BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

Optimasi Klasifikasi Kelayakan Perizinan Frekuensi Radio Menggunakan Varian Naïve Bayes

Deri Risyandi[*], Handoyo Widi Nugroho

Fakultas Ilmu Komputer, Institut Informatika Dan Bisnis Darmajaya, Bandar Lampung, Indonesia
Email: [1,*]deri.2421211025p@mail.darmajaya.ac.id, [2]handoyo.wn@darmajaya.ac.id
Email Penulis Korespondensi: deri.2421211025p@mail.darmajaya.ac.id

Abstrak−Pengelolaan spektrum frekuensi radio yang efisien membutuhkan sistem evaluasi perizinan yang cepat dan akurat untuk mencegah interferensi sinyal serta menjamin kepatuhan regulasi. Tingginya volume pengajuan sering kali menyebabkan hambatan pemrosesan pada layanan perizinan publik. Penelitian ini bertujuan untuk mengoptimalkan klasifikasi permohonan Izin Stasiun Radio (ISR) menjadi kelas Diterima (*Granted*) dan Ditolak (*Rejected*) menggunakan varian algoritma *Naïve Bayes*. Dataset riil yang digunakan mencakup 38.954 rekor data pengajuan dengan 26 fitur teknis, administrasi, dan geografis. Pra-pemrosesan data melibatkan eliminasi fitur kebocoran pasca-keputusan, imputasi median, label *encoding*, serta transformasi/skala fitur. Tiga varian algoritma dievaluasi: *Gaussian Naïve Bayes* Standar, *Gaussian Naïve Bayes* dengan *PowerTransformer* dan *Complement Naïve Bayes*. Hasil eksperimen menunjukkan bahwa *Gaussian Naïve Bayes* dengan *PowerTransformer* mencapai akurasi keseluruhan tertinggi sebesar 78,32% dan presisi sebesar 88,63% (5.112 *True Negative* dan 990 *True Positive*). Sebaliknya, *Complement Naïve Bayes* efektif menangani ketidakseimbangan kelas dengan mencapai *recall* yang signifikan lebih tinggi sebesar 73,71% dan *F1-Score* sebesar 64,21%. Kesimpulannya, transformasi fitur teknis *non-Gaussian* meningkatkan presisi, sedangkan *Complement Naïve Bayes* mengoptimalkan penangkapan kandidat izin yang layak. Temuan ini membuktikan bahwa varian *Naïve Bayes* teroptimasi menyediakan model yang efisien dan transparan untuk sistem pendukung keputusan otomatis dalam perizinan stasiun radio.

Kata Kunci: Complement Naïve Bayes; Izin Frekuensi Radio; Gaussian Naïve Bayes; Naïve Bayes; Klasifikasi Perizinan Spektrum

Abstract−Efficient management of the radio frequency spectrum requires a fast and accurate license evaluation system to avoid signal interference and ensure regulatory compliance. High submission volumes often cause processing bottlenecks in public licensing services. This study aims to optimize the classification of Radio Station License (ISR) applications into Granted and Rejected classes using variants of the Naïve Bayes algorithm. A real-world dataset comprising 38,954 application records with 26 technical, administrative, and geographic features was used. Data preprocessing involved removing post-decision leakage features, median imputation, label encoding, and feature scaling/transformation. Three algorithm variants were evaluated: Standard Gaussian Naïve Bayes, Gaussian Naïve Bayes with PowerTransformer, and Complement Naïve Bayes. Empirical results show that Gaussian Naïve Bayes with PowerTransformer achieved the highest overall accuracy of 78.32% and a high precision of 88.63% (5,112 True Negatives and 990 True Positives). Conversely, Complement Naïve Bayes effectively handled class imbalance by achieving a significantly higher recall of 73.71% and an F1-Score of 64.21%. In conclusion, transforming non-Gaussian technical features improves precision, while Complement Naïve Bayes optimizes candidate identification for approval. These findings demonstrate that optimized Naïve Bayes variants provide an effective and transparent model for automated decision support systems in spectrum licensing.

Keywords: Complement Naïve Bayes; Gaussian Naïve Bayes; Naïve Bayes; Radio Station License; Spectrum Licensing Classification

1. PENDAHULUAN

Spektrum frekuensi radio merupakan sumber daya alam terbatas yang memiliki peran sangat strategis dalam menopang seluruh infrastruktur telekomunikasi modern serta ekosistem digital nasional [1]. Penggunaan spektrum frekuensi diatur secara ketat oleh regulasi pemerintah guna memastikan tidak terjadinya gangguan atau interferensi sinyal antar pemancar radio yang berpotensi merusak kualitas layanan komunikasi [2],[3]. Setiap entitas penyelenggara yang mengoperasikan stasiun pemancar wajib memiliki Izin Stasiun Radio (ISR) yang diterbitkan melalui mekanisme verifikasi komprehensif terhadap parameter teknis, administratif, dan batas geografis lokasi pemancar. Seiring dengan pertumbuhan pesat teknologi komunikasi nirkabel dan ekspansi jaringan seluler, volume permohonan izin frekuensi radio mengalami peningkatan yang sangat signifikan setiap tahunnya. Kondisi ini memicu munculnya antrean panjang (*bottleneck*) pada sistem pelayanan publik terpadu, mengingat verifikasi kualifikasi berkas yang dilakukan secara manual oleh verifikator manusia membutuhkan waktu, kecermatan, dan ketelitian yang sangat tinggi. Proses evaluasi manual yang lambat tidak hanya menghambat efisiensi operasional regulator, tetapi juga merugikan pihak pelaku usaha yang membutuhkan kepastian izin secara cepat. Oleh karena itu, dibutuhkan sebuah solusi transformatif berbasis teknologi cerdas, yaitu pembangunan Sistem Pendukung Keputusan (*Decision Support System*) berbasis *machine learning* yang mampu mengklasifikasikan kelayakan berkas perizinan secara otomatis, presisi, dan transparan guna mempercepat alur pelayanan publik [4].

Penerapan metode *machine learning* dan *data mining* telah menjadi salah satu pendekatan paling efektif untuk membangun model otomatisasi penilaian risiko serta persetujuan dokumen perizinan [4]. Dalam berbagai penelitian klasifikasi biner, algoritma *Naïve Bayes* terbukti menjadi salah satu metode terpopuler berkat kompleksitas komputasinya yang rendah, kecepatan pelatihan data yang sangat tinggi, serta sifat prediksinya yang transparan berbasis kalkulasi probabilitas posterior [5],[6]. Beberapa studi terdahulu telah mencoba menerapkan algoritma ini dalam beragam domain klasifikasi. Sebagai contoh, Yoga et al. [7] menerapkan algoritma *Naïve Bayes* dan *Simple Kriging* untuk pemetaan klasifikasi curah hujan. Selanjutnya, Thoriq et al. [8] serta Hidayat et al. [9] menggunakan *Naïve Bayes* untuk memprediksi dan mengklasifikasikan kelayakan penerimaan beasiswa KIP Kuliah, namun studi tersebut mengabaikan ketidakseimbangan proporsi kelas target (*class imbalance*) yang menyebabkan penurunan performa pada kelas minoritas.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

