

# Analisis Pengaruh Seleksi Atribut Relief-F dan Gain Ratio Terhadap Performa Naïve Bayes Classifier

<sup>1</sup>Agung RM Alam, <sup>2</sup>Wanayumini, <sup>3</sup>Lili Tanti

<sup>1\*,2,3</sup>Program Studi Magister Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia

\*Korespondensi: [agungprakom@gmail.com](mailto:agungprakom@gmail.com)

Submit : 17 April 2026 | Diterima : 28 Mei 2026 | Terbit : 05 Jun 2026

## ABSTRACT

*Naïve Bayes Classifier (NBC) is one of the most popular probabilistic classification algorithms in data mining, known for its simplicity and efficiency. However, NBC performance tends to degrade when datasets contain irrelevant or noisy attributes. This study analyzes the effect of attribute selection using Relief-F and Gain Ratio methods on the performance improvement of NBC. Two benchmark datasets from the UCI Machine Learning Repository were selected to represent contrasting data characteristics: the House Vote dataset (435 records, symbolic attributes with balanced class distribution) and the Bank Marketing dataset (45,211 records, numeric and categorical attributes with severe class imbalance, approximately 88% majority class). All experiments were implemented in Google Colaboratory using Python. Three experimental scenarios were applied to each dataset: (1) NBC without attribute selection as baseline, (2) NBC with Relief-F attribute selection, and (3) NBC with Gain Ratio attribute selection. Performance evaluation used 10-fold cross-validation with metrics including accuracy, precision, recall, F1-score, and confusion matrix. Results show that on the House Vote dataset, Relief-F increased NBC accuracy from 90.11% to 93.79% (+3.68%), while Gain Ratio reduced accuracy to 89.43%. On the Bank Marketing dataset, Relief-F improved accuracy to 89.36% and improved minority class recall from 29.34% to 35.71%, while Gain Ratio yielded only marginal improvement. Overall, Relief-F proved more effective than Gain Ratio in enhancing NBC performance, particularly on datasets with clear classification patterns and imbalanced class distribution.*

**Keywords:** Attribute Selection; Classification; Data Mining; Gain Ratio; Naïve Bayes Classifier; Relief-F.

## ABSTRAK

Naïve Bayes Classifier (NBC) merupakan salah satu algoritma klasifikasi probabilistik yang paling banyak digunakan dalam data mining karena kesederhanaan dan efisiensinya. Namun, performa NBC cenderung menurun ketika dataset mengandung atribut yang tidak relevan atau mengandung noise. Penelitian ini bertujuan menganalisis pengaruh seleksi atribut menggunakan metode Relief-F dan Gain Ratio terhadap peningkatan performa NBC. Dua dataset dari UCI Machine Learning Repository dipilih untuk merepresentasikan karakteristik data yang kontras: dataset House Vote (435 data, atribut simbolik dengan distribusi kelas yang seimbang) dan dataset Bank Marketing (45.211 data, atribut numerik dan kategorikal dengan ketidakseimbangan kelas yang signifikan, sekitar 88% kelas mayoritas). Seluruh eksperimen diimplementasikan menggunakan Google Colaboratory dengan bahasa Python. Tiga skenario eksperimen diterapkan pada masing-masing dataset: (1) NBC tanpa seleksi atribut sebagai baseline, (2) NBC dengan seleksi atribut Relief-F, dan (3) NBC dengan seleksi atribut Gain Ratio. Evaluasi performa menggunakan 10-fold cross-validation dengan metrik akurasi, presisi, recall, F1-score, dan confusion matrix. Hasil penelitian menunjukkan bahwa pada dataset House Vote, Relief-F berhasil meningkatkan akurasi NBC dari 90,11% menjadi 93,79% (+3,68%), sedangkan Gain Ratio justru menurunkan akurasi menjadi 89,43%. Pada dataset Bank Marketing, Relief-F meningkatkan akurasi menjadi 89,36% dan memperbaiki recall kelas minoritas dari 29,34% menjadi 35,71%, sementara Gain Ratio hanya memberikan peningkatan marginal. Secara keseluruhan, Relief-F terbukti lebih efektif dibandingkan Gain Ratio dalam meningkatkan performa NBC, khususnya pada dataset dengan pola klasifikasi yang jelas dan distribusi kelas yang tidak seimbang.

