

Model Klasifikasi Penerima Bantuan Pangan Non-Tunai (BPNT) Menggunakan Algoritma Naïve Bayes Berbasis Data Sosial Ekonomi

Gunadi^{1*}, Juliandri Saputra², Edi Purwanto³

^{1,2}Fakultas Teknik dan Informatika, Universitas Sains dan Teknologi Indonesia

³Fakultas Ilmu Komputer, Universitas Lancang Kuning

Email: ¹gunadi@usti.ac.id, ²juliandri.saputra@gmail.com, ³edipurwanto@unilak.ac.id

*Penulis Korespondensi

ABSTRAK

Program Bantuan Pangan Non-Tunai (BPNT) merupakan salah satu program pemerintah yang bertujuan meningkatkan kesejahteraan masyarakat melalui pemberian bantuan pangan kepada keluarga yang memenuhi kriteria tertentu. Permasalahan yang masih sering dihadapi dalam penyaluran BPNT adalah ketidaktepatan sasaran akibat proses seleksi penerima yang masih bergantung pada penilaian manual sehingga berpotensi menimbulkan kesalahan inklusi maupun eksklusi. Penelitian ini bertujuan menerapkan algoritma Naïve Bayes untuk menentukan kelayakan penerima BPNT berdasarkan data sosial ekonomi masyarakat. Data yang digunakan berasal dari basis data Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Kabupaten Rokan Hulu tahun 2024 sebanyak 2.055 data rumah tangga yang mencakup berbagai atribut demografi, sosial, dan ekonomi. Metode penelitian meliputi tahapan pengumpulan data, praproses data, pembentukan model klasifikasi menggunakan algoritma Naïve Bayes, serta evaluasi kinerja model menggunakan confusion matrix dan nilai akurasi. Hasil penelitian menunjukkan bahwa penerapan SMOTEN dan optimasi parameter pada algoritma Naïve Bayes mampu meningkatkan kemampuan model dalam mengenali kelas penerima BPNT yang memiliki jumlah data lebih sedikit. Peningkatan nilai recall dari 53,33% menjadi 60,00% menunjukkan bahwa model lebih sensitif dalam mengidentifikasi rumah tangga yang berpotensi layak menerima bantuan. Hasil penelitian menunjukkan bahwa model yang dihasilkan layak digunakan sebagai pendukung keputusan dalam proses identifikasi penerima BPNT berbasis data sosial ekonomi.

Article Info

Kata Kunci:

Naïve Bayes

Klasifikasi

Bantuan sosial

Data mining

Pengambilan keputusan

Riwayat artikel:

Submit 4 Juli 2026

Revisi 8 Juli 2026

Diterima 14 Juli 2026

1. PENDAHULUAN

Program Bantuan Pangan Non-Tunai (BPNT) merupakan salah satu program prioritas pemerintah dalam upaya pengentasan kemiskinan dan peningkatan kesejahteraan masyarakat melalui pemenuhan kebutuhan pangan secara berkelanjutan. BPNT diberikan kepada Keluarga Penerima Manfaat (KPM) berdasarkan kriteria kelayakan tertentu yang merujuk pada kondisi sosial dan ekonomi rumah tangga [1]. Dalam implementasinya, proses penentuan kelayakan penerima BPNT masih menghadapi berbagai permasalahan, terutama terkait ketepatan sasaran penerima bantuan. Proses seleksi yang masih dilakukan secara konvensional dan melibatkan penilaian subjektif berpotensi menyebabkan kesalahan dalam menentukan masyarakat yang benar-benar membutuhkan bantuan [2].

Dalam beberapa tahun terakhir, pemerintah Indonesia terus melakukan penguatan pengelolaan data bantuan sosial melalui integrasi dan pemutakhiran basis data nasional. Pengembangan Data Tunggal Sosial Ekonomi Nasional (DTSEN) yang dibangun berdasarkan integrasi Regsosek, DTKS, dan P3KE bertujuan meningkatkan kualitas data melalui proses verifikasi dan validasi lintas lembaga pemerintahan [3]. Selain itu, Program Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) menyediakan data terpadu yang mencakup berbagai informasi seperti

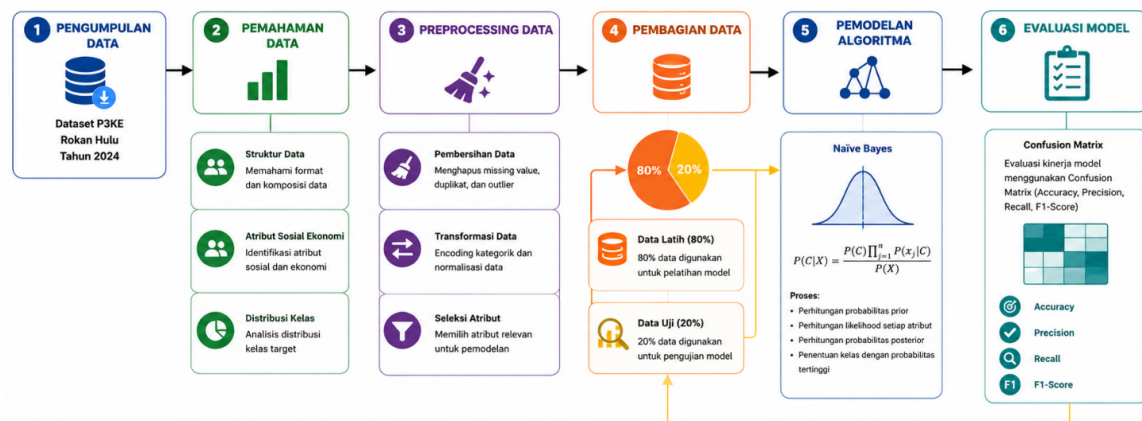
aspek demografi, sosial, ekonomi, dan kondisi kesejahteraan rumah tangga yang dapat digunakan sebagai dasar dalam pengambilan keputusan penerima bantuan [4]. Ketidaktepatan sasaran dalam distribusi bantuan sosial masih menjadi tantangan utama di berbagai negara berkembang. Permasalahan tersebut dapat berupa kesalahan inklusi, yaitu bantuan diberikan kepada kelompok yang tidak memenuhi kriteria, maupun kesalahan eksklusi, yaitu masyarakat yang memenuhi kriteria tetapi tidak memperoleh bantuan. Penelitian sebelumnya menunjukkan bahwa pemanfaatan machine learning mampu meningkatkan efektivitas penentuan target penerima bantuan melalui proses analisis pola data secara otomatis dibandingkan pendekatan manual [5].