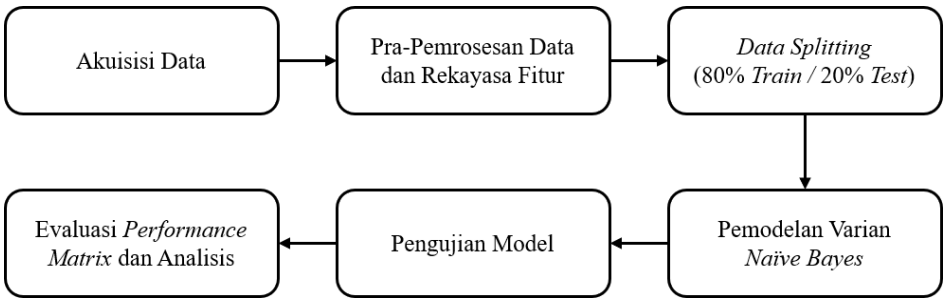
Sementara itu, Priyanto et al. [10] mengimplementasikan *Naïve Bayes* untuk penentuan penerima Bantuan Program Indonesia Pintar, di mana hasilnya menunjukkan bahwa asumsi independensi antar-fitur kerap menurunkan presisi pada data yang saling berkorelasi. Pada studi klasifikasi risiko kegagalan jantung, Subarkah et al. [11] membandingkan *Naïve Bayes* standar dengan teknik korelasi fitur dan menemukan bahwa seleksi atribut dapat meningkatkan akurasi, tetapi belum mengatasi masalah kemiringan (*skewness*) data kontinu. Adapun Alita et al. [12] serta Damari et al. [13] memanfaatkan *Naïve Bayes Classifier* dan penanganan *class imbalance* untuk pendukung keputusan dan mencatat bahwa ketidakseimbangan kelas memicu bias prediksi yang signifikan terhadap kelas mayoritas. Lebih lanjut, Binalla dan Villanueva [14] mencoba melakukan peningkatan algoritma *Gaussian Naïve Bayes* pada klasifikasi kualitas udara dengan mentransformasi data kontinu *skewed*, tetapi ruang lingkup variabelnya masih terbatas pada data parameter lingkungan sederhana. Di sisi lain, studi terbaru oleh Herminarimawati dan Kusrini [15] serta Siswanto et al. [16] mengevaluasi varian *Naïve Bayes* dan optimasi seleksi fitur untuk mengklasifikasikan data berdimensi tinggi, namun belum diuji secara empiris pada data perizinan publik terintegrasi dalam skala besar dengan varian *Complement Naïve Bayes*.

Berdasarkan penelusuran *state of the art* dari penelitian-penelitian terdahulu tersebut, terlihat adanya *gap analysis* yang nyata ketika algoritma *Naïve Bayes* diterapkan pada domain perizinan spektrum frekuensi radio. Parameter teknis frekuensi radio seperti daya pemancar (*Effective Isotropically Radiated Power* / EIRP), lebar pita (*bandwidth*), *gain* antena, dan ketinggian struktur memiliki karakteristik data kontinu numerik yang berdistribusi sangat miring (*skewed*) dan tidak berdistribusi normal (*non-Gaussian*) [17],[14]. Penerapan *Gaussian Naïve Bayes* standar yang mengasumsikan distribusi normal pada data teknis tersebut kerap menghasilkan estimasi probabilitas *likelihood* yang tidak akurat. Selain itu, dataset riil perizinan spektrum frekuensi radio secara inheren mengalami ketidakseimbangan kelas (*class imbalance*) yang tajam, di mana jumlah permohonan ditolak (*Rejected*) jauh lebih dominan dibandingkan dengan permohonan yang diterima (*Granted*) [18]. Keterbatasan ini menyebabkan model *Naïve Bayes* standar cenderung mengalami bias terhadap kelas mayoritas, yang berdampak pada rendahnya tingkat *recall* dalam mengidentifikasi berkas permohonan yang sebenarnya layak izin. Hingga saat ini, belum ada penelitian yang melakukan studi komparatif secara mendalam mengenai optimasi varian *Naïve Bayes* khususnya penggabungan transformasi fitur *PowerTransformer* (*Yeo-Johnson*) dan penggunaan *Complement Naïve Bayes* dalam mengatasi kombinasi masalah distribusi *non-Gaussian* serta *class imbalance* pada dataset perizinan frekuensi radio skala besar.

Guna mengisi celah penelitian (*research gap*) tersebut, penelitian ini mengajukan pendekatan optimasi klasifikasi kelayakan perizinan frekuensi radio dengan membandingkan secara empiris tiga varian algoritma *Naïve Bayes*, yaitu *Gaussian Naïve Bayes* Standar, *Gaussian Naïve Bayes* yang dikombinasikan dengan *PowerTransformer* (*Yeo-Johnson*), serta *Complement Naïve Bayes*. Kebaruan (*novelty*) utama dari riset ini terletak pada pemodelan dataset riil perizinan spektrum frekuensi radio skala besar yang melibatkan 38.954 rekor data historis permohonan dengan 26 atribut prediktor komprehensif (teknis, administrasi, dan geografis) setelah mengeliminasi fitur kebocoran pasca-keputusan (*post-decision data leakage*). Riset ini mengevaluasi secara ketat dampak transformasi fitur *non-Gaussian* terhadap peningkatan presisi serta efektivitas *Complement Naïve Bayes* dalam mengeliminasi bias kelas mayoritas. Selain itu, kontribusi utama dari penelitian ini mencakup aspek teoretis berupa pembuktian empiris efektivitas penanganan data teknis *skewed* dan *class imbalance* tanpa manipulasi data sintetis, serta aspek praktis berupa formulasi skenario operasional *Strict Compliance* dan *Fast Screening* bagi instansi regulator. Tujuan utama dari penelitian ini adalah untuk menghasilkan model klasifikasi otomatis yang memiliki keseimbangan metrik *precision*, *recall*, dan *F1-score* [19]. Harapan yang ingin dicapai dari hasil penelitian ini adalah terbentuknya sebuah model pendukung keputusan (*decision support system*) yang transparan, efisien, dan andal untuk diimplementasikan sebagai modul *screening* otomatis pada instansi pembina spektrum frekuensi radio, sehingga mampu mereduksi antrean verifikasi manual tanpa mengabaikan ketelitian regulasi dan kepatuhan hukum.

2. METODOLOGI PENELITIAN

Rancangan penelitian eksperimental ini disusun secara sistematis melalui enam tahapan utama yang saling berurutan guna mengoptimalkan kinerja klasifikasi kelayakan perizinan frekuensi radio. Tahapan tersebut mencakup akuisisi data, pra-pemrosesan data dan rekayasa fitur (*data preprocessing* dan *feature engineering*), pembagian dataset (*data splitting*), pemodelan varian *Naïve Bayes*, pengujian model pada data uji independen, serta evaluasi metrik performa. Prosedur penelitian secara menyeluruh ditunjukkan pada gambar 1.



Gambar 1. Diagram Alir Prosedur Penelitian

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

2.1 Tahapan Penelitian

Prosedur penelitian diawali dengan pengumpulan rekor historis pengajuan Izin Stasiun Radio (ISR) yang kemudian diproses melalui serangkaian pembersihan data untuk menghilangkan identitas unik dan fitur kebocoran pasca-keputusan (*post-decision data leakage*). Selanjutnya, dilakukan rekayasa fitur berupa imputasi median, *label encoding*, serta transformasi skala numerik. Dataset yang telah bersih dibagi menjadi data pelatihan (*training*) dan data pengujian (*testing*) menggunakan metode *Stratified Random Sampling*. Data pelatihan dimanfaatkan untuk membentuk tabel probabilitas *prior* dan *likelihood* pada tiga varian algoritma *Naïve Bayes*, yaitu *Gaussian Naïve Bayes* Standar, *Gaussian Naïve Bayes* dengan *PowerTransformer*, serta *Complement Naïve Bayes*. Ketiga model yang telah dilatih kemudian diuji menggunakan data *testing* yang belum pernah dilihat (*unseen data*) untuk menghasilkan estimasi kelas prediksi. Akhirnya, performa setiap varian dievaluasi berdasarkan *Confusion Matrix* dan metrik kuantitatif seperti akurasi, presisi, *recall*, serta *F1-Score* guna mengidentifikasi model pendukung keputusan yang paling optimal.