**Kata Kunci:** Algoritma Klasifikasi; Data Mining; Gain Ratio; Naïve Bayes Classifier; Relief-F; Seleksi Atribut.

## PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong pertumbuhan volume data secara eksponensial di berbagai bidang, mulai dari kesehatan, keuangan, pemasaran, hingga politik. Data dalam jumlah besar tersebut mengandung pola-pola tersembunyi yang dapat diekstrak menjadi pengetahuan berguna melalui teknik *data mining*. Salah satu teknik data mining yang paling banyak digunakan adalah klasifikasi, yaitu proses membangun model prediktif untuk menentukan label kelas data baru berdasarkan pola yang dipelajari dari data pelatihan (Alnuaimi & Albaldawi, 2024).

Di antara berbagai algoritma klasifikasi yang tersedia, Naïve Bayes Classifier (NBC) telah lama dikenal sebagai metode yang sederhana namun efektif. NBC menggunakan prinsip Teorema Bayes dengan asumsi independensi bersyarat antar atribut untuk menghitung probabilitas kelas suatu data (Lighthart et al., 2021). Algoritma ini memiliki sejumlah keunggulan: kebutuhan data pelatihan yang relatif kecil, kompleksitas waktu yang rendah, kemampuan menangani *missing value*, serta kinerja yang kompetitif untuk berbagai jenis data (Arfan Haqiqi et al., 2021). Meski demikian, NBC memiliki kelemahan mendasar; performanya dapat menurun secara signifikan ketika dataset mengandung atribut yang tidak relevan, redundan, atau mengandung *noise*.

Untuk mengatasi permasalahan tersebut, seleksi atribut (*feature selection*) menjadi langkah prapemrosesan yang krusial. Seleksi atribut bertujuan memilih subset atribut yang paling relevan terhadap variabel kelas sehingga model yang dihasilkan lebih akurat, efisien secara komputasi, dan lebih mudah diinterpretasikan (Bommert et al., 2022). Dalam penelitian ini, dua metode seleksi atribut dievaluasi secara komparatif, yaitu Relief-F dan Gain Ratio.

Relief-F merupakan pengembangan dari algoritma Relief yang dikembangkan Kononenko (1994). Algoritma ini menilai relevansi setiap atribut berdasarkan kemampuannya membedakan instance dari kelas yang sama (*nearest hit*) dan kelas yang berbeda (*nearest miss*). Pendekatan berbasis instance ini menjadikan Relief-F adaptif terhadap interaksi antar atribut dan efektif untuk data berdimensi tinggi (Yusra et al., 2021). Sementara itu, Gain Ratio merupakan pengembangan dari Information Gain yang menambahkan normalisasi melalui Split Information untuk mengurangi bias terhadap atribut bernilai banyak (Edusainstek et al., 2018).

Meskipun kedua metode telah banyak dikaji secara terpisah, belum banyak penelitian yang mengevaluasi dan membandingkan efektivitas Relief-F dan Gain Ratio secara langsung dalam konteks optimasi NBC, khususnya pada dataset dengan ketidakseimbangan kelas yang signifikan. Kesenjangan penelitian ini menjadi dasar dan motivasi utama studi yang disajikan dalam makalah ini.