Pemanfaatan data mining dalam pengolahan data sosial memungkinkan proses identifikasi pola penerima bantuan dilakukan secara lebih sistematis melalui pendekatan komputasi. Teknik klasifikasi dapat membantu mengelompokkan calon penerima berdasarkan karakteristik sosial ekonomi yang terdapat pada data historis. Teknik klasifikasi dalam data mining telah banyak diterapkan untuk menyelesaikan berbagai permasalahan prediksi dan pengelompokan, termasuk dalam bidang sosial dan pelayanan masyarakat [6]. Beberapa algoritma seperti Decision Tree, K-Nearest Neighbor, Random Forest, Support Vector Machine, dan Naïve Bayes telah digunakan dalam penelitian klasifikasi dengan tingkat performa yang berbeda sesuai karakteristik dataset yang digunakan [7]–[10]. Salah satu algoritma klasifikasi yang banyak digunakan adalah Naïve Bayes. Algoritma ini bekerja berdasarkan konsep probabilitas dengan asumsi independensi antar atribut dan memiliki keunggulan dalam efisiensi komputasi, kemudahan implementasi, serta kemampuan menangani data dengan atribut yang beragam [11]. Meskipun memiliki asumsi independensi sederhana, berbagai pengembangan dan penelitian terbaru menunjukkan bahwa Naïve Bayes masih mampu menghasilkan performa klasifikasi yang kompetitif pada berbagai kasus analisis data [12]. Penerapan algoritma Naïve Bayes pada bidang bantuan sosial telah banyak dilakukan. Namun, sebagian besar penelitian terdahulu masih berfokus pada pencapaian nilai akurasi sebagai indikator utama keberhasilan model. Pada kasus bantuan sosial dengan distribusi data tidak seimbang, penggunaan akurasi saja belum cukup karena model dapat menghasilkan akurasi tinggi dengan kecenderungan mengenali kelas mayoritas. Oleh karena itu, diperlukan pendekatan yang memperhatikan kemampuan model dalam mengenali kelas minoritas melalui evaluasi recall dan penanganan ketidakseimbangan data. Penelitian sebelumnya menunjukkan bahwa Naïve Bayes mampu digunakan untuk menentukan kelayakan penerima bantuan sosial berdasarkan atribut kondisi masyarakat dengan hasil klasifikasi yang baik [13]. Selain itu, optimasi dan perbandingan metode Naïve Bayes dengan algoritma lain menunjukkan bahwa metode ini tetap relevan sebagai pendekatan klasifikasi dalam sistem pendukung keputusan penerima bantuan [14], [15]. Berbeda dengan penelitian sebelumnya yang hanya berfokus pada peningkatan akurasi klasifikasi, penelitian ini mempertimbangkan permasalahan ketidakseimbangan kelas pada data penerima BPNT melalui penerapan SMOTEN serta evaluasi berbasis recall untuk mengurangi risiko kesalahan identifikasi penerima bantuan.

Berdasarkan permasalahan tersebut, rumusan masalah dalam penelitian ini adalah bagaimana membangun model klasifikasi yang mampu meningkatkan kemampuan identifikasi penerima BPNT pada kondisi dataset sosial ekonomi yang memiliki distribusi kelas tidak seimbang. Selain itu, penelitian ini menganalisis pengaruh penerapan SMOTEN dan optimasi parameter terhadap kinerja algoritma Naïve Bayes terutama dalam meningkatkan kemampuan deteksi terhadap kelas penerima bantuan data P3KE Kabupaten Rokan Hulu tahun 2024. Penelitian ini bertujuan menghasilkan model klasifikasi berbasis data mining yang mampu membantu proses identifikasi penerima bantuan secara lebih objektif, akurat, dan tepat sasaran.

2. METODE PENELITIAN

Penelitian ini menggunakan tahapan analisis data berbasis CRISP-DM sebagai pendekatan sistematis dalam membangun model klasifikasi penerima BPNT. Tahapan penelitian meliputi pemahaman data, persiapan data, pembentukan model, serta evaluasi hasil klasifikasi. Dalam penelitian ini, tahapan CRISP-DM disajikan dalam bentuk alur tahapan penelitian sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan penelitian ini dimulai dengan proses pengumpulan data menggunakan dataset Program Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Kabupaten Rokan Hulu tahun 2024. Dataset tersebut berisi informasi mengenai kondisi sosial dan ekonomi masyarakat yang digunakan sebagai dasar dalam menentukan kelayakan penerima Bantuan Pangan Non-Tunai (BPNT). Setelah data diperoleh, dilakukan tahap pemahaman data (data understanding) untuk mengetahui karakteristik dataset yang digunakan [16]. Pada tahap ini dilakukan analisis terhadap struktur data, identifikasi atribut sosial ekonomi, serta distribusi kelas penerima bantuan. Proses ini bertujuan untuk memahami pola data dan memastikan atribut yang tersedia sesuai dengan kebutuhan pemodelan [17]. Tahap berikutnya adalah preprocessing data yang bertujuan meningkatkan kualitas data sebelum dilakukan proses klasifikasi. Tahapan ini meliputi pembersihan data untuk menangani nilai kosong (missing value), data duplikat, maupun data tidak sesuai. Selanjutnya dilakukan transformasi data melalui proses encoding kategori dan normalisasi, serta seleksi atribut untuk memilih variabel yang relevan terhadap proses klasifikasi penerima BPNT.

