Jurnal Sosial dan Teknologi (SOSTECH)
p-ISSN 2774-5147 | e-ISSN 2774-5155

Improvement Naive Bayes Menggunakan Forward Selection, Information Gain dan Gain Ratio untuk Penanganan Independensi Fitur

Suprapto
Universitas PGRI Ronggolawe Tuban, Indonesia
Email: suprapto.tif.unirow@gmail.com*

Article Info:
Submitted:
10-04-2025
Final Revised:
16-04-2025
Accepted:
19-04-2025
Published:
22-04-2025

Abstrak

Penelitian ini bertujuan untuk menganalisis peningkatan kinerja algoritma *Naive Bayes* (NB) dalam menangani independensi fitur menggunakan metode Forward Selection, Information Gain, dan Gain Ratio. Naive Bayes merupakan algoritma klasifikasi yang sering digunakan karena efisiensi komputasinya yang tinggi, namun sering mengalami penurunan performa ketika ada ketergantungan antar fitur. Penelitian ini menggunakan pendekatan eksperimental dengan menerapkan beberapa algoritma, yakni Naive Bayes, Forward Selection Naive Bayes (FSNB), Forward Selection Information Gain Naive Bayes (FSIGNB), dan Forward Selection Information Gain Ratio Naive Bayes (FSIGRNB) pada dataset berdimensi tinggi. Metode validasi yang digunakan adalah 10-fold cross validation untuk mengukur akurasi setiap algoritma. Hasil penelitian menunjukkan bahwa algoritma FSNB dan FSIGNB berhasil meningkatkan akurasi secara signifikan dibandingkan dengan algoritma NB standar. FSNB memiliki akurasi rata-rata tertinggi sebesar 81,124%, diikuti oleh FSIGNB dan FSIGRNB. Implikasi dari penelitian ini adalah bahwa penerapan metode Forward Selection, Information Gain, dan Gain Ratio dapat meningkatkan akurasi klasifikasi Naive Bayes, terutama dalam dataset dengan dimensi fitur yang tinggi, serta memberikan kontribusi penting dalam pengembangan algoritma untuk menangani independensi fitur.

Kata kunci: naive bayes, independensi fitur, seleksi fitur, pembobotan fitur, forward selection, information gain, gain ratio naive bayes.

Abstract

This study aims to analyze the improvement of Naive Bayes (NB) algorithm performance in handling feature independence using Forward Selection, Information Gain, and Gain Ratio methods. Naive Bayes is a widely used classification algorithm due to its high computational efficiency, but it often experiences performance degradation when there is dependency between features. This research adopts an experimental approach by applying several algorithms, namely Naive Bayes, Forward Selection Naive Bayes (FSNB), Forward Selection Information Gain Naive Bayes (FSIGNB), and Forward Selection Information Gain Ratio Naive Bayes (FSIGRNB), on high-dimensional datasets. The validation method used is 10-fold cross-validation to measure the accuracy of each algorithm. The results show that FSNB and FSIGNB algorithms significantly improve accuracy compared to the standard NB algorithm. FSNB achieved the highest average accuracy of 81.124%, followed by

FSIGNB and FSIGRNB. This study implies that applying Forward Selection, Information Gain, and Gain Ratio methods can enhance Naive Bayes classification accuracy, especially in datasets with high feature dimensions, and contribute significantly to developing algorithms for handling feature independence.

Keywords: *Naive Bayes, feature independence, feature selection, feature weighting, Forward Selection, Information Gain, Gain Ratio Naive Bayes*

Corresponding: Suprapto
E-mail: suprapto.tif.unirow@gmail.com

PENDAHULUAN

Naïve Bayes adalah sebuah algoritma yang berasal dari teori probabilitas milik Thomas Bayes tahun 1950, yang digunakan dalam statistik dan pembelajaran mesin (*machine learning*). Algoritma ini bekerja dengan cara menghitung peluang (probabilitas) dari suatu kejadian berdasarkan kejadian sebelumnya, menggunakan prinsip *teorema Bayes* (Berrar, 2018; Rizki et al., 2021). *Naïve Bayes* merupakan algoritma klasifikasi yang menggunakan pendekatan statistik dalam menghitung probabilitas. Algoritma ini termasuk salah satu metode utama dalam data mining dan sering digunakan dalam berbagai permasalahan klasifikasi karena kinerjanya yang efektif. *Naïve Bayes* mampu memprediksi kelas suatu data dengan memperhitungkan probabilitas dari fitur-fitur yang ada, lalu memilih kelas yang paling sesuai atau optimal berdasarkan perhitungan tersebut (Annur, 2018; Damar Rani & Zuhri, 2020; Harungguan et al., 2023; Ramadhan et al., 2023).