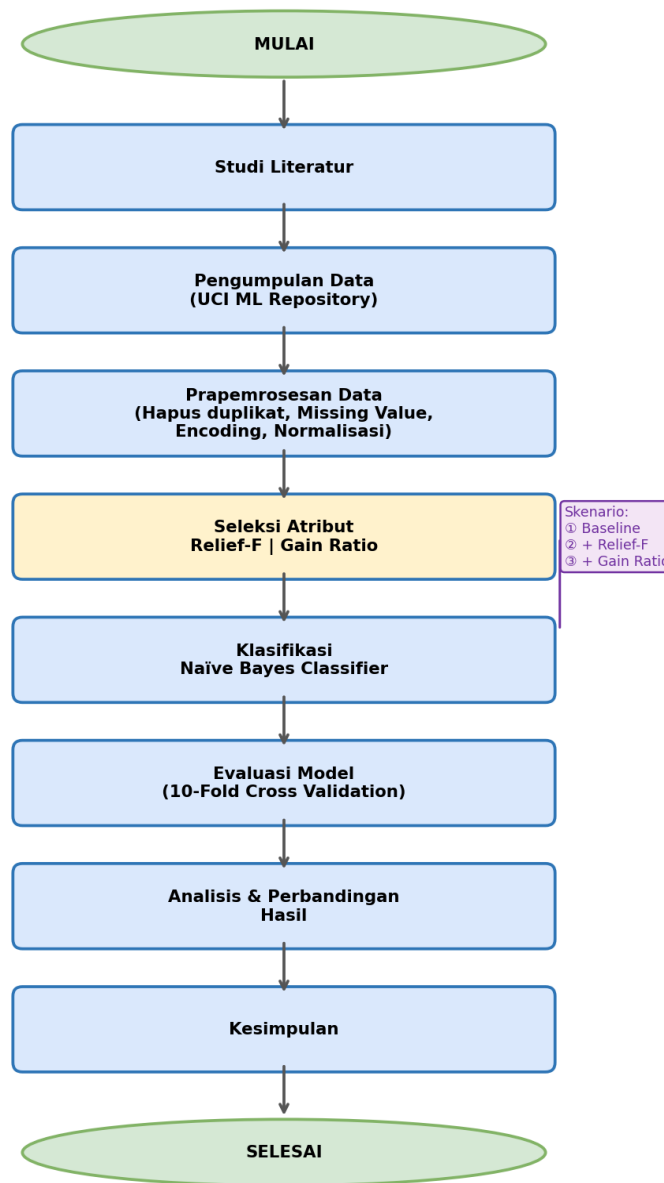
Penelitian ini menggunakan dua dataset benchmark dari UCI Machine Learning Repository dengan karakteristik berbeda: dataset House Vote (435 data, atribut simbolik) dan dataset Bank Marketing (45.211 data, atribut numerik dan kategorikal dengan distribusi kelas tidak seimbang). Tujuan penelitian adalah: (1) menganalisis pengaruh seleksi atribut Relief-F dan Gain Ratio terhadap performa NBC, (2) membandingkan efektivitas kedua metode seleksi atribut, dan (3) mengevaluasi konsistensi hasil pada dataset dengan karakteristik yang berbeda.

## METODE PENELITIAN

### Tahapan Penelitian

Secara keseluruhan, alur penelitian ini terdiri dari delapan tahapan utama yang dilaksanakan secara sistematis, mulai dari studi literatur hingga penarikan kesimpulan. Setiap tahapan saling berkaitan dan dirancang untuk memastikan validitas serta reliabilitas hasil yang diperoleh. Bagan alur penelitian disajikan pada Gambar 1 berikut.

### Bagan Alur Penelitian



Gambar 1. Bagan Alur Penelitian

#### Dataset yang Digunakan

Dataset pertama adalah House Vote yang terdiri atas 435 data dengan 16 atribut voting simbolik (yes/no/?) dan satu atribut kelas berupa afiliasi partai (democrat/republican). Dataset kedua adalah Bank Marketing yang terdiri atas 45.211 data dengan atribut numerik dan kategorikal yang merepresentasikan data demografis dan historis kampanye pemasaran bank, dengan variabel kelas berupa keputusan nasabah terhadap penawaran deposito berjangka (yes/no). Dataset Bank Marketing memiliki distribusi kelas yang tidak seimbang, dengan proporsi kelas "no" mencapai sekitar 88% dari total data.

#### Prapemrosesan Data

Sebelum proses klasifikasi, dilakukan prapemrosesan data yang meliputi: (1) penghapusan data duplikat untuk menghindari bias model, (2) penanganan *missing value* dengan penggantian menggunakan nilai rata-rata pada atribut numerik dan modus pada atribut kategorikal, (3) encoding atribut kategorikal menggunakan teknik *label encoding*, dan (4) normalisasi data numerik ke rentang [0,1].

