



JURNAL SISTEM INFORMASI DAN TEKNOLOGI (S I N T E K)



Situs Jurnal

<https://sintek.stmikku.ac.id/index.php/home>

KLASIFIKASI TINGKAT KESEJAHTERAAN MASYARAKAT DESA CIKUYA BERDASARKAN DATA SOSIAL EKONOMI MENGGUNAKAN ALGORITMA NAÏVE BAYES

Ramdan Irawan¹, Rudi Kurniawan², Bani Nurhakim³, Arif Rinaldi⁴, Fathurrahman⁵

¹Teknik Informatika, ³Manajemen Informatika, ⁴Rekayasa Perangkat Lunak
STMIK IKMI Cirebon,
Jl. Perjuangan No. 10B, Cirebon. Jawa Barat

Email: ¹ramdanirawan280@gmail.com, ²rudi226ikmi@gmail.com, ³baninurhakim@gmail.com,
⁴rinaldi21crb@gmail.com, ⁵faturrahman.ikmi@gmail.com

ABSTRAK

Penentuan tingkat kesejahteraan masyarakat memiliki peran penting dalam proses penyaluran bantuan sosial di tingkat desa. Namun, pendataan berbasis observasi manual masih menghadirkan potensi bias subjektif dan ketidakkonsistenan dalam pengambilan keputusan. Penelitian ini bertujuan mengembangkan model klasifikasi tingkat kesejahteraan masyarakat Desa Cikuya menggunakan algoritma Naïve Bayes sebagai pendekatan berbasis data yang lebih objektif. Tahapan penelitian meliputi pengumpulan data sosial ekonomi, pra-pemrosesan, encoding variabel kategorik, normalisasi variabel numerik, pelatihan model Gaussian Naïve Bayes, serta evaluasi menggunakan metrik akurasi, precision, recall, dan f1-score. Hasil penelitian menunjukkan bahwa model menghasilkan akurasi sebesar 98,33%, yang menunjukkan performa klasifikasi yang sangat baik. Analisis lebih lanjut mengindikasikan bahwa variabel pendapatan dan kondisi fisik rumah memiliki peranan paling dominan dalam membedakan kategori kesejahteraan. Model yang dikembangkan tidak hanya berfungsi sebagai alat klasifikasi, tetapi juga dapat dimanfaatkan sebagai sistem pendukung keputusan bagi pemerintah desa untuk menilai status kesejahteraan masyarakat secara lebih cepat, konsisten, dan bebas bias subjektif. Penelitian ini memberikan kontribusi pada pemanfaatan teknologi pembelajaran mesin dalam pemetaan kesejahteraan masyarakat, meskipun masih memiliki keterbatasan pada jumlah variabel dan cakupan data lokal. Temuan ini diharapkan dapat menjadi dasar pengembangan sistem penyaluran bantuan yang lebih tepat sasaran dan transparan.

Kata Kunci: *Naïve Bayes*, Klasifikasi Kesejahteraan, Data Sosial Ekonomi, *Machine Learning*, SDGs.

1. PENDAHULUAN

Pemanfaatan teknologi informasi dalam bidang sosial ekonomi telah menjadi aspek penting dalam mendukung kebijakan pemerintah, khususnya dalam penyaluran bantuan sosial dan pengukuran tingkat kesejahteraan masyarakat. Berbagai lembaga pemerintah dan akademisi kini mengandalkan pendekatan berbasis *machine learning* untuk mengolah data sosial-ekonomi dalam jumlah besar secara efisien dan akurat. [1]. Salah satu algoritma yang sering digunakan adalah Naïve Bayes classifier, yang dikenal karena kemampuannya

dalam menangani data kategorikal serta memberikan hasil klasifikasi dengan tingkat akurasi yang kompetitif (Agustyaningrum, 2023). Metode *Naïve Bayes* telah banyak diterapkan dalam analisis kesejahteraan sosial. [2].

