input ini menggambarkan berbagai atribut dari transaksi kartu kredit, seperti V1 hingga V28 yang mungkin mencerminkan berbagai fitur terkait transaksi tersebut. Misalnya, V1 dapat mengacu pada fitur-fitur seperti penggunaan kartu, V2 mungkin terkait dengan informasi tentang pembayaran, dan seterusnya. Selain itu, variabel amount menggambarkan jumlah transaksi. Variabel output, yang dinotasikan sebagai y atau class, adalah atribut yang ingin diprediksi atau diidentifikasi. Dalam konteks ini, class bertipe data integer dan mewakili kategori atau label untuk transaksi, seperti apakah itu transaksi yang sah (diwakili oleh nilai 0) atau transaksi yang curang (diwakili oleh nilai 1).

Alasan pemilihan dataset ini adalah karena dataset ini merupakan kumpulan data terbaru dan relevan yang mencerminkan tren terkini dalam transaksi kartu kredit dan penipuan. Selain itu, ukuran dataset yang besar memungkinkan pelatihan model yang lebih robust dan generalisasi yang lebih baik.

### 2.2 Skema Penelitian

Rancangan penelitian yang akan digunakan melibatkan komparasi model klasifikasi antara *naive bayes* dan *naive bayes* yang dianalisis menggunakan algoritma *bagging* dan *genetic algorithm* dalam mengklasifikasikan data transaksi kartu kredit. Ilustrasi berikut menggambarkan skema penelitian yang diterapkan.



Gambar 1. Skema Penelitian

Gambar 1. menggambarkan rencana penelitian yang digunakan untuk membandingkan model algoritma klasifikasi antara *naive bayes* dan *naive bayes* yang telah dianalisis menggunakan algoritma *bagging* dan *genetic algorithm* dalam mengklasifikasikan data transaksi kartu kredit. Pengujian dilakukan dalam empat tahap, yaitu:

pengujian klasifikasi melibatkan penerapan algoritma *naive bayes*, *naive bayes* yang dianalisis menggunakan *bagging*, *naive bayes* yang dianalisis menggunakan *genetic algorithm*, dan *naive bayes* dengan penggabungan analisis *bagging* dan *genetic algorithm*. Metode pengujian melibatkan validasi berulang, dan hasilnya dievaluasi menggunakan *confusion matrix* untuk mengukur akurasi, presisi, dan *recall*.

### 2.3 Algoritma Naïve Bayes

*Naive bayes classifier* adalah metode klasifikasi yang didasarkan pada Teorema Bayes [9]. *Naive bayes* menggunakan teorema bayes sebagai dasar untuk model machine learning berbasis probabilitas. Probabilitas posterior adalah vektor data kelas C yang diberikan. Ini diwakili dengan P(C|X). Persamaan 1 berikut menjelaskan algoritma *naive bayes* [10]:

$$P(C_i | X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{1}$$

Dalam Persamaan 1 dinyatakan bahwa nilai P(X) konstan untuk seluruh kelas, di mana nilai X melibatkan catatan pelatihan dengan atribut n X (X1, X2,..., Xn). Kelas data direpresentasikan oleh Ci, yang diwakili oleh C1, C2,..., Cn.

Keuntungan menggunakan pendekatan *naive bayes* adalah bahwa menghitung estimasi parameter yang diperlukan dalam proses pengklasifikasian hanya membutuhkan sedikit data pelatihan.

Selain itu, metode *naive bayes* dianggap sebagai salah satu dari sepuluh klasifikasi paling populer untuk *data mining*. Metode ini juga dianggap memiliki tingkat efisiensi komputasi dan akurasi yang lebih tinggi saat mengklasifikasikan dokumen [11].

Namun, kelemahan Naïve Bayes terletak pada asumsi independensi antar fitur, yang sering kali tidak sesuai dengan kondisi sebenarnya, sehingga dapat menurunkan akurasi dalam beberapa situasi. Meskipun asumsi independensi fitur jarang terpenuhi dalam data nyata, kinerja pengklasifikasian Naïve Bayes tetap relatif tinggi, seperti yang dibuktikan oleh berbagai penelitian empiris [12].