2.2 Akuisisi Data

Data yang digunakan dalam penelitian ini merupakan dataset sekunder riil berupa rekor data historis pengajuan Izin Stasiun Radio (ISR) yang diperoleh dari instansi pembina spektrum frekuensi radio. Dataset mentah terdiri dari 38.954 rekor data permohonan dengan 81 atribut mentah yang mencakup parameter teknis radio, identitas perangkat, batas geografis, serta variabel administrasi perizinan. Proporsi kelas target terdiri dari dua kategori utama:

a. Data permohonan dengan status Diterima (*Granted*) sebanyak 12.760 rekor data (32,76%).
b. Data permohonan dengan status Ditolak (*Rejected*) sebanyak 26.194 rekor data (67,24%).

2.3 Pra-Pemrosesan Data dan Rekayasa Fitur

Tahap pra-pemrosesan data dan rekayasa fitur dilakukan melalui empat tahapan utama untuk menjamin integritas dan kualitas data sebelum dimasukkan ke dalam model klasifikasi:

a. Pembersihan Data (*Data Cleaning*) dan Eliminasi Identitas: Menghapus atribut identitas unik yang tidak memiliki nilai prediktif, seperti CLNT_ID, NO_SIMF, APPL_ID, dan SITE_ID.
b. Eliminasi Kebocoran Data (*Post-Decision Data Leakage*): Menghapus 8 atribut yang nilainya baru terisi setelah izin diterbitkan yaitu CURR_LIC_NUM, LICENCE_DATE, VALIDITY_DATE, RENEWAL_DATE, ARCHIVE_DATE, MASA_LAKU, UMUR_ISR, dan BHP. Atribut AP_REQUEST_TYPE juga dieliminasi untuk menghindari bias korelasi instan akibat artefak sistemik ekspor data.
c. Seleksi Fitur Prediktor: Menghasilkan matriks akhir sebanyak 26 atribut prediktor (15 atribut numerik kontinu dan 11 atribut kategorikal) sebagaimana dirangkum pada Tabel 1.
d. Imputasi dan *Encoding*:
    1. Nilai kosong (*missing values*) pada atribut numerik diimputasi menggunakan nilai median masing-masing kolom untuk menjaga ketahanan dari pencilan (*outliers*).
    2. Variabel kategorikal diubah menjadi representasi numerik integer menggunakan teknik *Label Encoding*.
    3. Transformasi skala (*scaling*) menggunakan *PowerTransformer* (metode Yeo-Johnson) dan *MinMaxScaler* diterapkan pada variabel kontinu numerik untuk menstabilkan varians dan meminimalkan kemiringan (*skewness*) distribusi data teknis *non-Gaussian*.

Tabel 1. Tipe Data Dataset Perizinan Frekuensi Radio

Kode Atribut	Nama Atribut	Tipe Data	Deskripsi Operasional
X_1	FREQ	Numerik	Frekuensi pemancar (MHz)
X_2	FREQ_PAIR	Numerik	Pasangan frekuensi penerima (MHz)
X_3	ERP_PWR_DBM	Numerik	Daya pancar efektif / EIRP (dBm)
X_4	EQ_PWR	Numerik	Daya keluaran perangkat (Watt)
X_5	GAIN	Numerik	Penguatan antena (dBi)
X_6	LOSS	Numerik	Redaman saluran transmisi (dB)
X_7	BWIDTH	Numerik	Lebar pita frekuensi (kHz)
X_8	HGT_ANT	Numerik	Ketinggian struktur antena (m)
X_9	CIRCUIT_LEN	Numerik	Jarak lintasan komunikasi (km)
X_10	H_ASL	Numerik	Ketinggian lokasi di atas permukaan laut (m)
X_11-X_15	IB, IP, HDDP, HDLP, NUM_SETS	Numerik	Parameter teknis pendukung dan jumlah set
X_16-X_20	SERVICE, SUBSERVICE, TRANS_TYPE, ZONA, NAMA_UPT	Kategorikal	Jenis layanan, zona, dan UPT pengelola
X_21-X_26	EQ_MFR, ANT_MFR, PROVINCE, CITY, DISTRICT, MASTER_PLZN_CODE	Kategorikal	Merk perangkat, antena, & lokasi geografis
Y	STATUS_SIMF	Biner (Target)	Kelas Target: 1 = Granted, 0 = Rejected

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

2.4 Pembagian Dataset (Data *Splitting*)

Dataset yang telah bersih dan terenkode dibagi menjadi dua bagian independen menggunakan metode *Stratified Random Sampling* [20]:

a. Data *Training* (80%): Berjumlah 31.163 rekor data, digunakan untuk melatih dan membentuk tabel probabilitas *prior* serta *likelihood* model.
b. Data *Testing* (20%): Berjumlah 7.791 rekor data (5.239 rekor *Rejected* dan 2.552 rekor *Granted*), digunakan murni untuk pengujian dan validasi performa model.

Penggunaan teknik *stratified sampling* bertujuan menjaga agar rasio proporsi kelas target (32,76% *Granted* : 67,24% *Rejected*) tetap konsisten pada data training dan data testing, sehingga mencegah timbulnya *sampling bias*.

2.5 Pemodelan Varian *Naïve Bayes*

Teorema *Naïve Bayes* mengkalkulasi probabilitas posterior $P(Y \mid X)$ untuk menentukan kelas target $Y \in \{0,1\}$ berdasarkan kombinasi vektor atribut $X = (x_1, x_2, \dots, x_n)$ melalui persamaan (1):

$$P(Y \mid X) = \frac{P(Y)\prod_{i=1}^{n} P(x_i|Y)}{P(X)} \tag{1}$$

Tiga varian algoritma *Naïve Bayes* diterapkan untuk mengevaluasi pendekatan penanganan distribusi data yang berbeda:

a. *Gaussian Naïve Bayes* (GNB) Standar: Mengasumsikan bahwa seluruh atribut numerik kontinu mengikuti distribusi normal (*Gaussian*) dengan fungsi densitas probabilitas pada persamaan (2) [21],[22]:

$$P(x_i|Y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(x_i-\mu_Y)^2}{2\sigma_Y^2}\right) \tag{2}$$

di mana μ_Y adalah rata-rata (*mean*) dan σ_Y^2 adalah varians atribut x_i pada kelas Y.

b. *Gaussian Naïve Bayes* + *PowerTransformer* (GNB-PT): Menerapkan transformasi Yeo-Johnson pada atribut numerik sebelum pemodelan GNB untuk mendekatkan distribusi fitur yang miring (*skewed*) menjadi terdistribusi mendekati normal.
c. *Complement Naïve Bayes* (CNB): Varian yang dirancang khusus untuk mengatasi *class imbalance* dengan menghitung bobot probabilitas dari komplemen kelas target $\bar{c}$ melalui persamaan (3):

$$w_{ci} = \log \frac{\sum_{j\in\bar{c}} f_{ji} + \alpha}{\sum_{i\prime}\sum_{j\in\bar{c}} f_{ji\prime} + \alpha N} \tag{3}$$

di mana w_{ci} menyatakan bobot ter-sisa untuk atribut i pada kelas c, f_{ji} adalah nilai/frekuensi fitur i pada sampel j yang berada di luar kelas c ($j \in \bar{c}$), α adalah *smoothing parameter*, dan N adalah total jumlah atribut.