### Seleksi Atribut Relief-F

Relief-F bekerja dengan memilih secara acak sejumlah instance sebagai sampel target, kemudian mencari  $k$  tetangga terdekat dari kelas yang sama (*nearest hit*) dan kelas yang berbeda (*nearest miss*). Bobot setiap atribut diperbarui berdasarkan perbedaan nilai atribut antara sampel target dengan tetangga-tetangganya menggunakan formula:

$$W[A] = W[A] - \text{diff}(A, R, H) + \text{diff}(A, R, M)$$

Atribut yang nilai-nilainya berbeda antara *nearest hit* dan *nearest miss* mendapat bobot lebih tinggi karena dianggap mampu membedakan kelas dengan baik. Pada penelitian ini digunakan Top-K = 16 atribut terbaik untuk dataset House Vote.

### Seleksi Atribut Gain Ratio

Gain Ratio menghitung relevansi atribut melalui tiga tahap: (1) menghitung Entropy dataset sebagai ukuran ketidakpastian kelas, (2) menghitung Information Gain setiap atribut terhadap kelas target, dan (3) menghitung Gain Ratio dengan membagi Information Gain dengan Split Information untuk mengurangi bias. Rumus Gain Ratio adalah:

$$\text{Gain Ratio}(S, A) = \text{Gain}(S, A) / \text{SplitInfo}(S, A)$$

Pada penelitian ini digunakan Top-K = 10 atribut terbaik untuk kedua dataset.

### Algoritma Naïve Bayes Classifier

Setelah seleksi atribut, model NBC dilatih menggunakan atribut yang terpilih. NBC menghitung probabilitas posterior setiap kelas berdasarkan Teorema Bayes:  $P(C|X) = P(X|C) \times P(C) / P(X)$ . Dengan asumsi independensi bersyarat, probabilitas likelihood dihitung sebagai perkalian probabilitas masing-masing atribut:  $P(C|X) \propto P(C) \times P(X_1|C) \times P(X_2|C) \times \dots \times P(X_n|C)$ . Kelas dengan probabilitas posterior tertinggi dipilih sebagai prediksi.

### Evaluasi Model

Semua eksperimen menggunakan teknik 10-fold cross-validation untuk memperoleh estimasi performa yang stabil dan tidak bias. Dataset dibagi menjadi 10 bagian (fold); setiap fold secara bergantian dijadikan data uji sementara sembilan fold lainnya menjadi data latih, dan hasilnya dirata-ratakan. Metrik evaluasi yang digunakan meliputi akurasi, presisi, recall, F1-score, dan confusion matrix. Seluruh pengujian dijalankan menggunakan Google Colaboratory.

## HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengujian sesuai tahapan metodologi yang telah diuraikan, meliputi prapemrosesan data, seleksi atribut, klasifikasi NBC, dan evaluasi model. Pembahasan dilakukan secara komparatif pada dua dataset untuk menilai efektivitas masing-masing skenario.

### Prapemrosesan Data

Pada dataset House Vote, ditemukan sejumlah nilai yang hilang (missing value) pada beberapa atribut voting yang direpresentasikan dengan tanda tanya (?). Nilai-nilai tersebut digantikan dengan nilai modus kelas masing-masing atribut. Tidak ditemukan data duplikat pada dataset ini. Pada dataset Bank Marketing, ditemukan beberapa data duplikat yang dieliminasi sebelum proses selanjutnya, serta sejumlah atribut kategorikal seperti job, marital, education, dan contact yang diubah menjadi representasi numerik melalui label encoding. Normalisasi kemudian diterapkan pada atribut numerik seperti age, balance, dan duration.

### Seleksi Atribut

Pada dataset House Vote, proses Relief-F dengan Top-K = 16 mengidentifikasi atribut *physician-fee-freeze* sebagai atribut paling diskriminatif. Atribut ini memiliki bobot Relief-F tertinggi sekaligus Gain Ratio tertinggi (0,657), yang mengindikasikan relevansinya yang sangat kuat terhadap afiliasi partai. Sebaliknya, atribut *water-project-cost-sharing* dan *immigration* memiliki skor yang rendah pada keduanya sehingga tidak dipilih oleh Gain Ratio (Top-K = 10).