### 2.4 Bagging

*Bagging* merupakan singkatan dari pengumpulan dengan metode *bootstrap*, sebuah teknik *ensemble* yang diusulkan oleh Leo Breiman. Metode ini berguna untuk mengurangi variansi dari variabel independen dengan tujuan meningkatkan kualitas prediksi klasifikasi dari sebuah pohon klasifikasi tunggal. Ide pokok dari pendekatan *bagging* melibatkan penggunaan proses pengambilan sampel acak dengan pengembalian dari *dataset* asal untuk menghasilkan *dataset* yang baru [13].

Ada dua tahap utama dalam metode *bagging*, yang pertama adalah *bootstrap*, yang mengambil contoh dari data pembelajaran yang diperoleh, yang kedua adalah *aggregating*, yang menyatukan semua nilai estimasi menjadi satu nilai estimasi. Sebagai contoh, Langkah-langkah pembuatan estimasi dengan pohon *bagging* adalah sebagai berikut [14].



Akurasi adalah metrik evaluasi yang mengukur sejauh mana model membuat prediksi yang benar dari keseluruhan prediksi yang dilakukan. Presisi adalah metrik evaluasi yang mengukur sejauh mana model membuat prediksi yang benar untuk kelas positif dari seluruh prediksi positif yang dihasilkan. Adapun untuk memperoleh nilai untuk setiap kelas dapat dilihat pada persamaan 3 dan 4.

$$\text{Presisi (TP)} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{Presisi (TN)} = \frac{TN}{FN+TN} \tag{4}$$

Presisi total untuk semua kelas dapat dihitung menggunakan persamaan 5.

$$\text{Presisi} = \frac{\text{Presisi (TP)} + \text{Presisi (TN)}}{2} \tag{5}$$

Sedangkan *recall* adalah metrik evaluasi yang mengukur seberapa baik suatu model dalam mengenali kelas positif secara tepat. Adapun untuk memperoleh nilai untuk setiap kelas dapat dilihat pada persamaan 6 dan 7.

$$\text{Recall (TP)} = \frac{TP}{TP+FN} \tag{6}$$

$$\text{Recall (TN)} = \frac{TN}{FP+TN} \tag{7}$$

*Recall* total untuk semua kelas dapat dihitung menggunakan persamaan 8.

$$\text{Recall} = \frac{\text{Recall (TP)} + \text{Recall (TN)}}{2} \tag{8}$$

### III. HASIL DAN PEMBAHASAN

Dalam penelitian ini, sampel yang digunakan sebanyak 568.630 data transaksi kartu kredit dengan 31 label yaitu id, V1 sampai dengan V28, *amount*, dan *class*. Dataset yang digunakan dalam penelitian diambil melalui *platform Kaggle* yang berupa data tentang transaksi kartu kredit. Dalam penelitian ini melakukan pengujian nilai akurasi terhadap *naive bayes* dan *naive bayes* menggunakan *genetic algorithm* dan *bagging*.

#### 3.1 Hasil Penelitian

Pengujian sebuah model klasifikasi melibatkan penggunaan *cross validation* untuk membagi data training dan data testing. Dataset akan dibagi menjadi K partisi melalui *cross validation*, yang memiliki bagian untuk pelatihan dan pengujian data. Proses ini dijalankan dengan frekuensi sebanyak nilai K yang telah ditentukan, sehingga setiap partisi telah digunakan sebagai data uji. Dari pengujian klasifikasi kinerja algoritma *naive bayes* dengan algoritma *genetic algorithm* dan *bagging* terhadap data transaksi kartu kredit dapat di lihat pada tabel 4.