2.6 Pengujian Model

Pada tahap ini, tiga varian model *Naïve Bayes* yang telah dilatih menggunakan 80% data *training* dievaluasi kinerjanya menggunakan 20% data pengujian (*validation test set*) sebanyak 7.791 rekor data yang belum pernah dilihat (*unseen data*) oleh model. Proses pengujian dilakukan dengan memasukkan vektor atribut prediktor data *testing* X_{test} ke dalam masing-masing model untuk mengkalkulasi nilai probabilitas posterior $P(Y = c|X_{test})$. Hasil estimasi probabilitas tersebut kemudian dikonversi menjadi label kelas prediksi $Y \in \{0,1\}$, di mana 1 merepresentasikan *Granted* dan 0 merepresentasikan *Rejected*. Tahap pengujian ini berfungsi untuk mengukur kemampuan generalisasi model secara objektif tanpa terjadinya kebocoran data (*data leakage*), serta menghasilkan matriks luaran berupa pasangan label prediksi $\hat{Y}$ dan label aktual Y yang siap dianalisis pada tahap evaluasi performa.

2.7 Evaluasi *Performance Metrics* dan Analisis

Evaluasi dilakukan dengan memprediksi status kelayakan 7.791 rekor data *testing* independen. Hasil prediksi dibandingkan dengan label riil menggunakan *Confusion Matrix* sebagaimana ditunjukkan pada Tabel 2 [19].

Tabel 2. *Confusion Matrix*

Status Aktual	Prediksi Diterima ($Y = 1$)	Prediksi Ditolak ($Y = 0$)
Aktual Diterima ($Y = 1$)	*True Positive* (TP)	*False Negative* (FN)
Aktual Ditolak ($Y = 0$)	*False Positive* (FP)	*True Negative* (TN)

Metrik evaluasi kinerja komparatif dihitung berdasarkan empat indikator kuantitatif utama [19]:

a. Akurasi (*Overall Accuracy*): Rasio total prediksi benar dari seluruh data *testing*.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{4}$$

b. Presisi (*Precision* - Kelas *Granted*): Tingkat ketepatan prediksi saat model mengklasifikasikan izin diterima.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

c. *Recall* / Sensitivitas (*Granted*): Kemampuan model dalam mengidentifikasi seluruh permohonan yang aktualnya layak diterima.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

d. *F1-Score*: Nilai rata-rata harmonis antara presisi dan *recall*.

$$F1\text{-}Score = 2 \times \frac{\text{Presisi} \times Recall}{\text{Presisi} + Recall} \quad (7)$$

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen dan pembahasan komprehensif dari klasifikasi kelayakan permohonan Izin Stasiun Radio (ISR) menggunakan tiga varian algoritma *Naïve Bayes*, yaitu *Gaussian Naïve Bayes* Standar, *Gaussian Naïve Bayes* dengan *PowerTransformer*, serta *Complement Naïve Bayes*. Seluruh tahapan eksperimen diimplementasikan menggunakan bahasa pemrograman Python dengan pustaka *scikit-learn*, *pandas*, *numpy*, dan *seaborn*. Pemodelan dievaluasi secara objektif terhadap dataset pengujian (*testing set*) independen sebanyak 7.791 rekor data guna menguji ketepatan model dalam mengklasifikasikan permohonan ke dalam kelas Diterima (*Granted*) dan Ditolak (*Rejected*). Pemaparan bagian ini dibagi secara sistematis ke dalam tiga sub-bab utama: sub-bab 3.1 menguraikan penerapan metode dan pra-pemrosesan data riil, sub-bab 3.2 menyajikan hasil implementasi dan pengujian model, serta sub-bab 3.3 menyajikan pembahasan kritis, perbandingan dengan penelitian terdahulu, serta implikasi operasionalnya.

3.1 Penerapan Metode dan Eksekusi *Preprocessing* Data

Penerapan metode klasifikasi pada penelitian ini diawali dengan mengolah dataset sekunder riil hasil penggabungan dua berkas laporan utama perizinan spektrum frekuensi radio, yaitu dataset permohonan status *Granted* sebanyak 12.760 rekor data (32,76%) dan dataset permohonan status *Rejected* sebanyak 26.194 rekor data (67,24%), sehingga menghasilkan total dataset gabungan sebanyak 38.954 rekor data dengan 81 atribut mentah. Struktur awal dataset gabungan sebelum pembersihan disajikan pada Tabel 3. Sebelum data dimasukkan ke dalam algoritma klasifikasi, dilakukan serangkaian prosedur pra-pemrosesan data dan rekayasa fitur (*data preprocessing* dan *feature engineering*) secara sistematis untuk menjamin integritas data serta menghilangkan bias sistemik. Langkah-langkah teknis pra-pemrosesan data dan penerapan metode dijalankan melalui urutan prosedur sebagai berikut:

Tabel 3. Dataset Perizinan Frekuensi Radio

	CLNT_ID	NO_SIMF	APPL_ID	SITE_ID	CLNT_NAME	...	SA_SAT_NAME	SA_COMMENT	NAMA_UPT
0	*****	*****	*****	*****	TELEKOMUNIKASI SELULAR, PT.	...	NaN	NaN	BALMON KELAS II LAMPUNG
1	*****	*****	*****	*****	TELEKOMUNIKASI SELULAR, PT.	...	NaN	NaN	BALMON KELAS II LAMPUNG
2	*****	*****	*****	*****	SINAR LAUT LAMPUNG PERMAI, PT.	...	NaN	NaN	BALMON KELAS II LAMPUNG
...	...	...	...	...	...	...	...	...	...
38951	*****	*****	*****	*****	SAMPOERNA TELEKOMUNIKASI INDONESIA,PT	...	NaN	NaN	BALMON KELAS II LAMPUNG
38952	*****	*****	*****	*****	SAMPOERNA TELEKOMUNIKASI INDONESIA,PT	...	NaN	NaN	BALMON KELAS II LAMPUNG
38953	*****	*****	*****	*****	SAMPOERNA TELEKOMUNIKASI INDONESIA,PT	...	NaN	NaN	BALMON KELAS II LAMPUNG

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

a. Pembersihan Data dan Eliminasi Kebocoran Pasca-Keputusan (*Post-Decision Data Leakage*) Tahap awal pembersihan data difokuskan pada eliminasi atribut yang tidak memiliki nilai prediktif serta atribut yang berpotensi menyebabkan kebocoran data.
   1. Eliminasi Atribut Identitas Unik: Atribut seperti CLNT_ID, NO_SIMF, APPL_ID, dan SITE_ID dihapus karena hanya berfungsi sebagai kunci indeks identitas pendaftaran yang tidak berhubungan dengan kualifikasi teknis perizinan, sebagaimana dirangkum pada Tabel 4.