Data yang telah melalui tahap preprocessing kemudian dibagi menjadi dua bagian, yaitu data latih (training data) sebesar 80% dan data uji (testing data) sebesar 20%. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data baru. Tahap selanjutnya adalah pemodelan menggunakan algoritma Naïve Bayes. Algoritma ini bekerja dengan mempelajari pola hubungan antara atribut sosial ekonomi dengan status kelayakan penerima BPNT berdasarkan probabilitas dari data latih. Model yang terbentuk kemudian digunakan untuk melakukan klasifikasi apakah suatu data termasuk kategori layak atau tidak layak menerima bantuan. Tahap akhir penelitian adalah evaluasi model menggunakan metode confusion matrix. Evaluasi dilakukan untuk mengukur performa algoritma berdasarkan beberapa indikator seperti accuracy, precision, recall, dan F1-score. Hasil evaluasi digunakan untuk mengetahui tingkat keberhasilan algoritma Naïve Bayes dalam mendukung proses penentuan penerima BPNT secara lebih objektif dan berbasis data.

Berikut Rumus Naïve Bayes :

$$P(C|X) = \left(\frac{P(C) \cdot P(x_1|C) \cdot P(x_2|C) \cdot \dots \cdot P(x_n|C)}{P(x_1) \cdot P(x_2) \cdot \dots \cdot P(x_n)} \right) \quad (1)$$

Keterangan:

$P(C|X)$ = probabilitas kelas C berdasarkan atribut X
 $P(X|C)$ = probabilitas atribut X terhadap kelas tertentu
 $P(C)$ = probabilitas awal kelas
 $P(X)$ = probabilitas atribut data

Dalam penelitian ini, kelas C merepresentasikan status penerima BPNT, sedangkan X merupakan atribut sosial ekonomi rumah tangga yang digunakan dalam proses klasifikasi.

3. HASIL DAN DISKUSI

Pada bab ini disajikan hasil-hasil yang diperoleh dari tahapan penelitian, meliputi hasil analisis data, preprocessing data, pemodelan menggunakan algoritma Naïve Bayes, serta hasil pengujian dan evaluasi model.

3.1. Hasil Analisis Data

Dataset yang digunakan dalam penelitian ini merupakan data P3KE (Percepatan Penghapusan Kemiskinan Ekstrem) Kabupaten Rokan Hulu tahun 2024 dengan jumlah total 2055 data rumah tangga. Dataset ini terdiri dari 37 atribut yang merepresentasikan kondisi sosial, ekonomi, dan demografi rumah tangga. Variabel target dalam penelitian ini adalah status Penerima BPNT, yang dibagi ke dalam dua kelas, yaitu :

1. Penerima BPNT sebanyak 450 data ($\approx 22\%$).
2. Bukan Penerima BPNT sebanyak 1.605 data ($\approx 78\%$).

Distribusi kelas tersebut menunjukkan bahwa dataset memiliki kondisi ketidakseimbangan kelas (imbalanced data).

```
Distribusi target:
Penerima BPNT
Bukan Penerima    1605
Penerima          450
Name: count, dtype: int64
```

Gambar 2. Analisis Awal Dataset

Berdasarkan hasil eksekusi kode pada gambar 2, diperoleh informasi bahwa dataset memiliki 2055 baris data dengan distribusi kelas yang tidak seimbang, di mana data Bukan Penerima BPNT lebih dominan dibandingkan Penerima BPNT.

3.2. Hasil Preprocessing Data

Preprocessing data yang dilakukan meliputi pembersihan data, transformasi data, dan seleksi atribut. Atribut administratif dan identitas pribadi seperti nomor urut, ID keluarga, alamat, serta nomor identitas kependudukan dihapus karena tidak memiliki pengaruh langsung terhadap proses klasifikasi. Pembersihan data dilakukan dengan menghapus atribut yang bersifat administratif dan identitas pribadi karena tidak berpengaruh langsung terhadap penentuan kelayakan penerima BPNT serta berpotensi menimbulkan bias pada model klasifikasi. Atribut yang dihapus antara lain nomor urut, ID keluarga, alamat, dan nomor identitas kependudukan yang dapat dilihat pada gambar 3.

```
Dimukhtakhirkan Tahun Provinsi Kabupaten/Kota Kecamatan \
0 2023 RIAU ROKAN HULU RAMBAH
1 2023 RIAU ROKAN HULU RAMBAH
2 2023 RIAU ROKAN HULU RAMBAH
3 2023 RIAU ROKAN HULU RAMBAH
4 2023 RIAU ROKAN HULU RAMBAH

Desa/Kelurahan Kode Kemdagri Desil Kesejahteraan Persentil \
0 RAMBAH TENGAH UTARA 1406032002 1 1
1 RAMBAH TENGAH UTARA 1406032002 1 1
2 RAMBAH TENGAH UTARA 1406032002 1 2
3 RAMBAH TENGAH UTARA 1406032002 1 2
4 RAMBAH TENGAH UTARA 1406032002 1 2

Status Verval Keluarga Status Verval Individu ... Kualitas Dinding
0 Belum Terverifikasi Pemda Belum Terverifikasi Pemda ... Tidak Menjawab
1 Belum Terverifikasi Pemda Belum Terverifikasi Pemda ... Tidak Menjawab
2 Belum Terverifikasi Pemda Belum Terverifikasi Pemda ... Tidak Menjawab
3 Belum Terverifikasi Pemda Belum Terverifikasi Pemda ... Tidak Menjawab
4 Belum Terverifikasi Pemda Belum Terverifikasi Pemda ... Tidak Menjawab

Jenis Lantai Kualitas Lantai Sumber Penerangan Daya Listrik Terpasang \
0 Semen Tidak Menjawab Listrik PLN <= 900 watt
1 Semen Tidak Menjawab Listrik PLN <= 900 watt
2 Semen Tidak Menjawab Listrik PLN Tidak Menjawab
...
3 Ya, dengan Septictank Tidak Beresiko Stunting Bukan Penerima
4 Ya, dengan Septictank Beresiko Stunting Bukan Penerima

[5 rows x 32 columns]
```

Gambar 3. Hasil pembersihan data

Berdasarkan gambar 3, diperoleh data yang telah dibersihkan dari atribut administratif dan identitas. Data yang tersisa hanya atribut sosial ekonomi yang relevan dengan proses klasifikasi. Transformasi data dilakukan untuk menyesuaikan format data agar dapat diproses oleh algoritma Naïve Bayes. Atribut numerik seperti Desil Kesejahteraan dan Persentil dipertahankan dalam bentuk numerik dan dilakukan proses diskretisasi (binning). Atribut kategorikal seperti pekerjaan, kondisi rumah, dan akses fasilitas dasar dilakukan proses pengkodean (encoding) ke dalam bentuk numerik.