Pada dataset Bank Marketing, Gain Ratio memilih 10 atribut terbaik yaitu: *poutcome*, *duration*, *previous*, *contact*, *housing*, *month*, *loan*, *job*, *default*, dan *campaign*. Atribut-atribut ini mencerminkan karakteristik historis kampanye pemasaran yang memiliki pengaruh informatif terkuat secara individual. Relief-F pada dataset yang sama memilih subset yang sedikit berbeda dengan mempertimbangkan kedekatan antar instance sehingga lebih sensitif terhadap pola lokal data.

**Hasil Klasifikasi – Dataset House Vote**

Pengujian pada dataset House Vote dengan tiga skenario menghasilkan performa yang disajikan pada Tabel 1 berikut.

**Tabel 1. Perbandingan Performa Model pada Dataset House Vote**

Metrik	NBC Baseline	NBC + Relief-F	NBC + Gain Ratio
Akurasi	90,11%	93,79%	89,43%
Presisi (Democrat)	94,44%	96,88%	93,68%
Recall (Democrat)	89,14%	92,88%	88,76%
F1-score (Democrat)	91,71%	94,84%	91,15%
Presisi (Republican)	84,15%	89,39%	83,52%
Recall (Republican)	91,67%	95,24%	90,48%
F1-score (Republican)	87,75%	92,22%	86,86%
Macro Average F1-score	89,73%	93,53%	89,01%
Weighted Avg F1-score	90,18%	93,83%	89,49%

NBC baseline menghasilkan akurasi 90,11%, menunjukkan bahwa pola voting anggota Kongres cukup terstruktur untuk membedakan kedua kelas partai. Meski demikian, masih terdapat 29 data democrat yang salah diklasifikasikan sebagai republican dan 14 data republican yang salah diklasifikasikan sebagai democrat.

Penerapan seleksi atribut Relief-F dengan Top-K = 16 memberikan peningkatan yang signifikan: akurasi naik menjadi 93,79% (+3,68%), kesalahan pada kelas democrat berkurang dari 29 menjadi 19 data, dan kesalahan pada kelas republican berkurang dari 14 menjadi 8 data. Peningkatan ini konsisten pada seluruh metrik, termasuk Macro Average F1-score yang naik dari 89,73% menjadi 93,53%.

Sebaliknya, Gain Ratio dengan Top-K = 10 justru menurunkan akurasi menjadi 89,43%. Kondisi ini terjadi karena Gain Ratio menghilangkan sejumlah atribut yang secara individual kurang informatif, namun masih memberikan kontribusi probabilistik yang berarti bagi NBC ketika dikombinasikan bersama atribut lain.

**Hasil Klasifikasi – Dataset Bank Marketing**

Pengujian pada dataset Bank Marketing yang berskala besar dengan distribusi kelas tidak seimbang menghasilkan performa yang disajikan pada Tabel 2.

**Tabel 2. Perbandingan Performa Model pada Dataset Bank Marketing**

Metrik	NBC Baseline	NBC + Relief-F	NBC + Gain Ratio
Akurasi	88,74%	89,36%	88,83%
Presisi (No)	91,17%	91,88%	91,03%
Recall (No)	96,61%	96,31%	96,90%
F1-score (No)	93,81%	94,04%	93,87%
Presisi (Yes)	53,44%	56,20%	54,42%
Recall (Yes) – Kelas Minoritas	29,34%	35,71%	27,96%
F1-score (Yes)	37,89%	43,60%	36,94%
Macro Avg F1-score	65,85%	69,02%	65,41%
Weighted Avg F1-score	87,27%	88,01%	87,21%