Algoritma ini sangat cocok digunakan dalam berbagai aplikasi karena kesederhanaannya dalam proses pembangunan dan penerapan. *Naïve Bayes* tidak memerlukan proses pelatihan yang kompleks seperti penyesuaian parameter secara berulang, yang biasanya memakan waktu dan tenaga komputasi tinggi. Kemudahan ini membantu algoritma bekerja efisien pada dataset besar tanpa menurunkan performa. Selain itu, karena prinsip kerjanya berdasarkan logika probabilitas yang cukup mudah dijelaskan, hasil dari algoritma ini bisa dipahami bahkan oleh pengguna yang belum menguasai teknologi klasifikasi secara mendalam. Inilah yang menjadikannya ramah bagi pemula namun tetap kuat dalam performa (Mahmood, 2018). Salah satu alasan algoritma Naïve Bayes banyak digunakan adalah karena proses latihannya sangat cepat dan sederhana. Hal ini membuatnya hemat waktu dan sumber daya, terutama pada proyek dengan data besar. Meskipun begitu, algoritma ini bekerja dengan dasar probabilitas yang membutuhkan pengetahuan awal, karena tanpa informasi tersebut, proses pengambilan keputusan oleh algoritma bisa menjadi kurang akurat atau bias. Oleh karena itu, meskipun efisien, akurasi algoritma ini tetap bergantung pada ketersediaan dan kualitas data pelatihan yang mencerminkan pengetahuan awal tersebut (Asaad & Abdulhakim, 2021; Guohua & Francis, 2017; Mostafa & Mahmoud, 2022; Nazneentarannum & Rizvi, 2016; Prasdika & Sugiantoro, 2018). Teorema Bayes menjadi dasar dari algoritma *Naïve Bayes*, yang secara umum mampu bersaing dengan metode klasifikasi lain seperti Decision Tree dan Neural Network, terutama dalam konteks performa. Hasil penelitian menunjukkan bahwa *Naïve Bayes* mampu memberikan hasil klasifikasi yang akurat dalam waktu yang lebih singkat, terutama saat mengolah data dalam jumlah besar. Keunggulan inilah yang membuat *Naïve Bayes* sering dipilih ketika dibutuhkan solusi yang cepat dan efisien dalam analisis data berskala besar.

Algoritma *Naïve Bayes* didasarkan pada prinsip bahwa setiap fitur dalam data dianggap tidak saling memengaruhi atau disebut sebagai asumsi independensi.Tujuan dari pendekatan ini adalah untuk menyederhanakan perhitungan probabilitas, sehingga proses klasifikasi menjadi

jauh lebih efisien secara komputasi. Meskipun dalam praktiknya fitur-fitur tersebut mungkin saling terkait, asumsi ini tetap digunakan karena terbukti cukup efektif di banyak kasus dan menghasilkan akurasi yang kompetitif. Algoritma ini rentan terhadap dependensi antar fitur. Ketidaktepatan asumsi independensi pada *naïve bayes* seringkali menjadi masalah, yang mengakibatkan penurunan kinerja klasifikasi dan efektivitas hasil (Attamami et al., 2023; Depari et al., 2022; Ericha Apriliyani & Salim, 2022; Istighfar et al., 2023).

Kinerja *naïve bayes* dapat ditingkatkan melalui berbagai cara, seperti memodifikasi fitur, data, atau struktur algoritma. Dalam manipulasi fitur, teknik pembobotan dan seleksi terbukti efektif. Studi menunjukkan bahwa metode-metode ini berhasil mengurangi dampak dari asumsi independensi fitur yang seringkali tidak akurat pada *naïve bayes*.

Berdasarkan pandangan Lee dan Wehicle, fleksibilitas pembobotan dan seleksi fitur menjadikannya pilihan utama dibandingkan pendekatan lain. Sesuai dengan hal ini, penelitian ini mengkhususkan pada kedua teknik tersebut. Untuk seleksi fitur, digunakan metode *Wrapper* yang mencakup tiga strategi: *Forward Selection* (pemilihan maju), *Backward Elimination* (eliminasi mundur), dan *Recursive Feature Elimination* (eliminasi fitur rekursif) (Shafiee et al., 2021). Dalam pembobotan fitur, setiap fitur dievaluasi dan diberi nilai yang berbeda berdasarkan tingkat kepentingannya. Beberapa fitur dianggap lebih krusial daripada yang lain. Tujuan utama dari pemberian bobot yang bervariasi ini adalah untuk menghasilkan sekumpulan fitur dengan hierarki kepentingan yang jelas, sambil tetap mempertahankan semua fitur, termasuk yang mungkin kurang relevan. Pendekatan *Correlation Based Feature Selection* (CFS) dan *Forward Selection* merupakan cara yang paling sederhana untuk mengimplementasikan pembobotan fitur sebagai bagian dari manipulasi fitur (Zhang et al., 2016). Sebagai teknik manipulasi fitur berbasis korelasi, *correlation based feature selection* (CFS) bekerja dengan cara mengidentifikasi dan memilih subset fitur yang dianggap paling baik dari keseluruhan fitur. Selanjutnya, fitur-fitur yang terpilih ini akan diberikan bobot yang lebih besar.