Table 3. Hasil Akurasi Pengujian

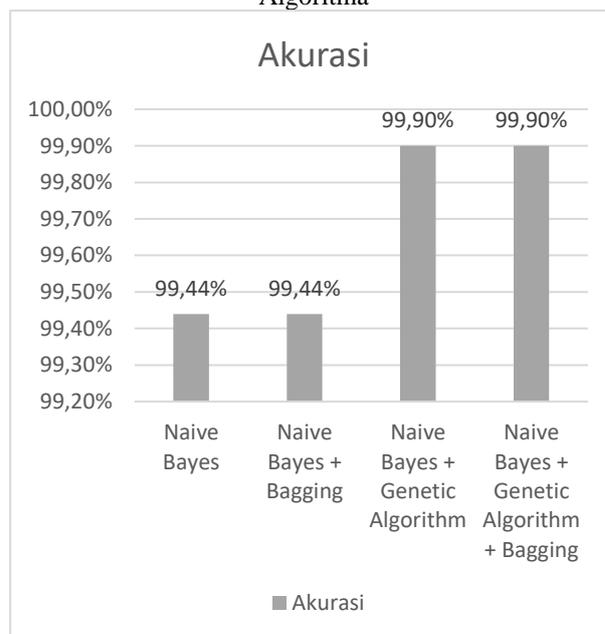
Algoritma	Akurasi
<i>Naive bayes</i>	99,44%

Algoritma	Akurasi
<i>Naive bayes + Bagging</i>	99,44%
<i>Naive bayes + Genetic algorithm</i>	99,90%
<i>Naive bayes + Bagging + Genetic algorithm</i>	99,90%

Berdasarkan Tabel 4, akurasi algoritma *naive bayes* mencapai 99,44%. Setelah menganalisis kinerja algoritma tersebut dengan metode *bagging*, keakuratannya memiliki nilai akurasi yang sama dengan *naive bayes* yaitu 99,44%. Selain itu, dengan menggunakan algoritma genetika untuk analisis kinerja, keakuratan meningkat drastis menjadi 99,90%. Namun, ketika kedua metode, yaitu *genetic algorithm* dan *bagging*, dikombinasikan, keakuratan juga mendapatkan hasil yang sama dengan analisis *naive bayes* yang menggunakan algoritma genetika yaitu 99,90%. Dari hasil pengujian ini, dapat disimpulkan bahwa menggunakan *genetic algorithm* dan kombinasi *genetic algorithm* dan *bagging* memberikan hasil terbaik untuk meningkatkan keakuratan algoritma *naive bayes*.

Hasil perbandingan tingkat ketepatan kinerja algoritma disajikan melalui grafik yang terdapat dalam gambar 3.

Gambar 3. Grafik Perbandingan Nilai Akurasi Kinerja Algoritma



Hasil penilaian kinerja dari setiap algoritma ditampilkan melalui *confusion matrix* yang tersedia dalam tabel 5.

Table 4. *Confusion Matrix* Kinerja *Naive bayes*

Akurasi = 99,44%	Prediction		Recall	
	positive	negative		
Actual	positive	284314	1	1,00
	negative	3175	281140	0,99
Precision		0,99	1,00	

Pada Tabel 5, dalam proses 10 kali iterasi pengujian pada *cross validation*, pada algoritma *naive bayes* dievaluasi menggunakan total 45211 data. Dari data tersebut, 284314 data kelas positif diprediksi dengan benar, sedangkan 1 data dari kelas positif salah diprediksi sebagai kelas negatif. Selain itu, 281140 data kelas negatif diprediksi dengan benar, namun 3175 data kelas negatif salah diprediksi sebagai kelas positif. Dari hasil tersebut, nilai presisi untuk kelas positif adalah 0,99%, dan untuk kelas negatif adalah 1,00%. Dengan demikian, rata-rata nilai presisinya adalah 0,99%. Sementara itu, nilai *recall* untuk kelas positif adalah 1,00%, dan untuk kelas negatif adalah 0,99%. Rata-rata nilai *recall* dari kedua kelas tersebut adalah 0,99%.

Berikutnya, Tabel 6 menampilkan *confusion matrix* yang menunjukkan kinerja algoritma *Naive bayes* yang telah dianalisis kinerja menggunakan algoritma *Bagging*.