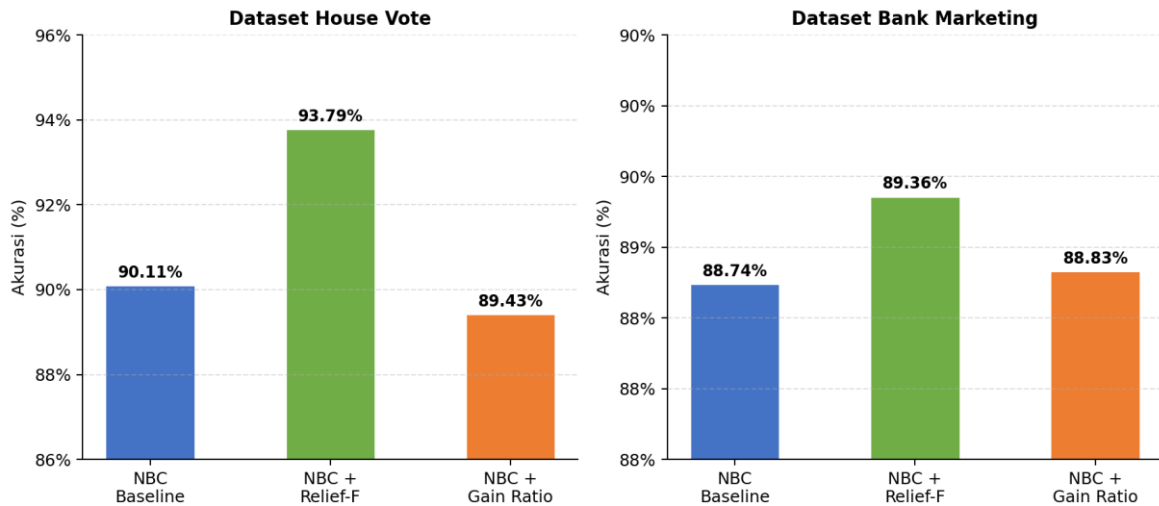
NBC baseline menghasilkan akurasi 88,74% yang secara angka tampak tinggi, namun analisis mendalam mengungkapkan bahwa performa ini didominasi oleh keberhasilan mengenali kelas mayoritas "no" (recall 96,61%), sedangkan kemampuan mendeteksi kelas

minoritas "yes" sangat rendah (recall hanya 29,34%). Kondisi ini merupakan konsekuensi langsung dari distribusi kelas yang timpang.

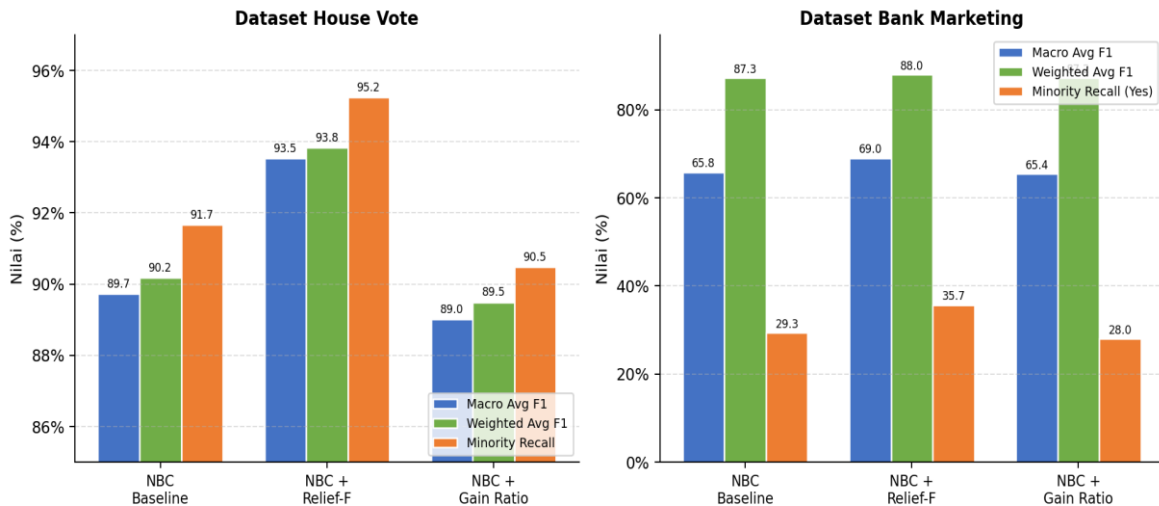
Relief-F memberikan perbaikan yang lebih bermakna dibandingkan Gain Ratio: akurasi meningkat menjadi 89,36%, dan yang lebih penting, recall kelas "yes" meningkat dari 29,34% menjadi 35,71%, serta Macro Average F1-score naik dari 65,85% menjadi 69,02%. Gain Ratio sebaliknya menghasilkan recall kelas "yes" yang bahkan lebih rendah dari baseline (27,96%), menandakan bahwa atribut yang dipilih Gain Ratio kurang sensitif terhadap kelas minoritas.