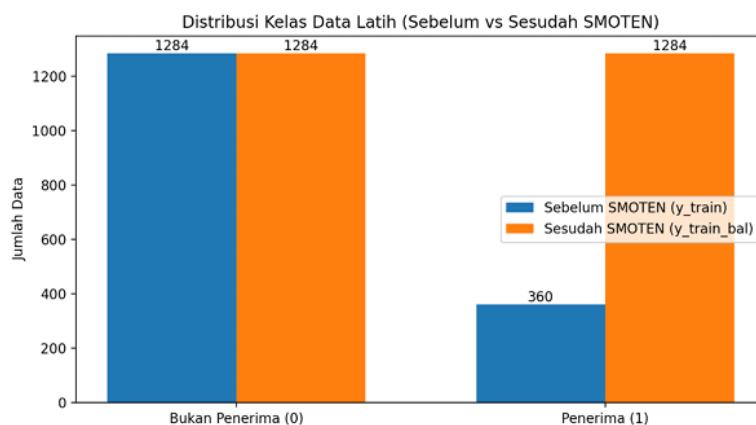
...	Dimuktakhirkan Tahun	Provinsi	Kabupaten/Kota	Kecamatan	Desa/Kelurahan	\
0	2023	0.0	0.0	3.0	16.0	
1	2023	0.0	0.0	3.0	16.0	
2	2023	0.0	0.0	3.0	16.0	
3	2023	0.0	0.0	3.0	16.0	
4	2023	0.0	0.0	3.0	16.0	
	Kode Kemdagri	Desil Kesejahteraan	Persentil	Status Verval Keluarga	\	
0	1406032002	1	1	0.0		
1	1406032002	1	1	0.0		
2	1406032002	1	2	0.0		
3	1406032002	1	2	0.0		
4	1406032002	1	2	0.0		
	Status Verval Individu	...	Kualitas Dinding	Jenis Lantai	\	
0	0.0	...	0.0	3.0		
1	0.0	...	0.0	3.0		
2	0.0	...	0.0	3.0		
3	0.0	...	0.0	3.0		
4	0.0	...	0.0	3.0		
	Kualitas Lantai	Sumber Penerangan	Daya Listrik Terpasang	\		
0	0.0	1.0	0.0			
1	0.0	1.0	0.0			
2	0.0	1.0	2.0			
...						
3	2.0	0.0				
4	0.0	0.0				

Gambar 4. Hasil transformasi data

Hasil transformasi data pada gambar 4 menunjukkan bahwa seluruh atribut telah berada dalam format numerik sehingga dapat digunakan dalam tahap pemodelan algoritma Naïve Bayes. Seleksi atribut dilakukan untuk memilih atribut yang paling relevan dalam menentukan status penerima BPNT. Atribut yang dipertahankan adalah atribut yang merepresentasikan kondisi sosial ekonomi rumah tangga, seperti tingkat kesejahteraan, kondisi tempat tinggal, akses fasilitas dasar, serta karakteristik kepala keluarga. Atribut yang tidak memiliki kontribusi langsung terhadap penentuan kelas target dikeluarkan untuk mengurangi kompleksitas model dan meningkatkan stabilitas hasil klasifikasi. Setelah dilakukan seleksi atribut, dataset akhir yang digunakan dalam pemodelan hanya terdiri dari atribut-atribut yang relevan dan siap digunakan pada tahap klasifikasi.

3.3. Distribusi Kelas Data Latih

Visualisasi distribusi kelas pada data latih sebelum dan sesudah penerapan metode SMOTEN ditunjukkan pada Gambar 5.



Gambar 5. Distribusi kelas data latih

Pada kondisi awal, jumlah data latih kelas Bukan Penerima BPNT jauh lebih banyak dibandingkan kelas Penerima BPNT. Ketidakseimbangan ini berpotensi menyebabkan model klasifikasi memiliki kecenderungan untuk memprediksi kelas mayoritas dan mengabaikan kelas minoritas. Setelah dilakukan penanganan imbalanced data menggunakan SMOTEN, distribusi kelas pada data latih menjadi lebih seimbang. Penambahan data sintesis pada kelas Penerima BPNT bertujuan untuk memberikan representasi yang lebih baik terhadap kelas minoritas, sehingga model dapat mempelajari karakteristik rumah tangga yang layak menerima bantuan secara lebih optimal. Grafik ini menunjukkan bahwa SMOTEN berhasil mengurangi ketimpangan distribusi kelas pada tahap pelatihan model.