Selain pada konteks bantuan sosial, metode ini juga diterapkan dalam mengukur aspek sosial lain seperti kepuasan masyarakat dan kesejahteraan umum. [3]. (Fauziah & Dana, 2023) menunjukkan bahwa *Naïve Bayes* dapat mengklasifikasikan status kesejahteraan masyarakat desa secara efektif dengan memanfaatkan data pendidikan, pendapatan, dan

aset keluarga. Hal ini memperlihatkan fleksibilitas algoritma ini dalam menganalisis berbagai jenis data sosial.

penelitian terdahulu juga membandingkan kinerja *Naïve Bayes* dengan algoritma lain seperti K-Nearest Neighbor (K-NN). (Kamali, 2022) menemukan bahwa meskipun *Naïve Bayes* lebih sederhana, hasil klasifikasinya pada penyaluran bansos COVID-19 memiliki tingkat akurasi yang sebanding dengan K-NN, bahkan lebih baik pada dataset dengan distribusi data tertentu. (Setiawan, 2021) menegaskan pentingnya penerapan *data mining* berbasis algoritma ini dalam menentukan calon penerima bantuan sosial, karena hasilnya dapat dijadikan dasar pengambilan keputusan yang objektif dan transparan. [4]. Dengan demikian, berbagai penelitian tersebut memperlihatkan bahwa algoritma *Naïve Bayes* memiliki potensi besar dalam membantu pengambilan keputusan sosial berbasis data. Selain efisien dan mudah diimplementasikan, metode ini mampu memberikan hasil yang akurat dalam mengklasifikasikan tingkat kesejahteraan dan kelayakan penerima bantuan. Oleh karena itu, pengembangan sistem klasifikasi berbasis *Naïve Bayes* untuk mendukung kebijakan sosial ekonomi menjadi penting guna mewujudkan tata kelola bantuan yang adil, transparan, dan berbasis bukti ilmiah. [5]

2. KAJIAN PUSTAKA

2.1 Landasan Teori

Konsep dasar klasifikasi dalam penambahan data dan pembelajaran mesin teorema Bayes, asumsi independensi bersyarat, pemilihan/pembobotan fitur, penanganan ketidakseimbangan kelas, kalibrasi model, interpretabilitas, dan mitigasi kewajaran/bias sangat penting saat membangun pengklasifikasi Bayes Naïf untuk tingkat kesejahteraan masyarakat (misalnya, Desa Cikuya). [5].

Cara kerja algoritma Naive Bayes didasarkan pada teorema Bayes, di mana probabilitas posterior suatu kelas berdasarkan fitur dihitung melalui kali probabilitas sebelumnya dengan asumsi independensi bersyarat fitur berdasarkan kelas tersebut; suku probabilitas difaktorkan menjadi produk probabilitas bersyarat kelas fitur individual, menyederhanakan pembelajaran dan inferensi (Sarker, 2021) Derivasi matematis menunjukkan bahwa aturan keputusan disederhanakan menjadi perbandingan produk prior kelas \times probabilitas fitur (atau logaritmanya) (Zheng et al., 2023). Asumsi independensi ini dikenal "naïf" karena dalam banyak dataset dunia nyata, fitur berkorelasi, namun Naive Bayes seringkali tetap

kompetitif karena sifat generatif dan estimasi parameternya yang efisien (Romano, 2024), Ekstensi menangani masalah probabilitas nol melalui penghalusan (misalnya, Laplace) dan fitur kontinu melalui asumsi kerapatan Gaussian atau bentuk parametrik lainnya (Wang et al., 2022). Penelitian terbaru menyelidiki koreksi ketergantungan fitur, menunjukkan peningkatan akurasi ketika asumsi independensi dilonggarkan (misalnya, NB yang ditambah korelasi) (Khalilpour Darzi et al., 2021). [6]