Sementara itu, *Forward Selection* merupakan teknik seleksi fitur yang baik untuk menangani isu independensi fitur. Metode ini bekerja secara iteratif, dimulai dengan model kosong (tanpa fitur). Di setiap langkah, fitur yang memberikan peningkatan performa model paling besar akan dipilih dan ditambahkan (Yang et al., 2020).

Pembobotan fitur dianggap sebagai cara yang potensial untuk mengatasi masalah asumsi independensi fitur. Penelitian ini mengusulkan penggunaan metode *Forward Selection, Information Gain,* dan *Gain Ratio*, yang dinilai efektif untuk dataset dengan dimensi yang besar. Studi ini menggunakan lima dataset dari repositori UCI, menerapkan validasi silang 10-*fold* sebagai standar evaluasi, dan menggunakan akurasi sebagai metrik utama untuk mengukur kinerja.

Mengacu pada latar belakang dan identifikasi masalah, penelitian ini mengajukan rumusan masalah: "Sejauh mana peningkatan algoritma *naïve bayes* melalui penggunaan *Forward Selection, Information Gain*, dan *Gain Ratio* dalam menangani asumsi independensi fitur akan mempengaruhi akurasi algoritma *Naïve Bayes*?" Tujuan penelitian ini adalah untuk meningkatkan akurasi algoritma *Naïve Bayes* dengan memanfaatkan algoritma *Feature Selection, Information Gain,* dan *Gain Ratio* untuk mengatasi keterbatasan asumsi independensi fitur.

Algoritma *Naive Bayes* (NB) sering digunakan dalam klasifikasi data karena efisiensinya yang tinggi, namun menghadapi masalah penurunan kinerja ketika ada ketergantungan antar fitur. Masalah ini terutama terjadi pada dataset dengan dimensi tinggi, di mana asumsi independensi fitur yang mendasari algoritma *Naive Bayes* tidak selalu valid. Oleh sebab itu, upaya pengembangan metode yang bertujuan untuk meningkatkan kinerja *Naive Bayes* dalam menangani relasi antar fitur menjadi suatu keharusan. Untuk mengatasi masalah tersebut, salah satu pendekatan yang dapat ditempuh adalah dengan menggunakan metode seleksi fitur yang dirancang untuk mengurangi dependensi antar fitur sehingga akurasi klasifikasi dapat meningkat.

Penelitian ini penting karena mengidentifikasi dan mengatasi masalah independensi fitur yang seringkali mengurangi efektivitas algoritma *Naive Bayes*, terutama pada dataset dengan dimensi tinggi. Penerapan metode seleksi fitur seperti *Forward Selection, Information Gain,* dan *Gain Ratio* diharapkan dapat meningkatkan akurasi algoritma *Naive Bayes* dan membuatnya lebih

efektif dalam menangani ketergantungan antar fitur. Dengan meningkatkan kinerja algoritma ini, penelitian ini dapat memberikan kontribusi dalam pengembangan algoritma klasifikasi yang lebih efisien dan akurat, terutama dalam aplikasi-aplikasi yang membutuhkan pengolahan data besar.