Tabel 4. Dataset Hasil Eliminasi Atribut Identitas Unik

	CLNT_NAME	STATUS_SIMF	AP_REQUEST_TYPE	SERVICE	SUBSERVICE	...	SA_SAT_NAME	SA_COMMENT	NAMA_UPT
0	TELEKOMUNIKASI SELULAR, PT.	Granted	N	Land Mobile (public)	LM Registered Stations	...	NaN	NaN	BALMON KELAS II LAMPUNG
1	TELEKOMUNIKASI SELULAR, PT.	Granted	N	Land Mobile (public)	LM Registered Stations	...	NaN	NaN	BALMON KELAS II LAMPUNG
2	SINAR LAUT LAMPUNG PERMAI, PT.	Granted	N	Land Mobile (private)	Standard	...	NaN	NaN	BALMON KELAS II LAMPUNG
...	...	...	...	...	...	...	...	...	...
38951	SAMPOERNA TELEKOMUNIKASI INDONESIA,PT	Rejected	A	Fixed Service	PP	...	NaN	NaN	BALMON KELAS II LAMPUNG
38952	SAMPOERNA TELEKOMUNIKASI INDONESIA,PT	Rejected	A	Fixed Service	PP	...	NaN	NaN	BALMON KELAS II LAMPUNG
38953	SAMPOERNA TELEKOMUNIKASI INDONESIA,PT	Rejected	A	Fixed Service	PP	...	NaN	NaN	BALMON KELAS II LAMPUNG

   2. Eliminasi Atribut Kebocoran Pasca-Keputusan: Delapan atribut yang nilainya baru terisi setelah izin diterbitkan oleh pihak regulator dihapus dari matriks fitur, meliputi CURR_LIC_NUM, LICENCE_DATE, VALIDITY_DATE, RENEWAL_DATE, ARCHIVE_DATE, MASA_LAKU, UMUR_ISR, dan BHP. Selain itu, atribut AP_REQUEST_TYPE dieliminasi untuk menghindari korelasi instan yang artifisial akibat mekanisme ekspor data historis (Tabel 5).

Tabel 5. Dataset Hasil Eliminasi Atribut Kebocoran Pasca-Keputusan

	FREQ	FREQ_PAIR	ERP_PWR_DBM	EQ_PWR	GAIN	...	MASTER_PLZN_CODE	NAMA_UPT	STATUS_SIMF
0	2127.5	1937.5	57.5303	20	17.52	...	V	BALMON KELAS II LAMPUNG	Granted
1	2132.5	1942.5	57.5303	20	17.52	...	V	BALMON KELAS II LAMPUNG	Granted

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

	FREQ	FREQ_ PAIR	ERP_PW R_DBM	EQ_PWR	GAIN	...	MASTER_ PLZN_CO DE	NAMA_UP T	STATUS_ SIMF
2	408.025	408.025	36.9897	5	0	...	V	BALMON KELAS II LAMPUNG	Granted
...	...	...	...	...	...	...	...	...	...
38951	7401	7240	61.1794	0.25	41.2	...	V	BALMON KELAS II LAMPUNG	Rejected
38952	7240	7401	61.1794	0.25	41.2	...	V	BALMON KELAS II LAMPUNG	Rejected
38953	7505	7666	63.2794	0.25	43.3	...	V	BALMON KELAS II LAMPUNG	Rejected

b. Imputasi Nilai Kosong (*Missing Values*) dan *Label Encoding* Setelah eliminasi fitur kebocoran data, dilakukan seleksi terhadap 26 atribut prediktor utama yang terdiri dari 15 atribut numerik kontinu dan 11 atribut kategorikal.

1. Imputasi *Median* Atribut Numerik: Nilai kosong pada 15 atribut numerik kontinu (seperti FREQ, FREQ_PAIR, ERP_PWR_DBM, EQ_PWR, GAIN, LOSS, BWIDTH, HGT_ANT, CIRCUIT_LEN, dan H_ASL) diisi menggunakan nilai *median* dari masing-masing atribut. Pemilihan nilai *median* dilakukan karena sifatnya yang tangguh (*robust*) terhadap keberadaan pencilan (*outliers*) pada data parameter teknis radio, dengan hasil akhir yang disajikan pada Tabel 6.

Tabel 6. Dataset Hasil Imputasi *Median* Atribut Numerik

	FREQ	FREQ_ PAIR	ERP_PW R_DBM	EQ_P WR	GAIN	LOSS	...	MASTER_ PLZN_CO DE	NAMA_UP T	STATU S_SIMF
0	2127.5	1937.5	57.5303	20	17.52	1	...	V	BALMON KELAS II LAMPUNG	Granted
1	2132.5	1942.5	57.5303	20	17.52	1	...	V	BALMON KELAS II LAMPUNG	Granted
2	408.025	408.025	36.9897	5	0	0	...	V	BALMON KELAS II LAMPUNG	Granted
...	...	...	...	...	...	...	...	...	...	...
38951	7401	7240	61.1794	0.25	41.2	4	...	V	BALMON KELAS II LAMPUNG	Rejected
38952	7240	7401	61.1794	0.25	41.2	4	...	V	BALMON KELAS II LAMPUNG	Rejected
38953	7505	7666	63.2794	0.25	43.3	4	...	V	BALMON KELAS II LAMPUNG	Rejected

2. *Label Encoding* Atribut Kategorikal: Sebanyak 11 atribut kategorikal (termasuk SERVICE, SUBSERVICE, TRANS_TYPE, ZONA, EQ_MFR, ANT_MFR, PROVINCE, CITY, DISTRICT, MASTER_PLZN_CODE, dan NAMA_UPT) diubah menjadi representasi numerik bertipe integer menggunakan teknik *Label Encoding* seperti ditunjukkan pada Tabel 7 agar dapat diproses oleh kalkulasi matematis algoritma *Naïve Bayes*.

Tabel 7. Dataset Hasil *Label Encoding* Atribut Kategorikal

	FREQ	FREQ_ PAIR	ERP_PW R_DBM	EQ_P WR	GAIN	LOSS	...	MASTER_ PLZN_CO DE	NAMA _UPT	STATU S_SIMF
0	2127.5	1937.5	57.5303	20	17.52	1	...	6	0	Granted
1	2132.5	1942.5	57.5303	20	17.52	1	...	6	0	Granted
2	408.025	408.025	36.9897	5	0	0	...	6	0	Granted
...	...	...	...	...	...	...	...	...	...	...
38951	7401	7240	61.1794	0.25	41.2	4	...	6	0	Rejected
38952	7240	7401	61.1794	0.25	41.2	4	...	6	0	Rejected
38953	7505	7666	63.2794	0.25	43.3	4	...	6	0	Rejected

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

c. Transformasi Skala Fitur *Non-Gaussian* dan Normalisasi Fitur numerik kontinu pada parameter teknis frekuensi radio memiliki karakteristik distribusi data yang miring (*skewed*) dan tidak berdistribusi normal.
   1. Transformasi *PowerTransformer* (*Yeo-Johnson*): Untuk varian *Gaussian Naïve Bayes* teroptimasi, dilakukan transformasi fitur numerik menggunakan metode *Yeo-Johnson* melalui *PowerTransformer*. Transformasi ini bertujuan untuk menstabilkan varians dan menggeser distribusi data yang *skewed* menjadi mendekati distribusi normal (*Gaussian*) sebagaimana dirangkum pada Tabel 8, sehingga asumsi dasar pada fungsi densitas probabilitas *Gaussian Naïve Bayes* dapat dipenuhi dengan lebih akurat.

Tabel 8. Dataset Hasil Transformasi *PowerTransformer* (*Yeo-Johnson*)

	FREQ	FREQ_PAIR	ERP_PWR_DBM	EQ_PWR	GAIN	LOSS	...	MASTER_PLZN_CODE	NAMA_UPT	STATUS_SIMF
0	-1.48328	-1.65548	-0.09908	2.06813	-1.98320	-0.02867	...	0.49016	-0.01133	Granted
1	-1.48192	-1.65383	-0.09908	2.06813	-1.98320	-0.02867	...	0.49016	-0.01133	Granted
2	-2.07353	-2.29167	-2.31685	1.86666	-2.42244	-1.85653	...	0.49016	-0.01133	Granted
...	...	...	...	...	...	...	...	...	...	...
38951	-0.37361	-0.35023	0.48750	-0.29727	0.92205	2.38843	...	0.49016	-0.01133	Rejected
38952	-0.40188	-0.31807	0.48750	-0.29727	0.92205	2.38843	...	0.49016	-0.01133	Rejected
38953	-0.35547	-0.26571	0.85499	-0.29727	1.34725	2.38843	...	0.49016	-0.01133	Rejected

   2. Normalisasi *MinMaxScaler*: Untuk varian *Complement Naïve Bayes*, seluruh matriks fitur yang telah terenkode ditransformasikan menggunakan *MinMaxScaler* ke dalam rentang nilai $[0,1]$. Transformasi skala ini diperlukan karena *Complement Naïve Bayes* membutuhkan masukan fitur numerik non-negatif dalam kalkulasi bobot probabilitas komplemennya (Tabel 9).