Jaringan Saraf Tiruan Dalam (DNN) memainkan peran krusial dalam tugas klasifikasi dengan menggunakan pembelajaran representasi hierarkis, ekstraksi fitur otomatis, dan pemodelan hubungan non-linier kompleks antara masukan dan label kelas, sehingga seringkali mengungguli metode tradisional dalam domain data berskala besar atau tidak terstruktur (H. Chen et al., 2021; Clever et al., 2022; Wang et al., 2022; Zheng et al., 2023). Secara spesifik, DNN mendukung hasil multikelas, menangkap interaksi fitur laten, dan memfasilitasi pembelajaran transfer serta penyempurnaan untuk klasifikasi spesifik domain. Mereka juga menghadirkan tantangan seperti interpretabilitas, overfitting, sensitivitas ketidakseimbangan kelas, dan biaya komputasi yang memerlukan mitigasi melalui regularisasi, pemilihan fitur, penyetulan arsitektur, dan praktik AI yang tepercaya (J. I.-Z. Chen & Pi, 2022). Dalam pemodelan klasifikasi tingkat kesejahteraan komunitas, DNN dapat menawarkan alternatif atau pelengkap *Naïve Bayes* dengan memanfaatkan set fitur yang lebih kaya dan mengungkap pola tingkat kesejahteraan yang kompleks. [7]

3. METODOLOGI PENELITIAN

3.1 Pengumpulan Data

Dalam prosedur penelitian ini menggunakan tahapan Knowledge Discovery in Database (KDD) dan dapat di jelaskan sebagai berikut:

a. Data Acquisition

Data yang digunakan adalah data sekunder yang diperoleh dari dokumentasi desa, seperti data jumlah anggota keluarga, pendapatan, jenis pekerjaan, kondisi rumah.

Berdasarkan hasil observasi diperoleh informasi karakteristik datanya sebagai berikut:

- i. Sumber data dalam penelitian ini diperoleh dari Kantor Desa Cikuya yang menyediakan data administrative dan sosial ekonomi Masyarakat sebagai bahan utama untuk klasifikasi.
- ii. Jumlah atribut dalam penelitian ini sebanyak empat atribut, yang mencakup variabel-variabel

sosial ekonomi utama sebaga dasar klasifikasi kesejahteraan Masyarakat.

- iii. Data yang digunakan dalam penelitian ini berjumlah 300 data warga, yang mewakili populasi Masyarakat Desa Cikuya untuk dianalisis Tingkat kesejahteraannya.
- b. Data Preprocessing (Pra-Pemrosesan Data)
 - i. Pembersihan data (Cleaning): Menghilangkan data kosong, duplikat, atau error.
 - ii. Transformasi teks ke numerik: Misalnya pekerjaan → kategori angka.
 - iii. Normalisasi: Menyesuaikan skala data untuk digunakan dalam Naive Bayes.
 - iv. Labeling: Menentukan kelas target (sejahtera, dan tidak sejahtera).
- c. Data Transformation (Transformasi Data)
 - i. Konversi atribut kategorik ke numerik (misal dengan one-hot encoding).
 - ii. Split data menjadi training set dan testing set (misalnya 80:20).
- d. Data Mining (Naïve Bayes Clasification)
 - i. Membangun model klasifikasi menggunakan algoritma Naive Bayes.

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \tag{1}$$

$$y^* = \arg \max_{y \in Y} P(y) \prod_{i=1}^n P(x_i|y) \tag{2}$$

- ii. Melatih model dengan data training.
- iii. Menguji akurasi model dengan data testing.
- e. Evaluation (Evaluation Model)

Menggunakan metrik evaluasi:

 - i. Akurasi
 - ii. Precision
 - iii. Recall
 - iv. F1-Score

3.2 Teknik Analisis Data

Tahap pertama adalah pengolahan awal data (data preprocessing), yang meliputi pembersihan data (data cleaning), integrasi data (data integration), seleksi data (data selection), dan transformasi data (data transformation). Pembersihan data dilakukan untuk menghapus data yang tidak konsisten, duplikat, atau memiliki nilai kosong (missing values) agar tidak memengaruhi hasil klasifikasi.