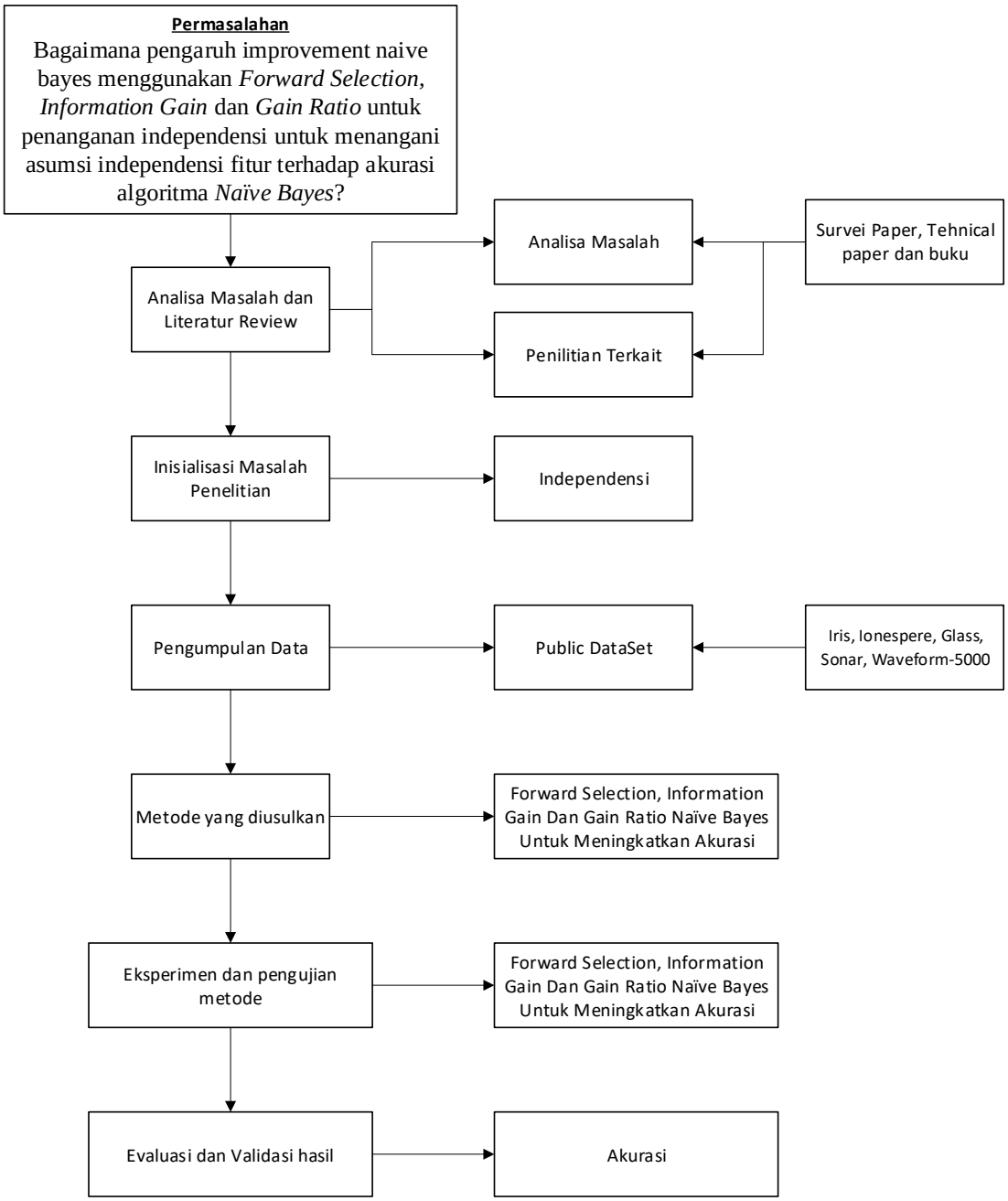
Penelitian oleh Zhang dan Sheng (2004) mengungkapkan bahwa metode pembobotan fitur dapat meningkatkan kinerja *Naive Bayes*, namun pada beberapa kasus, teknik ini masih terbatas pada pengolahan data dengan jumlah fitur yang kecil. Penelitian lain oleh Lee dan Wehicle (2010) menunjukkan bahwa *Forward Selection* efektif dalam mengatasi masalah independensi fitur dengan memilih subset fitur yang relevan, yang berkontribusi terhadap peningkatan akurasi klasifikasi. Selain itu, Jiang et al. (2016) menyatakan bahwa penggunaan Information Gain dalam seleksi fitur dapat meningkatkan akurasi *Naive Bayes* dengan memberikan bobot lebih pada fitur yang lebih signifikan, namun belum banyak penelitian yang mengkombinasikan metode ini dengan Gain Ratio dalam konteks dataset berdimensi tinggi.

Berbagai metode peningkatan *Naive Bayes* telah diteliti sebelumnya, namun belum ada kajian spesifik mengenai penerapan gabungan *Forward Selection*, *Information Gain*, dan *Gain Ratio* untuk mengatasi asumsi independensi fitur pada dataset dengan dimensi tinggi. Penelitian ini bertujuan untuk mengisi celah riset ini dan menambahkan kebaharuan dengan melakukan pengujian dan perbandingan kinerja beberapa metode seleksi fitur dalam upaya meningkatkan akurasi *Naive Bayes*. Tujuan utamanya adalah untuk meningkatkan kemampuan *Naive Bayes* dalam mengklasifikasi data dengan dimensi fitur yang tinggi.

Kemudian, penelitian ini diharapkan dapat berkontribusi signifikan pada kemajuan algoritma klasifikasi yang memiliki akurasi dan efisiensi yang lebih baik. Penelitian ini diharapkan dapat memberikan manfaat bagi pengembangan algoritma klasifikasi yang lebih efektif, khususnya dalam konteks pengolahan data dengan dimensi fitur tinggi. Temuan dari penelitian ini juga dapat memberikan implikasi penting untuk memperbaiki akurasi algoritma *Naive Bayes*, sehingga lebih optimal untuk diterapkan dalam berbagai aplikasi pengolahan data besar.