### Perbandingan Antar Metode

Untuk memudahkan analisis komparatif performa ketiga skenario pada kedua dataset, hasil evaluasi disajikan secara visual dalam dua grafik berikut.



Gambar 2. Perbandingan Akurasi Antar Metode pada Kedua Dataset



Gambar 3. Perbandingan Metrik Evaluasi Antar Metode pada Kedua Dataset

Gambar 2 secara jelas memperlihatkan bahwa NBC + Relief-F unggul dalam hal akurasi pada kedua dataset. Pada House Vote, peningkatan akurasi sebesar 3,68% sangat nyata secara visual. Pada Bank Marketing, perbedaan akurasi antar metode lebih kecil, mengindikasikan bahwa masalah utama di sana bukan semata pemilihan atribut melainkan ketidakseimbangan kelas.

Gambar 3 memperkuat temuan tersebut dengan menampilkan tiga metrik sekaligus: Macro Average F1, Weighted Avg F1, dan Minority Recall. Pada dataset Bank Marketing, perbedaan Minority Recall antar metode sangat mencolok: Relief-F mencapai 35,71% sementara Gain Ratio hanya 27,96%, bahkan lebih rendah dari baseline 29,34%. Hal ini menegaskan

superioritas Relief-F dalam konteks dataset tidak seimbang karena pendekatan berbasis instance-nya lebih sensitif terhadap pola lokal kelas minoritas.

Perbedaan mendasar antara kedua metode terletak pada paradigma evaluasi atribut. Relief-F mengevaluasi relevansi atribut secara kontekstual berdasarkan kedekatan antar instance data aktual, sehingga mampu menangkap interaksi antar atribut yang tidak dapat diidentifikasi oleh Gain Ratio. Sebaliknya, Gain Ratio mengukur kekuatan hubungan atribut-kelas secara global menggunakan teori informasi — lebih sesuai untuk pembentukan pohon keputusan (*decision tree*) daripada model probabilistik seperti NBC.

## KESIMPULAN

Penelitian ini telah menganalisis dan membandingkan efektivitas dua metode seleksi atribut, Relief-F dan Gain Ratio, dalam meningkatkan performa Naïve Bayes Classifier pada dua dataset benchmark yang merepresentasikan karakteristik data berbeda. Secara keseluruhan, Relief-F terbukti unggul secara konsisten dibandingkan Gain Ratio: pada dataset House Vote, Relief-F meningkatkan akurasi NBC sebesar 3,68% (dari 90,11% menjadi 93,79%), sedangkan Gain Ratio justru menurunkannya; pada dataset Bank Marketing dengan distribusi kelas yang timpang, Relief-F secara signifikan meningkatkan recall kelas minoritas dari 29,34% menjadi 35,71%, sementara Gain Ratio menghasilkan recall yang bahkan lebih rendah dari baseline. Keunggulan ini dapat dijelaskan dari perspektif teoretis: Relief-F mengevaluasi relevansi atribut secara kontekstual melalui pendekatan berbasis instance yang mampu menangkap interaksi antar atribut, sehingga menghasilkan subset atribut yang lebih komplementer terhadap asumsi probabilistik NBC; sebaliknya, Gain Ratio mengukur relevansi atribut secara individual menggunakan teori informasi—paradigma yang lebih sesuai untuk pohon keputusan daripada model berbasis Teorema Bayes. Secara praktis, temuan ini menegaskan bahwa pemilihan metode seleksi atribut harus mempertimbangkan kesesuaian paradigma evaluasi dengan arsitektur model yang digunakan, serta bahwa pada kasus ketidakseimbangan kelas yang ekstrem, seleksi atribut perlu dikombinasikan dengan teknik seperti SMOTE. Penelitian lanjutan disarankan untuk mengeksplorasi variasi nilai Top-K secara empiris, memperluas perbandingan dengan metode seleksi atribut lain seperti Mutual Information dan Chi-Square, serta menguji kombinasi metode pada beragam domain dataset.