Table 5. *Confusion Matrix* Kinerja *Naive bayes* dan *Bagging*

Akurasi = 99,44%	Prediction		<i>Recall</i>	
	positive	negative		
Actual	positive	284314	1	1,00
	negative	3175	281140	0,99
Precision		0,99	1,00	

Pada Tabel 6, dalam dalam proses 10 kali iterasi pengujian pada *cross validation*, pada algoritma *naive bayes* dengan *bagging* dievaluasi menggunakan total 45211 data. Dari data tersebut, 284314 data kelas positif diprediksi dengan benar, sedangkan 1 data dari kelas positif salah diprediksi sebagai kelas negatif. Selain itu, 281140 data kelas negatif diprediksi dengan benar, namun 3175 data kelas negatif salah diprediksi sebagai kelas positif. Dari hasil tersebut, nilai presisi untuk kelas positif adalah 0,99%, dan untuk kelas negatif adalah 1,00%. Dengan demikian, rata-rata nilai presisinya adalah 0,99%. Sementara itu, nilai *recall* untuk kelas positif adalah 1,00%, dan untuk kelas negatif adalah 0,99%. Rata-rata nilai *recall* dari kedua kelas tersebut adalah 0,99%.

Berikutnya, Tabel 7 menampilkan *confusion matrix* yang menunjukkan kinerja algoritma *Naive bayes* yang telah dianalisis kinerja menggunakan *Genetic algorithm*.

Table 6. *Confusion Matrix* Kinerja *Naive bayes* dengan *Genetic algorithm*

Akurasi = 99,90%	Prediction		<i>Recall</i>	
	positive	negative		
Actual	positive	284315	0	1,00
	negative	566	283749	1,00
Precision		1,00	1,00	

Pada Tabel 7, dalam dalam proses 10 kali iterasi pengujian pada *cross validation*, pada algoritma *Naive bayes* dengan *Genetic algorithm* dievaluasi menggunakan total 45211 data. Dari data tersebut, 284315 data kelas positif diprediksi dengan benar, sedangkan 0 data dari kelas positif salah diprediksi sebagai kelas negatif. Selain itu, 283749 data kelas negatif diprediksi dengan benar, namun 566 data kelas negatif salah diprediksi sebagai kelas positif. Dari hasil tersebut, nilai presisi untuk kelas positif adalah 1,00%, dan untuk kelas negatif adalah 1,00%. Dengan demikian, rata-rata nilai presisinya adalah 1,00%. Sementara itu, nilai *recall* untuk kelas positif adalah 1,00%, dan untuk kelas negatif adalah 1,00%. Rata-rata nilai *recall* dari kedua kelas tersebut adalah 1,00%.

Berikutnya, Tabel 8 menampilkan *confusion matrix* yang menunjukkan kinerja algoritma *Naive bayes* yang telah dianalisis kinerja menggunakan teknik kombinasi dengan *Genetic algorithm* dan *Bagging*.

Table 7. *Confusion Matrix* Kinerja *Naive bayes* dengan *Genetic algorithm* dan *Bagging*

Akurasi = 99,90%	Prediction		<i>Recall</i>	
	positive	negative		
Actual	positive	284315	0	1,00
	negative	566	283749	1,00
Precision		1,00	1,00	

Pada Tabel 8, dalam dalam proses 10 kali iterasi pengujian pada *cross validation*, pada algoritma *naive bayes* kombinasi *genetic algorithm* dan *bagging* dievaluasi menggunakan total 45211 data. Dari data tersebut, 284315 data kelas positif diprediksi dengan benar, sedangkan 0 data dari kelas positif salah diprediksi sebagai kelas negatif. Selain itu, 283749 data kelas negatif diprediksi dengan benar, namun 566 data kelas negatif salah diprediksi sebagai kelas positif. Dari hasil tersebut, nilai presisi untuk kelas positif adalah 1,00%, dan untuk kelas negatif adalah 1,00%. Dengan demikian, rata-rata nilai presisinya adalah 1,00%. Sementara itu, nilai *recall* untuk kelas positif adalah 1,00%, dan untuk kelas negatif adalah 1,00%. Rata-rata nilai *recall* dari kedua kelas tersebut adalah 1,00%.