METODE PENELITIAN

Untuk meneliti hubungan sebab-akibat antar variabel, penelitian ini menggunakan metode eksperimen, di mana peneliti secara langsung melakukan intervensi dan observasi. Metode eksperimen yang digunakan dalam penelitian ini berarti peneliti secara aktif membuat skenario atau percobaan tertentu guna melihat bagaimana suatu variabel memengaruhi variabel lainnya. Dalam pelaksanaannya, peneliti menguji langsung model yang dikembangkan menggunakan komputer. Perhitungan dan analisis dilakukan dengan bantuan perangkat lunak RapidMiner versi 9.6, yang berguna untuk menjalankan model dan mengevaluasi seberapa akurat hasil prediksinya dibandingkan dengan data sebenarnya. *Naïve Bayes* bekerja dengan asumsi bahwa setiap fitur dalam dataset bersifat saling bebas atau independen satu sama lain. Untuk menguji seberapa baik algoritma ini bekerja berdasarkan asumsi tersebut, peneliti membutuhkan dataset yang umum dan telah banyak digunakan sebagai standar pembanding. UCI Repository merupakan sumber terkenal yang menyediakan berbagai dataset yang telah dikaji dalam banyak penelitian, termasuk empat dataset yang relevan dengan pengujian *Naïve Bayes*. Di samping itu, Datahub Enterprise juga menyediakan satu dataset yang sering digunakan dalam konteks serupa. Penggunaan kelima dataset ini bertujuan untuk memastikan bahwa metode yang diusulkan oleh peneliti dapat dinilai secara adil dan konsisten terhadap metode lain dalam literatur. Informasi mengenai dataset yang digunakan dalam penelitian ini, meliputi nama, jumlah data, jumlah fitur, status *missing value* (ada/tidak), dan tipe data numerik, dapat dilihat pada Tabel 3.2. Dataset-dataset ini diperoleh dari UCI *machine learning repository*. Dalam proses persiapan data, *missing value* yang ditemukan diisi menggunakan *unsupervised attribute filter replace missing value* melalui *software RapidMiner*. Selanjutnya, penelitian ini melibatkan serangkaian eksperimen yang mengikuti tahapan-tahapan yang telah dirancang.



HASIL DAN PEMBAHASAN

Hasil Eksperimen Algoritma *Naive Bayes*

Penelitian ini mengimplementasikan perhitungan algoritma *Naïve Bayes* secara manual pada dataset Sonar yang bersumber dari UCI Repository. Langkah ini dilakukan untuk memfasilitasi pemahaman mendetail mengenai setiap tahapan perhitungan algoritma *Naïve Bayes*. Dataset Sonar memiliki ciri-ciri sebagai berikut: multivariat, tipe atribut real, jumlah atribut 61, jumlah baris 208, jumlah kelas 2 (Mines, Rock), dan tidak terdapat *missing value*. Metode validasi yang digunakan adalah 10-*fold cross validation*, yang membagi dataset menjadi 10 bagian, di mana 9 bagian digunakan untuk pelatihan dan 1 bagian untuk pengujian. Contoh dataset Sonar yang akan digunakan dalam eksperimen ini dapat dilihat pada Tabel 1.

Tabel 1. Contoh Dataset Sonar

X56	X57	X58	X59	X60	X61
0,0029	0,0013	0,001	0,0032	0,0047	**M**
0,0085	0,0101	0,0016	0,0028	0,0014	**R**
0,0004	0,0018	0,0049	0,0024	0,0016	**R**
0,0093	0,0042	0,0003	0,0053	0,0036	**R**
0,0068	0,006	0,0045	0,0002	0,0029	**M**
0,0062	0,0024	0,0063	0,0017	0,0028	**M**
0,0084	0,0037	0,0024	0,0034	0,0007	**M**
0,0036	0,0026	0,0036	0,0006	0,0035	**R**
0,0019	0,0034	0,0034	0,0051	0,0031	**R**
0,0068	0,0041	0,0052	0,0194	0,0105	**R**

Dataset Sonar yang sudah difilter untuk kelas *Mines (M)* berjumlah 4 untuk kelas *Mine (M).* dataset Sonar yang sudah difilter untuk kelas *Rock (R)* berjumlah 6 untuk kelas *Rock (R).*

Hasil Eksperimen Algoritma Forward Selection

Dataset Sonar yang diperoleh dari UCI repository, dengan 208 sampel, 60 fitur (diidentifikasi sebagai X1 hingga X60), dan satu variabel kelas (X61 yang merepresentasikan 'M' atau 'R'), akan digunakan untuk melakukan perhitungan *Forward Selection*. Langkah awal dalam proses ini adalah mengukur hubungan korelasi antara masing-masing dari 60 fitur dengan variabel kelas. Validasi hasil analisis akan dilakukan menggunakan metode 10-*fold cross validation*.

Hasil Ekspirimen Algoritma *Information Gain Naive Bayes*

Penelitian ini mengimplementasikan perhitungan algoritma *Naive Bayes* secara manual dengan menggunakan pembobotan fitur *Information Gain* pada dataset Sonar yang bersumber dari UCI repository. Langkah ini dilakukan untuk memfasilitasi pemahaman mendetail mengenai setiap tahapan perhitungan algoritma. Metode validasi yang digunakan adalah 10-*fold cross validation*, yang membagi dataset menjadi 10 bagian, di mana 9 bagian digunakan untuk pelatihan dan 1 bagian untuk pengujian.