3.4. Hasil Klasifikasi Naïve Bayes (Model Baseline)

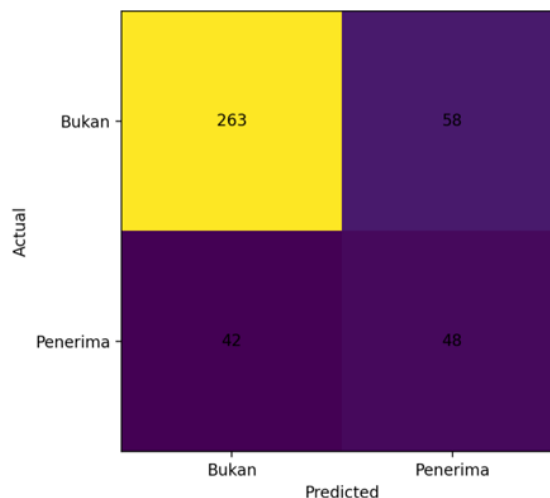
Model klasifikasi baseline dibangun menggunakan algoritma Naïve Bayes tanpa penanganan ketidakseimbangan data. Dataset yang telah melalui tahap preprocessing dibagi menjadi data latih dan data uji dengan rasio 80:20 dengan menjaga proporsi kelas tetap sama seperti di data asli.

```
[INFO] Shift applied: 0
[INFO] Encoded train shape: (1644, 29)

=== BASELINE (Tanpa Balancing) ===
Accuracy : 0.7566909975669099
Precision: 0.4528301886792453
Recall   : 0.5333333333333333
F1-score : 0.4897959183673469
...
      accuracy                0.76      411
      macro avg      0.66      0.68      0.67      411
      weighted avg   0.77      0.76      0.76      411
```

Gambar 6. Hasil klasifikasi tanpa optimasi (Model Baseline)

Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan data dengan cukup baik, namun masih mengalami kesalahan pada kelas minoritas. Berdasarkan hasil pengujian terhadap data uji, diperoleh nilai evaluasi model baseline sebagai berikut: akurasi sebesar 75,67%, presisi sebesar 45,28%, recall sebesar 53,33%, dan F1-score sebesar 48,98%. Nilai akurasi menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Namun, nilai presisi dan recall yang masih relatif rendah mengindikasikan bahwa model baseline masih mengalami kesulitan dalam membedakan rumah tangga yang benar-benar layak menerima BPNT dengan yang tidak layak, terutama pada kelas minoritas.



Gambar 7. Confusion Matrix (Model Baseline)

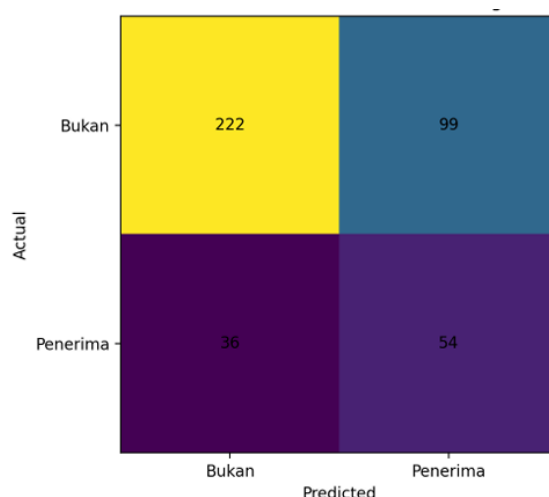
Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa masih terdapat sejumlah data Penerima BPNT yang salah diklasifikasikan sebagai Bukan Penerima BPNT. Kondisi ini dipengaruhi oleh distribusi kelas yang tidak seimbang, sehingga model cenderung lebih banyak memprediksi kelas mayoritas.

3.5. Hasil Klasifikasi Setelah Optimasi (SMOTEN + Tuning)

Teknik resampling seperti oversampling sering dipakai untuk mengatasi ketidakseimbangan kelas, tetapi pemakaiannya perlu kehati-hatian karena dapat mengubah struktur data latih. Pada kondisi tertentu, model justru dapat belajar mengenali noise/jejak dari proses oversampling (data sintetis), bukan pola asli kelas minoritas (Boudegz dame et al., 2024). Selain itu, dilakukan tuning parameter smoothing (alpha) pada algoritma Naïve Bayes untuk memperoleh konfigurasi model yang memberikan kinerja terbaik. Proses tuning dilakukan dengan pendekatan

validasi silang untuk memastikan bahwa parameter yang dipilih mampu meningkatkan performa model secara konsisten.

Hasil pengujian model setelah dilakukan SMOTEN dan tuning parameter menunjukkan nilai akurasi sebesar 67,15%, presisi sebesar 35,29%, recall sebesar 60,00%, dan F1-score sebesar 44,44%. Terjadi peningkatan nilai recall pada kelas Penerima BPNT dibandingkan dengan model baseline, yang menunjukkan bahwa model menjadi lebih baik dalam mendeteksi rumah tangga yang layak menerima bantuan. Namun demikian, peningkatan recall tersebut diikuti dengan penurunan nilai akurasi dan presisi, yang menandakan adanya peningkatan kesalahan prediksi pada kelas Bukan Penerima BPNT.



Gambar 8. Confusion Matrix setelah optimasi (Smote + Tuning)

Berdasarkan Gambar 8 dapat dilihat bahwa jumlah data Penerima BPNT yang berhasil diklasifikasikan dengan benar (true positive) meningkat dibandingkan dengan model baseline. Di sisi lain, jumlah data Bukan Penerima BPNT yang salah diklasifikasikan sebagai Penerima BPNT (false positive) juga mengalami peningkatan. Hal ini menunjukkan bahwa penerapan SMOTEN membuat model lebih sensitif terhadap kelas minoritas, sehingga mampu mengurangi kesalahan tidak terdeteksinya penerima bantuan, meskipun dengan konsekuensi meningkatnya kesalahan prediksi pada kelas mayoritas.

3.6. Pembahasan

Berdasarkan hasil pengujian menggunakan model Naïve Bayes baseline, diperoleh nilai akurasi sebesar 75,67%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Namun, apabila ditinjau dari nilai precision dan recall, model masih memiliki keterbatasan dalam mengidentifikasi rumah tangga yang benar-benar layak menerima Bantuan Pangan Non-Tunai (BPNT). Nilai recall sebesar 53,33% menunjukkan bahwa hampir setengah dari data penerima BPNT belum berhasil dikenali oleh model. Kondisi ini disebabkan oleh distribusi dataset yang tidak seimbang, di mana jumlah data Bukan Penerima BPNT jauh lebih banyak dibandingkan Penerima BPNT. Pada kondisi seperti ini, algoritma Naïve Bayes cenderung mempelajari pola dari kelas mayoritas sehingga probabilitas posterior lebih sering mengarah pada kelas mayoritas. Akibatnya, jumlah false negative meningkat dan kemampuan model dalam mengenali kelas penerima bantuan menjadi menurun.