### 3.2 Analisa Hasil Pengujian

Berdasarkan hasil pengujian, peningkatan akurasi kinerja algoritma klasifikasi *naive bayes* tercatat sebesar 0,46%, yang didapatkan oleh penambahan *genetic algorithm* pada *naive bayes*. Namun, penggunaan *bagging* tidak menghasilkan peningkatan lebih lanjut, sehingga akurasi yang diperoleh sama dengan *naive bayes* tanpa *bagging*. Evaluasi menggunakan *confusion matrix* menunjukkan bahwa presisi dan *recall* dari *naive bayes* dengan *bagging* tidak berbeda dengan *naive bayes* tanpa *bagging*. Hal ini disebabkan oleh rendahnya variansi pada *naive bayes*, di mana tambahan *bagging* mungkin tidak memberikan keuntungan yang signifikan.

*Naive bayes*, sebagai algoritma sederhana dan efektif, bekerja optimal pada data yang seimbang, sehingga risiko overfitting yang biasanya diatasi oleh *bagging* menjadi kurang relevan. Selain itu, kestabilan *naive bayes* dan ketiadaan variansi yang tinggi mengurangi manfaat penggunaan *bagging*, yang umumnya digunakan untuk mengatasi overfitting pada algoritma dengan variansi tinggi. Evaluasi terhadap kombinasi *naive bayes* dengan *genetic algorithm*, serta kombinasi *naive bayes* dengan *genetic algorithm* dan *bagging*, menunjukkan hasil yang sama dalam hal akurasi, presisi, dan recall.

Model klasifikasi yang digunakan dalam penelitian ini menunjukkan bahwa penggunaan *genetic algorithm*, baik secara mandiri maupun dalam kombinasi dengan *naive bayes* dan *bagging*, tetap menghasilkan nilai akurasi, presisi, dan recall yang sama. Oleh karena itu, penambahan *bagging* pada *naive bayes*, bahkan ketika digabungkan dengan *genetic algorithm*, tidak memberikan kontribusi yang berarti. Secara keseluruhan, meskipun *genetic algorithm* menunjukkan potensi dalam mengoptimalkan kinerja *naive bayes*, integrasi *bagging* tidak membawa keuntungan tambahan. Hasil penelitian ini menekankan perlunya eksplorasi lebih lanjut terhadap metode optimasi dan teknik ensemble lainnya yang mungkin lebih efektif dalam meningkatkan kinerja model klasifikasi. Kombinasi alternatif atau pendekatan yang berbeda mungkin diperlukan untuk mencapai peningkatan yang signifikan dalam akurasi, presisi, dan recall pada klasifikasi data transaksi kartu kredit.

### 3.3 Implikasi Hasil

Penerapan *genetic algorithm* pada model *naive bayes* menunjukkan peningkatan akurasi yang signifikan (99,90% dibandingkan 99,44% pada *naive bayes* saja), menunjukkan bahwa *genetic algorithm* efektif dalam mengoptimalkan parameter model atau fitur yang lebih baik. Namun, penambahan *bagging* pada *naive bayes* tidak memberikan peningkatan akurasi yang berarti, yang dapat diatribusikan pada keseimbangan kelas dalam dataset, di mana *naive bayes* sudah cukup efektif tanpa adanya risiko overfitting yang tinggi. *Bagging* sering digunakan untuk model dengan variansi tinggi, sedangkan *naive bayes* memiliki variansi yang relatif rendah. Meskipun kombinasi *genetic algorithm* dan *bagging* juga menghasilkan akurasi 99,90%, tidak ada keuntungan tambahan yang diperoleh dibandingkan hanya menggunakan *genetic algorithm*, mengindikasikan bahwa manfaat *bagging* tidak dirasakan pada model dengan variansi rendah seperti *naive bayes* dalam dataset penelitian ini.