Perbandingan Kinerja Algoritma

Tahapan selanjutnya dalam penelitian ini adalah membandingkan kinerja dari berbagai algoritma klasifikasi. Algoritma yang dibandingkan adalah *Naive Bayes* (NB), *Forward Selection Naive Bayes* (FSNB), *Forward Selection Information Gain Naive Bayes* (FSIGNB), dan *Forward Selection Information Gain Ratio Naive Bayes* (FSIGRNB). Tujuan dari perbandingan ini adalah untuk mengidentifikasi algoritma dengan kinerja terbaik. Rincian perbandingan akurasi untuk masing-masing algoritma ditampilkan pada Tabel 2.

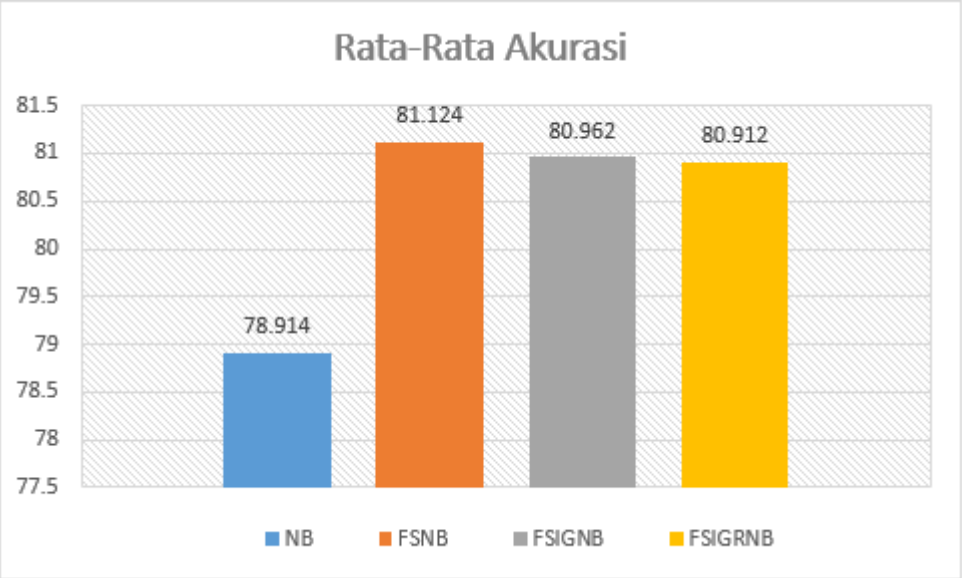
Tabel 2. Perbandingan Akurasi Tiap Algoritma

Dataset	Naive Bayes	FSNB	FSIGNB	FSIGRNB
Glass	54.67%	60.39%	59.37%	58.81%
Ionesphare	90.60%	91.74%	91.48%	91.75%
Sonar	73.08%	75.07%	75.48%	75.60%

Dataset	Naive Bayes	FSNB	FSIGNB	FSIGRNB
Waveform5000	80.22%	82.42%	82.48%	82.40%

Tabel 2 menampilkan perbandingan tingkat akurasi dari empat algoritma yang diuji pada lima dataset yang berbeda. Khusus untuk dataset Glass, algoritma FSNB menunjukkan performa terbaik dengan tingkat akurasi tertinggi mencapai 60,39%, yang ditandai dengan warna pink dalam tabel. Sementara itu, algoritma lain menunjukkan hasil yang sedikit lebih rendah, yaitu FSIGNB dengan akurasi 59,37%, diikuti oleh FSIGR sebesar 58,81%, dan algoritma NB dengan akurasi terendah yaitu 54,67%. Pada dataset Ionosphere, yang merupakan dataset kedua, algoritma FSIGRNB mencatatkan akurasi tertinggi sebesar 91,75%, yang diberi penanda warna biru dalam tabel. Sementara itu, algoritma FSNB memiliki akurasi yang sangat dekat yaitu 91,74%, diikuti oleh FSIGNB dengan akurasi 91,48%, dan algoritma NB dengan akurasi paling rendah di antara keempatnya yaitu 90,60%. Pada dataset Sonar, yang merupakan dataset ketiga dalam perbandingan di Tabel 2, algoritma FSIGRNB memperoleh akurasi tertinggi sebesar 75,60%, yang diberi penanda warna biru. Algoritma lainnya menunjukkan hasil yang cukup dekat, dengan FSIGNB mencatatkan akurasi 75,48%, diikuti oleh FSNB yang memiliki akurasi 75,07%, dan algoritma NB yang mencatatkan akurasi terendah sebesar 73,08%. Pada dataset WaveForm5000, yang merupakan dataset keempat dalam perbandingan di Tabel 2, algoritma FSIGNB mencatatkan akurasi tertinggi sebesar 82,48%, yang diberi penanda warna teal. Algoritma lainnya memiliki hasil yang sangat mendekati, dengan FSNB memperoleh akurasi 82,42%, diikuti oleh FSIGRNB yang mencatatkan akurasi 82,40%, sementara algoritma NB mencatatkan akurasi 80,22%.