## REFERENSI

- Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, 97, 1–24. <https://doi.org/10.1051/bioconf/20249700133>
- Arfan Haqiqi, R., Istiqomah Dwi Andari, & Siti Fatimah. (2021). Prediction of medical actions for Covid patients using Naïve Bayes method. *Elkom: Jurnal Elektronika Dan Komputer*, 14(2), 261–267. <https://doi.org/10.51903/elkom.v14i2.464>
- Bommert, A., Welchowski, T., & Schmid, M. (2022). Benchmark of filter methods for feature selection in high-dimensional gene expression survival data. *Briefings in Bioinformatics*, 23(August 2021), 1–13.
- Edusainstek, S. N., Kurniawan, M. F., & Irawan, J. N. (2018). Peningkatan performa algoritma Naive Bayes dengan seleksi atribut. *Prosiding Seminar Nasional*, 53–61.
- Fitriani, I., Basuki, S., & Minarno, A. E. (2020). Seleksi fitur Relief-F pada klasifikasi malware Android menggunakan Support Vector Machine (SVM). *Jurnal Repositor*, 2(11), 1529. <https://doi.org/10.22219/repositor.v2i11.901>
- Freda, P. J., Ye, S., Zhang, R., Moore, J. H., & Urbanowicz, R. J. (2024). Assessing the limitations of relief-based algorithms in detecting higher-order interactions. *BioData Mining*. <https://doi.org/10.1186/s13040-024-00390-0>
- Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: A tertiary study. *Artificial Intelligence Review*, 54(7). <https://doi.org/10.1007/s10462-021-09973-3>
- Purba, B., & Syahputra, R. (2021). Implementasi metode Naive Bayes Classifier pada evaluasi kepuasan mahasiswa terhadap pembelajaran daring. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*. <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/4352>

- Ramadandi, S. J. (2020). Klasifikasi gaya belajar mahasiswa menggunakan metode Naïve Bayes Classifier. *Jurnal Teknologi Dan Informasi (JATI)*, 10(September), 170–179. <https://doi.org/10.34010/jati.v10i2>
- Riany, A. F., & Testiana, G. (2023). Penerapan data mining untuk klasifikasi penyakit. *Prosiding Seminar Nasional*, 297–305.
- Saelan, M. R. R., Sahputra, D. A., Widiastuti, W., & Gata, W. (2020). Komparasi algoritma klasifikasi untuk prediksi minat sekolah tinggi pelajar pada Students Alcohol Consumption. *Jurnal Sains Dan Informatika*, 6(2), 120–129. <https://doi.org/10.34128/jsi.v6i2.236>
- Sari, S. I. P., Pranoto, W. J., & Verdikha, N. A. (2023). Analisis pengaruh Gain Ratio untuk algoritma K-Nearest Neighbor pada klasifikasi data banjir di Kota Samarinda. *Jurnal Sains Komputer Dan Teknologi Informasi*, 6(1), 54–59. <https://doi.org/10.33084/jsakti.v6i1.5472>
- Setiyorini, T., & Asmono, R. T. (2020). Implementation of Gain Ratio and K-Nearest Neighbor for classification of student performance. *Jurnal Pilar Nusa Mandiri*, 16(1), 19–24. <https://doi.org/10.33480/pilar.v16i1.813>
- Xi, J., Jiang, Q., Liu, H., & Gao, X. (2023). Lithological mapping research based on feature selection model of Relief-F-RF. *Applied Sciences*.
- Yusra, R. N., Sitompul, O. S., & Sawaluddin. (2021). Kombinasi K-Nearest Neighbor (KNN) dan Relief-F untuk meningkatkan akurasi pada klasifikasi data. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*, 1, 0–5.
- Zhao, M., & Ye, N. (2024). High-dimensional ensemble learning classification: An ensemble learning classification algorithm based on high-dimensional feature space reconstruction. *Applied Sciences (Switzerland)*, 14(5). <https://doi.org/10.3390/app14051956>