Hasil confusion matrix pada model baseline memperlihatkan masih terdapat sejumlah data Penerima BPNT yang salah diklasifikasikan sebagai Bukan Penerima BPNT (false negative). Dalam konteks penyaluran bantuan sosial, kesalahan ini memiliki konsekuensi yang cukup serius karena rumah tangga yang seharusnya memperoleh bantuan berpotensi tidak teridentifikasi oleh sistem. Oleh karena itu, meskipun model baseline memiliki akurasi yang relatif tinggi, model tersebut belum sepenuhnya sesuai digunakan sebagai sistem pendukung keputusan karena masih menghasilkan tingkat kesalahan yang cukup besar pada kelas penerima bantuan.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan metode SMOTEN sebagai teknik oversampling pada data kategorikal serta melakukan tuning parameter pada algoritma Naïve Bayes. Hasil pengujian menunjukkan bahwa penerapan kedua metode tersebut meningkatkan nilai recall dari 53,33% menjadi 60,00%. Peningkatan recall terjadi karena SMOTEN membangkitkan data sintesis pada kelas minoritas sehingga distribusi data latih menjadi lebih seimbang. Kondisi ini memberikan kesempatan yang lebih besar bagi algoritma Naïve

Bayes untuk mempelajari karakteristik rumah tangga penerima BPNT secara lebih representatif. Dengan representasi kelas minoritas yang lebih baik, jumlah false negative berhasil dikurangi sehingga lebih banyak rumah tangga yang layak menerima bantuan dapat dikenali oleh model.

Meskipun demikian, peningkatan nilai recall diikuti oleh penurunan accuracy menjadi 67,15% dan precision menjadi 35,29%. Penurunan tersebut terjadi karena model menjadi lebih sensitif terhadap kelas minoritas sehingga lebih banyak data Bukan Penerima BPNT yang diprediksi sebagai Penerima BPNT (false positive). Fenomena ini merupakan karakteristik umum pada permasalahan imbalanced classification, di mana peningkatan sensitivitas terhadap kelas minoritas sering kali diikuti dengan meningkatnya jumlah kesalahan pada kelas mayoritas. Dengan kata lain, model mengalami trade-off antara kemampuan mendeteksi penerima bantuan (recall) dan ketepatan prediksi secara keseluruhan (precision dan accuracy).

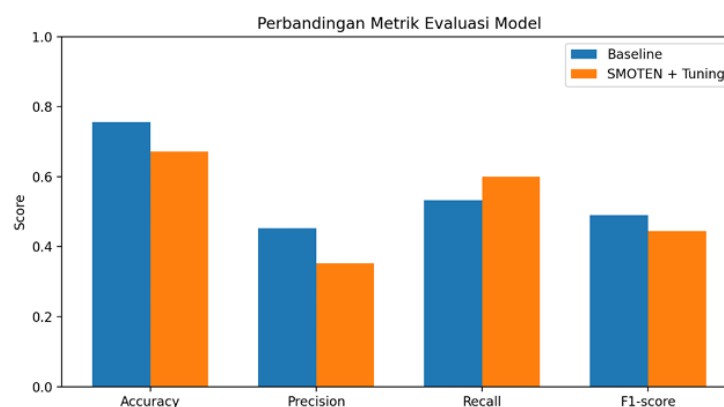
Hasil penelitian ini sejalan dengan penelitian Fatmawati et al. [13] yang menunjukkan bahwa algoritma Naïve Bayes mampu digunakan untuk klasifikasi kelayakan penerima bantuan sosial. Namun, penelitian tersebut belum mempertimbangkan pengaruh ketidakseimbangan distribusi kelas sehingga evaluasi lebih berfokus pada nilai akurasi. Penelitian Saputra et al. [14] juga menunjukkan bahwa optimasi algoritma Naïve Bayes mampu meningkatkan performa klasifikasi. Berbeda dengan kedua penelitian tersebut, penelitian ini menambahkan pendekatan SMOTEN untuk menangani ketidakseimbangan kelas dan menggunakan recall sebagai indikator utama keberhasilan model. Pendekatan ini lebih sesuai diterapkan pada kasus bantuan sosial karena tujuan utama sistem bukan hanya memperoleh akurasi tinggi, tetapi juga meminimalkan risiko rumah tangga yang layak menerima bantuan tidak teridentifikasi oleh model.

Dalam konteks penyaluran bantuan sosial, kesalahan false positive memiliki dampak yang relatif lebih ringan dibandingkan false negative. Rumah tangga yang diprediksi layak menerima bantuan tetapi sebenarnya tidak memenuhi kriteria masih dapat diverifikasi kembali melalui proses validasi lapangan. Sebaliknya, kesalahan false negative menyebabkan rumah tangga yang benar-benar membutuhkan bantuan justru tidak memperoleh haknya. Oleh karena itu, peningkatan recall pada penelitian ini memiliki nilai praktis yang lebih penting dibandingkan peningkatan akurasi semata.

Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma Naïve Bayes dengan penerapan SMOTEN dan tuning parameter mampu meningkatkan kemampuan model dalam mendeteksi calon penerima BPNT pada kondisi dataset yang tidak seimbang. Kontribusi penelitian ini menunjukkan bahwa pada sistem pendukung keputusan bantuan sosial, evaluasi model sebaiknya tidak hanya berorientasi pada nilai akurasi, tetapi juga mempertimbangkan kemampuan model dalam mengenali kelas minoritas melalui indikator recall. Pendekatan ini diharapkan dapat menghasilkan proses penentuan penerima bantuan yang lebih adil, objektif, dan tepat sasaran.

3.7. Perbandingan Matrik Evaluasi Model

Perbandingan kinerja antara model baseline dan model hasil optimasi ditunjukkan melalui grafik perbandingan metrik evaluasi yang meliputi accuracy, precision, recall, dan F1-score.