Tabel 9. Dataset Hasil Normalisasi *MinMaxScaler*

	FREQ	FREQ_PAIR	ERP_PWR_DBM	EQ_PWR	GAIN	LOSS	...	MASTER_PLZN_CODE	NAMA_UPT	STATUS_SIMF
0	0.05517	0.05837	0.60344	0.00100	0.35110	0.25000	...	0.66667	0.00000	Granted
1	0.05530	0.05852	0.60344	0.00100	0.35110	0.25000	...	0.66667	0.00000	Granted
2	0.01057	0.01219	0.38799	0.00025	0.00000	0.00000	...	0.66667	0.00000	Granted
...	...	...	...	...	...	...	...	...	...	...
38951	0.19195	0.21847	0.64171	0.00001	0.82565	1.00000	...	0.66667	0.00000	Rejected
38952	0.18777	0.22333	0.64171	0.00001	0.82565	1.00000	...	0.66667	0.00000	Rejected
38953	0.19464	0.23133	0.66374	0.00001	0.86774	1.00000	...	0.66667	0.00000	Rejected

Setelah tahap pra-pemrosesan data selesai, dataset dikonfigurasi ke dalam tiga varian algoritma *Naïve Bayes* untuk mengevaluasi pendekatan penanganan distribusi data dan ketidakseimbangan kelas yang berbeda:

a. *Gaussian Naïve Bayes* Standar: Diterapkan pada data yang ditransformasi menggunakan *StandardScaler* tanpa penanganan khusus terhadap kemiringan distribusi data kontinu *non-Gaussian*.
b. *Gaussian Naïve Bayes + PowerTransformer*: Diterapkan pada data yang telah ditransformasi menggunakan metode *Yeo-Johnson* untuk menguji sejauh mana perbaikan distribusi fitur dapat meningkatkan presisi klasifikasi.
c. *Complement Naïve Bayes*: Diterapkan pada data ter-*scale MinMaxScaler* dengan memanfaatkan estimasi probabilitas dari komplemen kelas target guna mengeliminasi bias terhadap kelas mayoritas (*Rejected*).

3.2 Hasil Implementasi dan Pengujian Model

Model yang telah dilatih menggunakan 80% data pelatihan (31.163 rekor data) diuji kinerjanya menggunakan 20% data pengujian independen (*validation test set*) sebanyak 7.791 rekor data yang belum pernah dilihat oleh model (*unseen data*). Data uji ini terdiri dari 5.239 rekor data dengan status aktual *Rejected* (67,24%) dan 2.552 rekor data dengan status aktual *Granted* (32,76%).

Hasil pengujian dari ketiga varian *Naïve Bayes* dievaluasi berdasarkan empat metrik kinerja utama: Akurasi (*Overall Accuracy*), Presisi (*Precision*), *Recall* (Sensitivitas), dan *F1-Score* pada kelas target *Granted* ($Y = 1$). Ringkasan rekapitulasi evaluasi performa kuantitatif disajikan pada tabel 10, sedangkan rincian nilai *Confusion Matrix* (*True Negative*, *False Positive*, *False Negative*, dan *True Positive*) dirangkum pada tabel 11 dan visualisasinya ditampilkan pada Gambar 2.

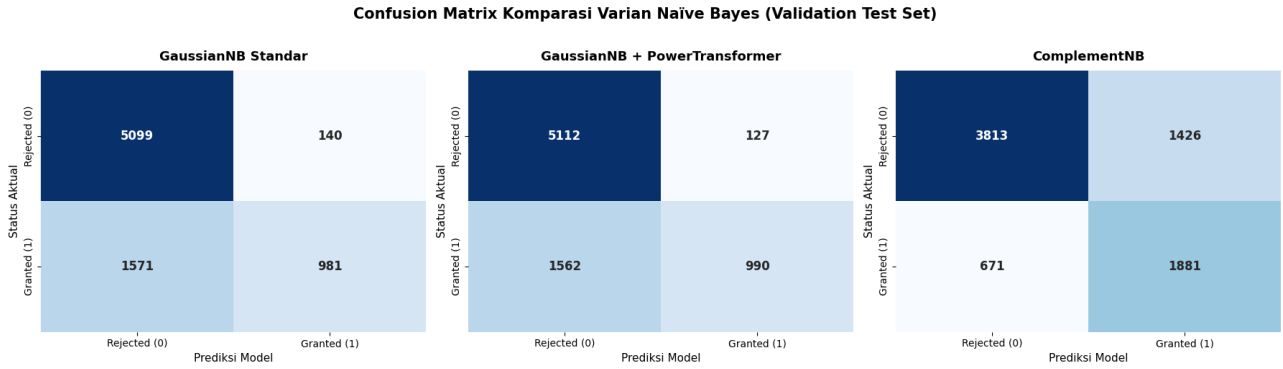
BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1954-1964
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1297

Tabel 10. Hasil Evaluasi Performa Varian Algoritma *Naïve Bayes*

Varian Model	Akurasi (%)	Presisi (%)	*Recall* (%)	*F1-Score* (%)
Gaussian Naïve Bayes Standar	78,04	87,51	38,44	53,42
Gaussian Naïve Bayes + PowerTransformer	78,32	88,63	38,79	53,97
Complement Naïve Bayes	73,08	56,88	73,71	64,21

Tabel 11. *Confusion Matrix* Pengujian (N = 7.791)

Varian Model	*True Negative* (TN)	*False Positive* (FP)	*False Negative* (FN)	*True Positive* (TP)
Gaussian Naïve Bayes Standar	5.099	140	1.571	981
Gaussian Naïve Bayes + PowerTransformer	5.112	127	1.562	990
Complement Naïve Bayes	3.813	1.426	671	1.881



Gambar 2. Hasil Evaluasi *Confusion Matrix* Ketiga Model pada Data Uji

Gambar 2 menunjukkan peta matriks konfusi (*Confusion Matrix*) komparatif untuk ketiga varian model pada dataset validasi independen. Berdasarkan hasil uji numerik pada tabel 10 dan tabel 11, ketiga varian *Naïve Bayes* memperlihatkan karakteristik performa yang berbeda secara signifikan sesuai dengan teknik *preprocessing* dan formulasi algoritma yang diterapkan.

Model *Gaussian Naïve Bayes* Standar menghasilkan akurasi keseluruhan sebesar 78,04%, presisi sebesar 87,51%, *recall* sebesar 38,44%, dan *F1-Score* sebesar 53,42%. Dari total 7.791 data uji, model ini berhasil memprediksi 5.099 data *True Negative* (TN) dan 981 data *True Positive* (TP). Namun, model ini menghasilkan *False Negative* (FN) yang cukup tinggi yaitu sebanyak 1.571 berkas permohonan yang layak izin tetapi diprediksi ditolak.

Pengombinasian *Gaussian Naïve Bayes* dengan transformasi *PowerTransformer* (*Yeo-Johnson*) berhasil meningkatkan kinerja klasifikasi di seluruh metrik evaluasi. Varian ini mencapai akurasi tertinggi sebesar 78,32%, presisi terunggul sebesar 88,63%, *recall* sebesar 38,79%, dan *F1-Score* sebesar 53,97%. Dibandingkan dengan GNB Standar, penerapan *PowerTransformer* mampu meningkatkan *True Negative* dari 5.099 menjadi 5.112 berkas dan *True Positive* dari 981 menjadi 990 berkas, sekaligus menekan jumlah kesalahan *False Positive* dari 140 menjadi hanya 127 berkas.