Grafik akurasi menggambarkan perbandingan kinerja algoritma pada dataset Glass yang memiliki 10 fitur, 7 kelas, dan 214 baris data. Dari grafik tersebut terlihat bahwa algoritma FSNB memberikan akurasi tertinggi sebesar 60.39%. Sementara itu, akurasi algoritma lainnya pada dataset Glass adalah FSIGNB sebesar 59.37%, FSIGR sebesar 58.81%, dan NB sebesar 54.67%. Grafik akurasi pada dataset Ionosphere yang memiliki 34 fitur, 2 kelas, dan 351 baris data. Dari grafik tersebut terlihat bahwa algoritma FSIGRNB memberikan akurasi tertinggi sebesar 91.75%. Sementara itu, akurasi algoritma lainnya pada dataset Ionosphere adalah FSNB sebesar 91.74%, FSIGNB sebesar 91.48%, dan NB sebesar 90.60%. Pada dataset Sonar, yang terdiri dari 60 fitur, 2 kelas, dan 208 baris, algoritma FSIGRNB mencatatkan akurasi tertinggi sebesar 75,60%. Diikuti oleh FSIGNB dengan akurasi 75,48%, FSNB dengan akurasi 75,07%, dan algoritma NB yang mencatatkan akurasi terendah sebesar 73,08%. Sedangkan untuk dataset Waveform-5000, yang memiliki 41 fitur, 3 kelas, dan 1000 baris, grafik akurasi menggambarkan bahwa dataset ini jauh lebih besar dan lebih kompleks dibandingkan dengan dataset Sonar. Pada dataset Waveform5000, algoritma FSIGNB mencatatkan akurasi tertinggi sebesar 82,48%, diikuti oleh FSNB dengan akurasi 82,42%, FSIGRNB yang memperoleh akurasi 82,40%, dan algoritma NB dengan akurasi terendah sebesar 80,22%. Jika dilihat dari nilai rata-rata akurasi, algoritma FSNB memiliki nilai tertinggi dengan 81,12%, diikuti oleh FSIGNB dengan nilai 80,96%, FSIGRNB dengan 80,91%, dan algoritma NB dengan nilai rata-rata 78,91%.



Gambar 1. Grafik Rata-rata Akurasi tiap Algoritma

Berdasarkan hasil perbandingan yang telah disampaikan antara algoritma *Naive Bayes*, *Forward Selection Naive Bayes, Forward Selection Information Gain Naive Bayes*, dan *Forward Selection Information Gain Ratio Naive Bayes*, dapat dilihat bahwa penerapan teknik-teknik ini berhasil meningkatkan akurasi algoritma *Naive Bayes*. Hal ini dapat dilihat pada Tabel 4.19 yang menunjukkan perbandingan akurasi masing-masing algoritma dan Gambar 4.6 yang menggambarkan grafik rata-rata akurasi tiap algoritma. Pendekatan seleksi fitur menggunakan *Forward Selection* dan pembobotan fitur dengan *Weight Information Gain* serta *Gain Ratio* pada *Naive Bayes* terbukti efektif, karena mampu menilai dan menentukan bobot setiap fitur dengan baik. Keberhasilan ini terjadi karena teknik-teknik tersebut menggabungkan nilai seleksi fitur terbaik dan bobot fitur dalam rumus klasifikasi *Naive Bayes*. Oleh karena itu, penelitian ini mendukung bahwa algoritma *Forward Selection* dan *Weighted Information Gain* serta *Gain Ratio Naive Bayes* dapat menghasilkan performa yang lebih baik dibandingkan dengan algoritma *Naive Bayes* standar.

KESIMPULAN

Kesimpulan penelitian menunjukkan bahwa algoritma *Forward Selection, Information Gain, dan Gain Ratio* (FSIGRNB) memberikan peningkatan kinerja yang signifikan dibandingkan dengan *Naïve Bayes* standar, terutama dalam menangani data dengan jumlah besar dan fitur yang banyak. Berdasarkan pemberian bobot yang lebih tepat pada setiap fitur melalui seleksi dan pembobotan yang tepat, algoritma ini mampu mengatasi masalah independensi fitur yang sering menurunkan akurasi *Naïve Bayes*. Hasil penelitian menunjukkan bahwa algoritma FSIGRNB menghasilkan akurasi yang lebih baik, menjadikannya lebih efektif dalam pengolahan data besar dengan dimensi fitur tinggi. Menurut hasil tersebut, penelitian ini menyarankan agar algoritma FSIGRNB diterapkan lebih luas dalam berbagai bidang yang membutuhkan pengolahan data besar dengan banyak fitur, seperti analisis teks, pengenalan pola, dan prediksi data. Ke depan, penelitian selanjutnya dapat dilakukan pengembangan lebih lanjut untuk memperbaiki algoritma ini dengan teknik pembobotan atau seleksi fitur lainnya yang dapat meningkatkan performa dalam kondisi yang lebih variatif. Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan pada pengembangan algoritma klasifikasi yang lebih efisien dan akurat.