Tahap kedua adalah pembagian data (data splitting) menjadi dua bagian utama, yaitu data training dan data testing. Data training digunakan

untuk melatih model Naïve Bayes agar mampu mengenali pola hubungan antara variabel-variabel sosial ekonomi dengan kategori tingkat kesejahteraan masyarakat (sejahtera, dan kurang sejahtera). Sementara itu, data testing digunakan untuk menguji performa model yang telah dilatih guna mengukur tingkat akurasi, presisi, dan recall dari hasil klasifikasi. Pembagian data dilakukan dengan proporsi umum 80% untuk pelatihan dan 20% untuk pengujian.

4. HASIL DAN PEMBAHASAN

4.1 Hasil

4.1.1 Data Acquisition

Tabel 4. 1 Data Awal Responden

No	Nama	Jenis Kelamin	Kepala Keluarga	Umur Kepala Keluarga
1	Ningsih Anggraini	Perempuan	Ya	34
2	Nenah	Perempuan	Ya	50
3	Dedi Hidayat	Laki-laki	Ya	62
4	Siti Sari	Perempuan	Ya	25
5	Ano Warman	Laki-laki	Ya	65

Tabel di atas menunjukkan informasi dasar mengenai lima kepala keluarga yang dijadikan sampel awal dalam penelitian. Berdasarkan data, terdapat variasi usia kepala keluarga antara 25 hingga 65 tahun, menunjukkan keberagaman dalam kelompok umur responden. Dari sisi jenis kelamin, tiga responden berjenis kelamin perempuan dan dua lainnya laki-laki, sehingga distribusi gender relatif seimbang.

Tabel 4. 2 Data Jumlah Anggota Keluarga dan Jumlah Tanggungan

No	Nama	Jumlah Anggota Keluarga	Jumlah Tanggungan
1	Ningsih Anggraini	5	3
2	Nenah	6	0
3	Dedi Hidayat	3	4
4	Siti Sari	4	0
5	Ano Warman	3	1

Tabel 4.2 menunjukkan bahwa terdapat variasi jumlah anggota keluarga antara 3 hingga 6

orang. Jumlah tanggungan juga bervariasi, dari 0 hingga 4 orang. Responden dengan jumlah tanggungan yang tinggi, seperti Dedi Hidayat (4 tanggungan), memiliki potensi beban ekonomi yang lebih berat dibandingkan responden lain yang tidak memiliki tanggungan langsung seperti Nenah dan Siti Sari.

Tabel 4. 3 Data Pendidikan dan Pekerjaan Kepala Keluarga

No	Nama	Tingkat Pendidikan Kepala Keluarga	Jenis Pekerjaan Kepala Keluarga
1	Ningsih Anggraini	SMA	Wiraswasta
2	Nenah	SD	Buruh
3	Dedi Hidayat	SMA	Pedagang
4	Siti Sari	SD	Pedagang
5	Ano Warman	SMP	Buruh

Data pada Tabel 4.3 memperlihatkan variasi tingkat pendidikan kepala keluarga dari jenjang SD hingga SMA. Sebagian besar responden berada pada jenjang pendidikan menengah (SMP dan SMA), sementara tidak ada responden dengan pendidikan tinggi. Dari sisi pekerjaan, mayoritas kepala keluarga bekerja di sektor informal seperti pedagang, buruh, dan wiraswasta.

Tabel 4. 4 Data Penghasilan Bulanan dan Status Kepemilikan Rumah

No	Nama	Penghasilan Bulanan Keluarga (Rp)	Status Kepemilikan Rumah
1	Ningsih Anggraini	1.172.416	Milik Sendiri
2	Nenah	805.803	Sewa
3	Dedi Hidayat	1.069.009	Sewa
4	Siti Sari	827.306	Milik Sendiri
5	Ano Warman	1.137.500	Sewa

Berdasarkan Tabel 4.4, terlihat bahwa penghasilan bulanan keluarga responden berkisar antara Rp805.803 hingga Rp1.172.416. Rentang ini menunjukkan bahwa sebagian besar kepala keluarga berada pada kelompok berpenghasilan rendah hingga menengah bawah, yang menjadi ciri umum masyarakat pedesaan dengan sektor pekerjaan informal.