Varian *Complement Naïve Bayes* memberikan pergeseran karakteristik performa yang sangat kontras. Meskipun akurasi keseluruhannya sedikit menurun menjadi 73,08% dan presisinya menjadi 56,88%, CNB berhasil meningkatkan nilai *recall* secara drastis dari 38,79% menjadi 73,71% serta meningkatkan *F1-Score* keseluruhan menjadi 64,21%. CNB secara sukses mengidentifikasi 1.881 berkas permohonan yang layak diterima (*True Positive*) dari total 2.552 berkas aktual *Granted*, yang berarti menekan jumlah kesalahan *False Negative* dari 1.562 menjadi hanya 671 berkas.

3.3 Pembahasan Kritis dan Analisis Komparatif

Hasil eksperimen yang diperoleh menunjukkan bahwa karakteristik distribusi fitur dan ketidakseimbangan proporsi kelas target memiliki pengaruh yang sangat dominan terhadap perilaku prediktif varian *Naïve Bayes* pada dataset perizinan frekuensi radio skala besar.

a. Pengaruh Transformasi Fitur *Non-Gaussian* terhadap Presisi Klasifikasi
   Keberhasilan varian *Gaussian Naïve Bayes + PowerTransformer* dalam meraih presisi tertinggi sebesar 88,63% membuktikan pentingnya penanganan kemiringan distribusi data kontinu numerik. Pada parameter teknis radio seperti daya pemancar (ERP_PWR_DBM), lebar pita (BWIDTH), *gain* antena (GAIN), dan ketinggian antena (HGT_ANT), nilai-nilai ekstrim dan distribusi miring (*skewed*) kerap memicu varians yang sangat besar pada estimasi kurva *Gaussian* standar. Ketika fitur-fitur teknis ini ditransformasikan menggunakan metode *Yeo-Johnson*, distribusi data menjadi lebih simetris dan mendekati kurva normal. Akibatnya, kalkulasi fungsi densitas probabilitas *likelihood* $P(x_i|Y)$ menjadi jauh lebih presisi. Dari sudut pandang keselamatan penerbitan izin, angka presisi 88,63% dengan

False Positive hanya 127 berkas (1,63% dari total data pengujian) mengindikasikan bahwa ketika model ini menyarankan status Diterima (*Granted*), tingkat kepastian kebenaran regulasinya sangat tinggi. Hal ini meminimalkan risiko lolosnya pemancar radio ilegal atau berkas yang cacat teknis yang berpotensi memicu interferensi sinyal radio fatal di lapangan.

b. Efektivitas *Complement Naïve Bayes* dalam Mengatasi *Class Imbalance*
Meskipun varian *Gaussian* memiliki presisi tinggi, nilai *recall*-nya yang rendah (38,79%) mengungkapkan adanya kelemahan mendasar berupa bias terhadap kelas mayoritas (*Rejected*). Karena kelas *Rejected* mencakup 67,24% dari total data, model *Gaussian* cenderung mengklasifikasikan data ambigu ke dalam kelas *Rejected* untuk memaksimalkan peluang probabilitas *prior*. Keterbatasan ini berhasil diatasi secara efektif oleh *Complement Naïve Bayes* (CNB). Algoritma CNB bekerja dengan mengkalkulasi bobot probabilitas dari komplemen kelas target (sampel di luar kelas c), sehingga estimasi parameter tidak didominasi oleh jumlah sampel kelas mayoritas. Hasilnya, CNB mampu melonjakkan nilai *recall* hingga 73,71% (meningkat sebesar 34,92% dibandingkan varian *Gaussian*). Kemampuan ini menjadikan CNB sangat peka dalam mengenali berkas-berkas permohonan yang sebenarnya telah memenuhi kualifikasi izin, sehingga mengurangi jumlah permohonan layak yang salah tertolak (*False Negative* berkurang dari 1.562 menjadi 671 berkas).

c. Perbandingan dengan Penelitian Terdahulu (*State of the Art Comparison*)
Temuan dalam penelitian ini memperkuat sekaligus melengkapi hasil-hasil studi terdahulu mengenai optimasi *Naïve Bayes* pada berbagai domain klasifikasi. Penelitian oleh Yoga et al. [7] pada klasifikasi curah hujan dan Thoriq et al. [8] pada penerimaan beasiswa mencatat tingkat performa yang memadai, namun kedua studi tersebut tidak mengevaluasi metrik *recall* dan *F1-Score* pada dataset yang mengalami ketidakseimbangan kelas. Penelitian Binalla dan Villanueva [14] membuktikan bahwa modifikasi distribusi pada *Gaussian Naïve Bayes* dapat meningkatkan akurasi data kontinu, sejalan dengan temuan riset ini di mana transformasi *PowerTransformer* terbukti menaikkan presisi hingga 88,63%. Lebih lanjut, hasil penerapan *Complement Naïve Bayes* dalam penelitian ini konsisten dengan temuan Siswanto et al. [16] serta Herminarimawati dan Kusrini [15], yang mengonfirmasi bahwa pendekatan komplementer penanganan varian *Naïve Bayes* sangat unggul dalam meningkatkan nilai *recall* dan *F1-Score* pada data kelas minoritas tanpa memerlukan teknik *resampling* buatan (seperti SMOTE [18]) yang berpotensi menimbulkan *noise* pada data teknis sensitif.

d. Implikasi Operasional pada Sistem Pendukung Keputusan (*Decision Support System*)
Perbedaan karakteristik antara varian *Gaussian Naïve Bayes + PowerTransformer* dan *Complement Naïve Bayes* memberikan implikasi strategis dalam implementasi Sistem Pendukung Keputusan (SPK) otomatis pada instansi pembina spektrum frekuensi radio:
   1. Skenario Kepatuhan Ketat (*Strict Compliance / High Precision Mode*): Jika instansi regulator mengutamakan prinsip kehati-hatian maksimal untuk memitigasi risiko interferensi sinyal radio, varian *Gaussian Naïve Bayes + PowerTransformer* merupakan pilihan model terbaik. Model ini berfungsi sebagai sistem verifikasi akhir yang menjamin bahwa setiap izin yang diterbitkan secara otomatis memiliki tingkat kepatuhan regulasi yang sangat tinggi.
   2. Skenario *Screening* Cepat (*Fast Screening / High Recall Mode*): Jika fokus utama instansi regulator adalah mengatasi antrean panjang pemrosesan berkas publik dengan mereduksi beban verifikasi manual verifikator manusia, maka varian *Complement Naïve Bayes* merupakan solusi yang paling optimal. Model ini dapat dipasang sebagai modul filter awal (*early-stage filter*) untuk secara cepat meloloskan 73,71% kandidat izin yang layak, sementara berkas sisanya dilarikan ke jalur pemeriksaan manual. Pendekatan hibrida ini terbukti mampu memangkas waktu pelayanan perizinan secara signifikan tanpa mengorbankan akurasi verifikasi.