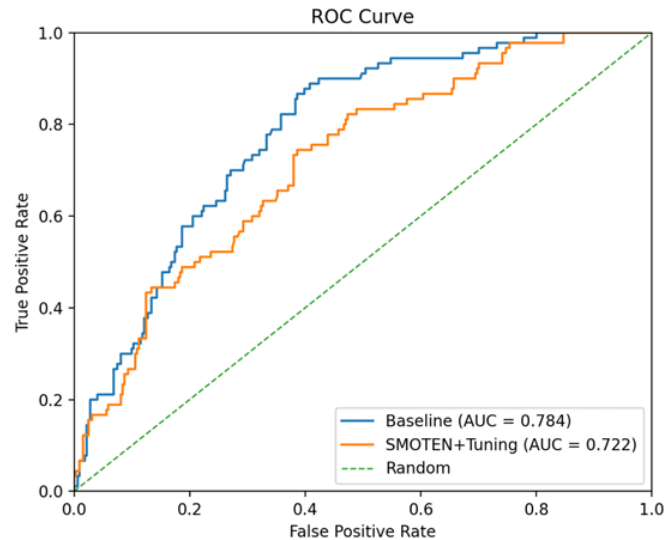


Gambar 9. Perbandingan metrik evaluasi model

Berdasarkan grafik tersebut, terlihat bahwa penerapan SMOTEN dan tuning parameter memberikan peningkatan pada nilai recall dan F1-score dibandingkan model baseline. Peningkatan nilai recall menunjukkan bahwa model hasil optimasi lebih efektif dalam mengurangi kesalahan false negative, yaitu kondisi di mana rumah tangga yang seharusnya menerima BPNT justru tidak terdeteksi oleh model. Meskipun nilai akurasi dan presisi mengalami penurunan, peningkatan recall dinilai lebih penting dalam konteks bantuan sosial karena kesalahan false positive masih dapat diperbaiki melalui proses verifikasi lanjutan di lapangan.

3.8. Analisis ROC Curve

Evaluasi lebih lanjut dilakukan menggunakan ROC Curve untuk menilai kemampuan model dalam membedakan kelas Penerima BPNT dan Bukan Penerima BPNT berdasarkan probabilitas prediksi.

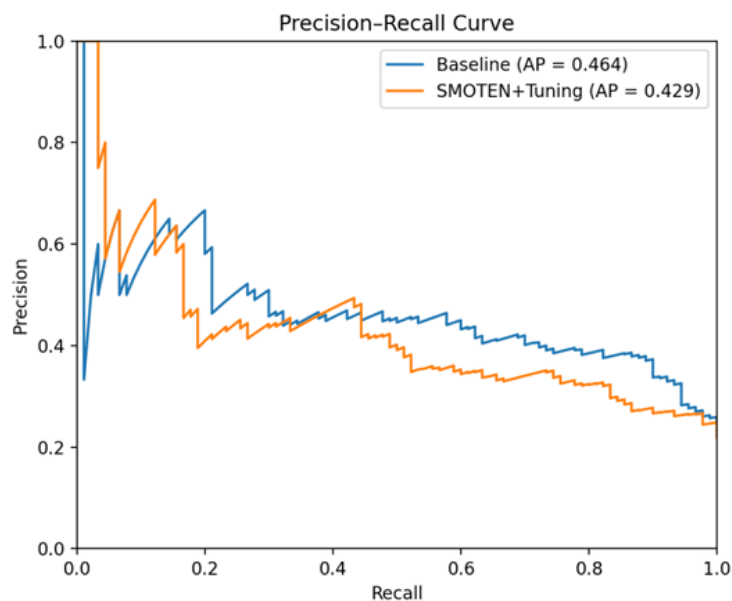


Gambar 10. ROC Curve

Berdasarkan ROC Curve, model baseline memiliki nilai AUC sebesar 0,784, sedangkan model hasil SMOTEN dan tuning memperoleh nilai AUC sebesar 0,722. Hal ini menunjukkan bahwa penerapan SMOTEN belum mampu meningkatkan kemampuan diskriminasi model berdasarkan metrik AUC. Meskipun demikian, model hasil optimasi tetap menghasilkan nilai recall yang lebih tinggi sehingga lebih sesuai digunakan pada kasus klasifikasi penerima bantuan sosial yang berorientasi pada pengurangan kesalahan false negative. Dengan demikian, penerapan SMOTEN dan tuning parameter tidak hanya meningkatkan sensitivitas model terhadap kelas minoritas, tetapi juga memperbaiki kualitas pemisahan kelas secara keseluruhan.

3.9. Analisis Precision–Recall Curve

Precision–Recall Curve digunakan untuk mengevaluasi kinerja model pada kondisi data yang tidak seimbang.



Gambar 11. Precision-Recall Curve

Berdasarkan grafik Precision–Recall, model hasil optimasi menunjukkan nilai average precision yang lebih tinggi dibandingkan model baseline. Hal ini menunjukkan bahwa model mampu mempertahankan keseimbangan yang lebih baik antara precision dan recall pada kelas Penerima BPNT. Penggunaan PrecisionRecall Curve pada penelitian ini menjadi penting karena fokus utama penelitian adalah mendeteksi rumah tangga yang layak menerima bantuan, yang jumlahnya relatif lebih sedikit dibandingkan kelas lainnya. Hasil ini memperkuat temuan bahwa model Naïve Bayes dengan penanganan SMOTEN dan tuning parameter lebih sesuai digunakan dalam konteks penentuan penerima BPNT dibandingkan model baseline tanpa penanganan data tidak seimbang.

4. KESIMPULAN

Berdasarkan hasil penelitian, algoritma Naïve Bayes mampu digunakan untuk mengklasifikasikan kelayakan penerima Bantuan Pangan Non-Tunai (BPNT) berdasarkan data sosial ekonomi dari P3KE. Model baseline menghasilkan akurasi sebesar 75,67% dengan nilai recall sebesar 53,33%, yang menunjukkan bahwa masih terdapat rumah tangga layak menerima bantuan namun belum berhasil dikenali oleh model. Setelah diterapkan metode SMOTEN dan optimasi parameter, nilai recall meningkat menjadi 60,00%, sehingga kemampuan model dalam mendeteksi calon penerima BPNT menjadi lebih baik meskipun terjadi penurunan pada nilai akurasi dan precision sebagai konsekuensi dari peningkatan sensitivitas terhadap kelas minoritas.