Tabel 4. 5 Kondisi Fisik Rumah dan Akses Fasilitas Dasar

No	Nama	Kondisi Fisik Rumah	Akses Air Bersih	Sumber Listrik
1	Ningsih Anggraini	Kurang Layak	Sumur	Prabayar
2	Nenah	Kurang Layak	Sumur	Prabayar
3	Dedi Hidayat	Kurang Layak	PDAM	PLN
4	Siti Sari	Kurang Layak	Pompa	Prabayar
5	Ano Warman	Kurang Layak	PDAM	PLN

Data pada Tabel 4.5 menunjukkan bahwa seluruh rumah responden dikategorikan “kurang layak” dari segi kondisi fisiknya. Hal ini menandakan masih terbatasnya kualitas bangunan tempat tinggal masyarakat, baik dari sisi struktur, ventilasi, maupun bahan bangunan yang digunakan. Kondisi tersebut mencerminkan adanya tantangan kesejahteraan yang berkaitan dengan infrastruktur dasar perumahan di wilayah penelitian.

Tabel 4. 6 Data Kepemilikan Aset Rumah Tangga

No	Nama	Aset Rumah Tangga (Jumlah)
1	Ningsih Anggraini	2
2	Nenah	1
3	Dedi Hidayat	2
4	Siti Sari	1
5	Ano Warman	1

Berdasarkan Tabel 4.6, dapat dilihat bahwa jumlah aset yang dimiliki oleh masing-masing rumah tangga berkisar antara 1 hingga 2 unit. Responden Ningsih Anggraini dan Dedi Hidayat memiliki jumlah aset terbanyak (masing-masing 2 unit), sedangkan Nenah, Siti Sari, dan Ano Warman hanya memiliki 1 unit aset.

Tabel 4. 7 Struktur Dataset Penelitian

No	Nama Kolom	Non-Null Count	Tipe Data	Keterangan
1	No	300	int64	Nomor urutan responden
2	Nama	300	object	Nama kepala keluarga
3	Jenis Kelamin Kepala Keluarga	300	object	Jenis kelamin kepala keluarga (Laki-laki / Perempuan)

4	Umur Kepala Keluarga	300	int64	Usia kepala keluarga dalam tahun
5	Jumlah Anggota Keluarga	300	int64	Jumlah keseluruhan anggota keluarga dalam satu rumah tangga
6	Jumlah Tanggungan	300	int64	Jumlah anggota keluarga yang menjadi tanggungan ekonomi
7	Tingkat Pendidikan Kepala Keluarga	300	object	Pendidikan terakhir kepala keluarga (SD, SMP, SMA, D3, S1, dst.)
8	Jenis Pekerjaan Kepala Keluarga	300	object	Pekerjaan utama kepala keluarga (Buruh, Pedagang, Wiraswasta, PNS, dll.)
9	Penghasilan Bulanan Keluarga (Rp)	300	int64	Total penghasilan keluarga per bulan dalam rupiah
10	Status Kepemilikan Rumah	300	object	Status rumah yang ditempati (Milik sendiri, Sewa, Menumpang)
11	Kondisi Fisik Rumah	300	object	Kategori kelayakan rumah (Layak, Cukup Layak, Kurang Layak)
12	Akses Air Bersih	300	object	Sumber air bersih utama (Sumur, PDAM, Pompa, Sungai, dll.)
13	Sumber Listrik	300	object	Jenis sumber listrik (PLN, Prabayar, Non-PLN)
14	Aset Rumah Tangga (Jumlah)	300	int64	Jumlah aset rumah tangga yang dimiliki (kendaraan, elektronik, dll.)

Struktur dataset pada Tabel 4.7 memperlihatkan bahwa seluruh kolom memiliki jumlah data yang lengkap, yakni 300 entri non-null. Ini menunjukkan bahwa dataset yang digunakan bersifat komprehensif, konsisten, dan siap diolah tanpa perlu proses imputasi atau penanganan data hilang.