DAFTAR PUSTAKA

Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, *10*(2). https://doi.org/10.33096/ilkom.v10i2.303.160-165

Asaad, R. R., & Abdulhakim, R. M. (2021). The Concept of Data Mining and Knowledge Extraction Techniques. *Qubahan Academic Journal*, *1*(2). https://doi.org/10.48161/qaj.v1n2a43

Attamami, N., Triayudi, A., & Aldisa, R. T. (2023). Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *7*(2). https://doi.org/10.35870/jtik.v7i2.756

Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vols. 1–3). https://doi.org/10.1016/B978-0-12-809633-8.20473-1

Damar Rani, H. A., & Zuhri, S. (2020). Sistem Prediksi Kondisi Kelahiran Bayi menggunakan Klasifikasi Naïve Bayes. *Joined Journal (Journal of Informatics Education)*, *3*(2). https://doi.org/10.31331/joined.v3i2.1432

Depari, D. H., Widiastiwi, Y., & Santoni, M. M. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *Informatik : Jurnal Ilmu Komputer*, *18*(3). https://doi.org/10.52958/iftk.v18i3.4694

Ericha Apriliyani, & Salim, Y. (2022). Analisis performa metode klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset. *Indonesian Journal of Data and Science*, *3*(2). https://doi.org/10.56705/ijodas.v3i2.45

Guohua, W., & Francis, E. H. (2017). Data Mining: Concept, Aplications and Techniques. *ASEAN Journal on Science and Technology for Development*, *17*(1). https://doi.org/10.29037/ajstd.134

Harungguan, A. R., Napitupulu, H., & Firdaniza, F. (2023). Analisis Sentimen Dengan Metode Klasifikasi Naïve Bayes dan Seleksi Fitur Chi-Square. *In Search*, *22*(2). https://doi.org/10.37278/insearch.v22i2.762

Istighfar, F., Negara, A. B. P., & Tursina, T. (2023). Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma Naive Bayes. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, *11*(1). https://doi.org/10.26418/justin.v11i1.52402

Jiang, L., Li, C., Wang, S., & Zhang, L. (2016). Deep feature weighting for naive Bayes and its application to text classification. *Engineering Applications of Artificial Intelligence*, *52*. https://doi.org/10.1016/j.engappai.2016.02.002

Mahmood, H. A. (2018). Network Intrusion Detection System (NIDS) in Cloud Environment based on Hidden Naïve Bayes Multiclass Classifier. *Al-Mustansiriyah Journal of Science*, *28*(2). https://doi.org/10.23851/mjs.v28i2.508

Mostafa, A. A. N., & Mahmoud, H. E. A. (2022). Review of Data Mining Concept and its Techniques. *International Journal of Academic Research in Business and Social Sciences*, *12*(6). https://doi.org/10.6007/ijarbss/v12-i6/13135

Nazneentarannum, M., & Rizvi, S. H. (2016). A Systematic Overview On Data Mining: Concepts And Techniques. *International Journal of Research in Computer & Information Technology (IJRCIT)*, *1*.

Prasdika, P., & Sugiantoro, B. (2018). A Review Paper on Big Data and Data Mining Concepts and Techniques. *IJID (International Journal on Informatics for Development)*, *7*(1). https://doi.org/10.14421/ijid.2018.07107

Ramadhan, T. D., Wahiddin, D., & Awal, E. E. (2023). Klasifikasi Sentimen Terhadap Pinjaman Online (Pinjol) Menggunakan Algoritma Naive Bayes. *Scientific Student Journal for Information, Technology and Science*, *IV*(1).

Rizki, M., Arhami, M., & Huzeni, H. (2021). Perbaikan Algoritma Naive Bayes Classifier Menggunakan Teknik Laplacian Correction. *Jurnal Teknologi*, *21*(1). https://doi.org/10.30811/teknologi.v21i1.2209

Shafiee, S., Lied, L. M., Burud, I., Dieseth, J. A., Alsheikh, M., & Lillemo, M. (2021). Sequential

forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Computers and Electronics in Agriculture, 183*. https://doi.org/10.1016/j.compag.2021.106036

Yang, C., Zhu, X., Qiao, J., & Nie, K. (2020). Forward and backward input variable selection for polynomial echo state networks. *Neurocomputing, 398*. https://doi.org/10.1016/j.neucom.2020.02.034

Zhang, L., Jiang, L., Li, C., & Kong, G. (2016). Two feature weighting approaches for naive Bayes text classifiers. *Knowledge-Based Systems, 100*. https://doi.org/10.1016/j.knosys.2016.02.017