Penelitian ini berhasil menunjukkan bahwa pada permasalahan klasifikasi penerima bantuan sosial dengan distribusi data yang tidak seimbang, pendekatan berbasis SMOTEN lebih sesuai dibandingkan hanya berfokus pada peningkatan nilai akurasi. Peningkatan recall memberikan manfaat yang lebih besar karena mampu mengurangi risiko rumah tangga yang berhak menerima bantuan tetapi tidak teridentifikasi oleh sistem (false negative). Dengan demikian, penelitian ini memberikan kontribusi ilmiah berupa penerapan strategi penanganan ketidakseimbangan kelas pada algoritma Naïve Bayes sebagai pendekatan yang lebih relevan untuk mendukung sistem pendukung keputusan penentuan penerima BPNT secara lebih adil, objektif, dan tepat sasaran.

Penelitian selanjutnya dapat mengembangkan model dengan membandingkan algoritma klasifikasi lainnya, seperti Random Forest, XGBoost, atau LightGBM, serta menerapkan teknik penanganan data tidak seimbang yang lebih beragam untuk meningkatkan performa klasifikasi pada data bantuan sosial.

REFERENSI

- [1] W. W. Arupandani, F. Taufik, and R. Mahyuni, “Implementasi Data Mining Menentukan Penerimaan Bantuan Sosial Pangan (BSP) Menggunakan Algoritma C4.5,” *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 2, no. 5, p. 705, 2023, doi: 10.53513/jursi.v2i5.5612.
- [2] A. P. Siregar, D. Irmayani, and M. N. Sari, “Analysis of the Naïve Bayes Method for Determining Social Assistance Eligibility Public,” *Sinkron*, vol. 8, no. 2, pp. 805–817, 2023, doi: 10.33395/sinkron.v8i2.12259.
- [3] L. Dasar, P. Implementasi, and R. Daerah, “Risalah Kebijakan Optimalisasi Pemutakhiran Data Tunggal Sosial Ekonomi Nasional,” pp. 1–4, 2025.
- [4] T. Srimurni and M. Sholihah, “Implementasi Program Bantuan Pangan Non Tunai (BPNT) Desa Duren Kecamatan Klakah Kabupaten Lumajang,” *KNOWLEDGE: Jurnal Inovasi Hasil Penelitian dan Pengembangan*, vol. 2, no. 4, pp. 309–316, 2022.
- [5] E. Aiken, S. Bellue, D. Karlan, C. Udry, and J. E. Blumenstock, “Machine learning and phone data can improve targeting of humanitarian aid,” *Nature*, vol. 603, no. 7903, pp. 864–870, 2022, doi: 10.1038/s41586-022-04484-9.
- [6] C. Pete, J. Julian, R. Randy, T. Thomas, T. Reinartz, C. Shearer, and R. Wirth, “CRISP-DM 1.0: Step-by-step data mining guide,” *CRISP-DM Consortium*, p. 76, 2000.
- [7] A. I. Perhan, I. Yustiana, and I. Sanjaya, “Implementation of Machine Learning Using Decision Tree Method for Social Assistance Recipient Classification,” *Bit-Tech*, vol. 8, no. 1, pp. 901–909, 2025, doi: 10.32877/bt.v8i1.2755.
- [8] A. Hasibuan, “Penerapan Algoritma C4.5 Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan,” 2021.
- [9] A. Putri, C. S. Hardiana, E. Novfuja, F. T. P. Siregar, Rahmaddeni, Y. Fatma, and R. Wahyuni, “Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.

- [10] M. Safrudin, Martanto, and U. Hayati, “Perbandingan Kinerja Naïve Bayes dan Support Vector Machine untuk Klasifikasi Sentimen Ulasan Game Genshin Impact,” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 3, pp. 3182–3188, 2024, doi: 10.36040/jati.v8i3.8415.
- [11] A. L. R. Entini and K. Handoko, “Implementasi Data Mining dengan Algoritma Naive Bayes,” Jurnal Comasie, vol. 3, pp. 343–351, 2023.
- [12] X. Q. Liu, X. C. Wang, L. Tao, F. X. An, and G. R. Jiang, “Alleviating conditional independence assumption of Naive Bayes,” Statistical Papers, vol. 65, no. 5, pp. 2835–2863, 2024, doi: 10.1007/s00362-023-01474-5.
- [13] A. Fatmawati, A. I. Purnamasari, and I. Ali, “Implementasi Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Sosial,” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 1, pp. 745–750, 2024, doi: 10.36040/jati.v8i1.8714.
- [14] A. K. Saputra, W. Susanty, and A. S. Kurniawan, “Optimasi Naïve Bayes Classifier Menggunakan Particle Swarm Optimization pada Klasifikasi Data Penerima Bantuan Sosial,” EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi, vol. 14, no. 2, p. 80, 2024, doi: 10.36448/expert.v14i2.3951.
- [15] Nilawati, “Perbandingan Tingkat Akurasi Metode Weighted Naïve Bayes dengan Random Forest dalam Mengklasifikasi Penerima Program Keluarga Harapan (PKH),” Program Studi Statistika Universitas Sulawesi Barat, 2024.
- [16] F. Nugraha and A. Surahmat, “Sistem Otomatis Ringkasan Laporan Keuangan Berbasis PDF Menggunakan Metode NLP Transformer”, FORMAT, vol. 14, no. 2, pp. 238–244, Oct. 2025.
- [17] Y. Yanuardi, F. F. Basri, F. F. Basri, M. L. Aksani, and M. L. Aksani, “Klasifikasi Kepribadian Berdasarkan Dimensi Ekstraversi Berbasis Data Mining Menggunakan Extremely Randomized Trees”, FORMAT, vol. 14, no. 2, pp. 229–237, Sep. 2025.