### 3.4 Perbandingan Metode

- Naive bayes* dan *genetic algorithm*: *genetic algorithm* memberikan peningkatan akurasi yang jelas dibandingkan *naive bayes* saja.
- Naive bayes* dan *bagging*: *bagging* tidak memberikan peningkatan akurasi pada *naive bayes* yang sudah memiliki performa tinggi.
- Naive bayes* + *genetic algorithm* dan *naive bayes* + *genetic algorithm* + *bagging*: kombinasi dengan

*bagging* tidak meningkatkan hasil dibandingkan hanya menggunakan *genetic algorithm*.

## IV. KESIMPULAN

Berdasarkan hasil penelitian, penggunaan *bagging* untuk menganalisis kinerja *naive bayes* dalam klasifikasi data transaksi kartu kredit menghasilkan akurasi yang sama dengan penggunaan *naive bayes* saja. Hal ini disebabkan oleh variansi rendah dari model *naive bayes*, yang mengurangi manfaat dari penerapan *bagging*. Dengan kata lain, dalam kasus ini, algoritma *bagging* tidak memberikan peningkatan akurasi yang signifikan.

Penelitian menemukan bahwa performa terbaik dalam mengoptimalkan *naive bayes* untuk mendeteksi penipuan kartu kredit dicapai dengan menggunakan *genetic algorithm*, dengan akurasi mencapai 99,90%. Akurasi tersebut meningkat sebanyak 0,46% dibandingkan dengan penggunaan algoritma *naive bayes* saja. Namun, ketika *bagging* digabungkan dengan *genetic algorithm*, tidak ada kontribusi signifikan terhadap peningkatan akurasi.

Evaluasi menggunakan confusion matrix menunjukkan bahwa meskipun *genetic algorithm* meningkatkan akurasi, presisi dan recall dari *naive bayes* dengan atau tanpa *bagging* tetap tinggi dan serupa. *Bagging* yang umumnya digunakan untuk mengurangi variansi dan meningkatkan performa model klasifikasi, tidak memberikan efek signifikan dalam kasus ini. Hal ini menunjukkan bahwa rendahnya variansi pada model *naive bayes* dalam dataset mengurangi manfaat tambahan dari *bagging*.

Meskipun *bagging* biasanya dapat meningkatkan kinerja model klasifikasi dengan mengurangi variansi, efeknya tidak signifikan dalam konteks penelitian ini. Oleh karena itu, diperlukan penelitian lebih lanjut untuk menemukan kombinasi metode yang lebih efektif. Penelitian mendatang diharapkan dapat mengeksplorasi strategi optimasi tambahan, seperti penggunaan Support Vector Machine (SVM), yang mampu menangani keseimbangan kelas dengan lebih efektif, untuk meningkatkan performa *naive bayes* dalam klasifikasi data transaksi kartu kredit.

## DAFTAR PUSTAKA

- [1] T. S. Lestari dan D. A. N. Sirodj, "Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest," *J. Ris. Stat.*, vol. 1, no. 2, hal. 160–167, 2022, doi: 10.29313/jrs.v1i2.525.
- [2] I. Sugiyarto, "Perbandingan Kinerja Algoritma Data Mining Prediksi Persetujuan Kartu Kredit," *Fakt. Exacta*, vol. 12, no. 3, hal. 180, 2019, doi: 10.30998/faktorexacta.v12i3.4310.
- [3] F. Stevanus Alvian, S. Carbini, S. Likmi Bandung, dan S. Mardira Indonesia, "PREDIKSI Kelayakan Pemberian Fasilitas Kartu Kredit Kepada Nasabah Dengan Metode Klasifikasi Data Mining (Studi Kasus : Bank XYZ)," *J. Comput. Bisnis*, vol. 14, no. 2, hal. 123–128, 2020, [Daring]. Tersedia pada: www.forbes.com
- [4] D. A. Manalu dan G. Gunadi, "Implementasi Metode Data Mining K-Means Clustering

- Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi,” *Infotech J. Technol. Inf.*, vol. 8, no. 1, hal. 43–54, 2022, doi: 10.37365/jti.v8i1.131.
- [5] E. Fitri, Y. Yuliani, S. Rosyida, dan W. Gata, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naive bayes* , Random Forest Dan Support Vector Machine,” vol. 18, no. 1, hal. 71–80, 2020.
- [6] L. B. Adzy, A. Pambudi, U. M. Sukabumi, P. Bantuan, I. Jaminan, dan S. K. Sukabumi, “Algoritma *Naive Bayes* Untuk Klasifikasi Kelayakan Penerima,” vol. 6, no. 1, hal. 1–10, 2023.
- [7] S. F. Pane, R. Maulana Awangga, E. V. Rahmadani, dan S. Permana, “Implementasi Algoritma Genetika Untuk Optimalisasi Pelayanan Kependudukan,” *J. Tekno Insentif*, vol. 13, no. 2, hal. 36–43, 2019, doi: 10.36787/jti.v13i2.130.
- [8] L. M. Cendani dan A. Wibowo, “Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes,” *J. Masy. Inform.*, vol. 13, no. 1, hal. 33–44, 2022, doi: 10.14710/jmasif.13.1.42912.
- [9] N. Nurdin, M. Suhendri, Y. Afrilia, dan R. Rizal, “Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode *Naive bayes* Classifier (NBC),” *Sistemasi*, vol. 10, no. 2, hal. 268, 2021, doi: 10.32520/stmsi.v10i2.1193.
- [10] A. Nugroho dan Y. Religia, “Analisis Optimasi Algoritma Klasifikasi *Naive bayes* menggunakan *Genetic algorithm* dan *Bagging*,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, hal. 504–510, Jun 2021, doi: 10.29207/resti.v5i3.3067.
- [11] U. Fitriani Dan A. Wibowo, “Penerapan Algoritme *Naive Bayes* Untuk Memprediksi Application Of *Naive Bayes* Algorithm To Predict Graduation Of Budi Luhur University Students Based On,” Vol. 2, no. September, hal. 745–753, 2023.
- [12] Syarli dan A. A. Muin, “Metode *Naive bayes* Untuk Prediksi Kelulusan,” *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, hal. 22–26, 2016, [Daring]. Tersedia pada: <https://media.neliti.com/media/publications/283828-metode-naive-bayes-untuk-prediksi-kelulu-139fcfea.pdf>
- [13] B. Dan, R. Forest, dan P. Klasifikasi, “Analisis perbandingan kinerja cart konvensional, *bagging* dan random forest pada klasifikasi objek: hasil dari dua simulasi,” vol. 12, no. 2, hal. 1–12, 2019, doi: 10.14710/medstat.12.1.1-12.
- [14] D. A. Wulandari dan D. Kusnandar, “*Bagging* Classification Trees Untuk Klasifikasi Tingkat Kesejahteraan Rumah Tangga Di Kalimantan Barat,” vol. 08, no. 4, hal. 765–772, 2019.
- [15] S. Saleh, N. Umar, dan M. A. Nur, “Implementasi Algoritma Genetika Untuk Penjadwalan Ujian Pada Universitas Handayani Makassar,” hal. 154–156.
- [16] D. Ariadi, “Aplikasi Algoritma Genetika Dalam Mengoptimasi Tuned Mass Damper Untuk Mereduksi Getaran Pada Gedung Akibat Beban Gempa,” *J. Kacapuri J. Keilmuan Tek. Sipil*, vol. 4, no. 1, hal. 19, 2021, doi: 10.31602/jk.v4i1.5125.
- [17] R. Fitria, D. Yulisda, dan M. Ula, “Data Mining Classification Algorithms for Diabetes Dataset Using Weka Tool,” *Sisfo J. Ilm. Sist. Inf.*, vol. 5, no. 2, hal. 117–124, 2021, doi: 10.29103/sisfo.v5i2.6236.