4. KESIMPULAN

Penelitian ini secara komprehensif berhasil menjawab seluruh rumusan masalah mengenai efisiensi dan keakuratan sistem evaluasi perizinan stasiun pemancar berbasis pembelajaran mesin. Pertama, hambatan distribusi parameter teknis yang miring dan berdistribusi *non-Gaussian* pada data riil berhasil diatasi melalui pengombinasian *Gaussian Naïve Bayes* dengan transformasi *PowerTransformer* (Yeo-Johnson), yang terbukti menghasilkan akurasi keseluruhan sebesar 78,32% serta presisi tertinggi mencapai 88,63% dengan menekan angka kesalahan *False Positive* hingga 1,63% (127 berkas). Kedua, permasalahan ketidakseimbangan kelas (*class imbalance*) yang tajam antara permohonan disetujui dan ditolak diselesaikan secara efektif oleh *Complement Naïve Bayes*, di mana algoritma ini mampu melonjakkan nilai *recall* secara drastis dari 38,79% menjadi 73,71% serta mengoptimalkan *F1-score* hingga 64,21% tanpa memerlukan manipulasi data sintetis. Ketiga, ketersendatan alur pelayanan publik akibat tingginya volume antrean verifikasi manual berkas perizinan terjawab melalui formulasi arsitektur sistem pendukung keputusan dua skenario, yakni *Strict Compliance* untuk pemastian kepatuhan regulasi tingkat tinggi dan *Fast Screening* guna menyeleksi kandidat berkas layak izin secara instan. Kendati demikian, keterbatasan model yang masih mengandalkan asumsi independensi antar-fitur prediktor perlu ditindaklanjuti pada riset mendatang melalui pengembangan teknik seleksi atribut hibrida, penerapan metode *hybrid resampling*, serta pengujian algoritma *ensemble* seperti *Random Forest* atau *Gradient Boosting* demi mewujudkan otomatisasi evaluasi kelayakan perizinan spektrum skala nasional yang semakin adaptif, presisi, transparan, dan andal.

REFERENCES

[1] J. Manco, I. Dayoub, A. Nafkha, M. Alibakhshikenari, and H. Ben Thameur, "Spectrum Sensing Using Software Defined Radio for Cognitive Radio Networks: A Survey," *IEEE Access*, vol. 10, pp. 131887–131908, 2022, doi: 10.1109/ACCESS.2022.3229739.

[2] M. A. Hossain *et al.*, "Machine learning-based cooperative spectrum sensing in dynamic segmentation enabled cognitive radio vehicular network," *Energies*, vol. 14, no. 4, pp. 1–29, 2021, doi: 10.3390/en14041169.

[3] I. A. Bartsiokas, P. K. Gkonis, D. I. Kaklamani, and I. S. Venieris, "ML-Based Radio Resource Management in 5G and Beyond Networks: A Survey," *IEEE Access*, vol. 10, pp. 83507–83528, 2022, doi: 10.1109/ACCESS.2022.3196657.

[4] C. Wang and H. Yu, "Intelligent Assessment of Personal Credit Risk Based on Machine Learning," *Systems*, vol. 13, no. 2, p. 112, 2025, doi: 10.3390/systems13020112.

[5] T. Kim and J. S. Lee, "Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers," *IEEE Access*, vol. 10, pp. 22724–22736, 2022, doi: 10.1109/ACCESS.2022.3155231.

[6] Z. Ahmed, B. Issac, and S. Das, "Ok-NB: An Enhanced OPTICS and k-Naive Bayes Classifier for Imbalance Classification With Overlapping," *IEEE Access*, vol. 12, pp. 57458–57477, 2024, doi: 10.1109/ACCESS.2024.3391749.

[7] I. K. Yoga, S. S. Prasetyowati, and Y. Sibaroni, "Prediction and Mapping Rainfall Classification Using Naive Bayes and Simple Kriging," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 4, pp. 1244–1253, 2022, doi: 10.29100/jipi.v7i4.3264.

[8] M. Thoriq, F. Maulana, Y. S. Eirlangga, N. Hayati, and M. A. Madani, "Implementasi Algoritma Naïve Bayes dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa KIP di Universitas Adzkia," *J. FASILKOM*, vol. 15, no. 1, pp. 108–114, 2025, doi: 10.37859/jf.v15i1.8959.

[9] H. Hidayat, M. KH, I. Kusmanto, M. Kadir, and K. Kristian, "Klasifikasi Mahasiswa Calon Penerima Beasiswa KIP Menggunakan Algoritma Naive Bayes di Universitas Tomakaka Mamuju," *J. Fasilkom*, vol. 15, no. 3, pp. 456–464, 2025, doi: 10.37859/jf.v15i3.10784.

[10] I. Priyanto, E. M. Dewanti, T. Tundo, M. Nurdin, and R. Kasiono, "Penerapan Algoritma Metode Naïve Bayes Untuk Penentuan Penerimaan Bantuan Program Indonesia Pintar (PIP)," *J. Manajamen Inform. Jayakarta*, vol. 4, no. 2, pp. 162–174, 2024, doi: 10.52362/jmijayakarta.v4i2.1355.

[11] P. Subarkah, W. R. Damayanti, and R. A. Permana, "Comparison of Correlated Algorithm Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Classification of Heart Failure," *Ilk. J. Ilm.*, vol. 14, no. 2, pp. 1–6, 2022, doi: 10.33096/ilkom.v14i2.1148.120-125.

[12] D. Alita, I. Sari, A. Rahman Isnain, and Stywati, "Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa," *Jdmsi*, vol. 2, no. 1, pp. 17–23, 2021, doi: 10.33365/jdmsi.v2i1.1028.

[13] A. Damari, T. Azhima Yoga Siswa, and W. Joko Pranoto, "Implementation of the PSO-SMOTE Method on the Naive Bayes Algorithm to Address Class Imbalance in Landslide Disaster Data," *INOVTEK Polbeng - Seri Inform.*, vol. 10, no. 1, pp. 332–343, 2025, doi: 10.35314/7wcvrb72.

[14] M. C. Binalla and M. A. F. Villanueva, "An Enhancement of the Gaussian Naive Bayes Algorithm Applied to Air Quality Classification," *Int. J. Multidiscip. Res.*, vol. 6, no. 6, pp. 1–20, 2024, doi: 10.36948/ijfmr.2024.v06i06.33366.

[15] E. Herminarimawati and Kusrini, "Performance Comparison of Naive Bayes, PSO-Naive Bayes, and PCA-PSO-Naive Bayes for the Classification of Elementary School Students' Interests," *J. Pendidik. Inform. dan Sains*, vol. 15, no. 1, pp. 77–87, 2026, doi: 10.31571/saintek.v15i1.10914.

[16] Siswanto, I. Kurniawan, and S. A. Thamrin, "Naive Bayes Algorithm with Feature Selection Using Particle Swarm Optimization," *J. Varian*, vol. 7, no. 2, pp. 127–136, 2024, doi: 10.30812/varian.v7i2.2409.

[17] X. Zeng and E. Pinsky, "Stable Distribution Naive Bayes Achieves Higher Accuracy Than Traditional Naive Bayes Classification," *Int. J. Cybern. Informatics*, vol. 14, no. 2, pp. 71–86, 2025, doi: 10.5121/ijci.2025.140205.

[18] Z. Wang and Q. Liu, "Imbalanced Data Classification Method Based on LSSASMOTE," *IEEE Access*, vol. 11, pp. 32252–32260, 2023, doi: 10.1109/ACCESS.2023.3262460.

[19] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLoS One*, vol. 18, no. 10, pp. 1–27, 2023, doi: 10.1371/journal.pone.0291908.

[20] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, vol. 102, no. 2021. doi: 10.1007/978-1-0716-1418-1.

[21] P. Srisuradetchai, "Posterior averaging with Gaussian naive Bayes and the R package RandomGaussianNB for big-data classification," *Front. Big Data*, vol. 8, pp. 1–14, 2025, doi: 10.3389/fdata.2025.1706417.

[22] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. 2022. doi: 10.7551/mitpress/8291.003.0026.