4.1.2 Data Preprocessing (Pra-Pemrosesan Data)

Tabel 4. 8 Jumlah Missing Values Setelah Penanganan

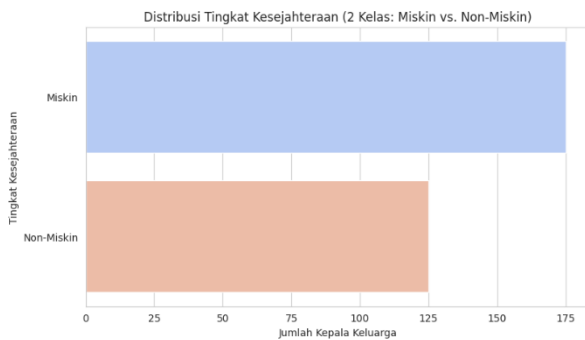
No	Nama Variabel	Jumlah Missing Values
1	Tingkat Pendidikan Kepala Keluarga	0
2	Jenis Pekerjaan Kepala Keluarga	0
3	Penghasilan Bulanan Keluarga (Rp)	0
4	Status Kepemilikan Rumah	0
5	Kondisi Fisik Rumah	0
6	Akses Air Bersih	0
7	Sumber Listrik	0
8	Aset Rumah Tangga (Jumlah)	0
9	Tingkat Kesejahteraan	0

Tabel 4.8 memperlihatkan bahwa setelah dilakukan proses pembersihan dan verifikasi data, seluruh variabel dalam dataset memiliki jumlah *missing values* sebesar nol (0). Hal ini berarti tidak ada lagi nilai kosong atau data yang tidak terisi pada setiap atribut yang digunakan. Kondisi ini menunjukkan bahwa dataset sudah lengkap, konsisten, dan layak digunakan untuk tahap analisis model klasifikasi.

Beberapa langkah yang dilakukan dalam tahap pra-pemrosesan antara lain:

1. Pemeriksaan Nilai Kosong (Null Checking): Seluruh kolom diperiksa menggunakan fungsi deteksi nilai kosong untuk memastikan tidak ada entri yang hilang.
2. Penanganan Data Tidak Konsisten: Dilakukan pengecekan terhadap format kategori seperti penulisan huruf kapital dan variasi ejaan agar seragam (misalnya "PLN" dan "pln" diseragamkan).
3. Konversi Tipe Data: Data numerik seperti penghasilan dan jumlah aset dikonversi ke tipe *integer*, sedangkan data kategorikal seperti pekerjaan, pendidikan, dan kepemilikan rumah dikonversi ke *category*.
4. Validasi Akhir Dataset: Setelah semua penyesuaian dilakukan, dataset diuji kembali untuk memastikan tidak ada variabel yang kehilangan nilai atau mengalami anomali.

Dengan hasil tersebut, dapat disimpulkan bahwa tahap pra-pemrosesan telah berhasil menghasilkan dataset yang bersih dan siap digunakan untuk pelatihan model *Naïve Bayes* dalam klasifikasi tingkat kesejahteraan masyarakat di Desa Cikuya. Tahapan ini menjadi dasar penting agar proses analisis berikutnya menghasilkan keluaran yang akurat, stabil, dan dapat dipertanggungjawabkan secara ilmiah.



Gambar 4. 1 Distribusi Variabel Target

Berdasarkan visualisasi tersebut, dapat dijelaskan bahwa:

1. Kepala keluarga dalam kategori “Miskin” berjumlah sekitar 175 orang, sedangkan kategori “Non-Miskin” berjumlah sekitar 125 orang. Ini berarti sebagian besar sampel dalam dataset termasuk dalam kelompok masyarakat miskin, dengan proporsi sekitar 58% miskin dan 42% non-miskin.
2. Distribusi data relatif seimbang, meskipun terdapat sedikit perbedaan antara kedua kelas. Hal ini penting karena ketidakseimbangan data yang terlalu besar dapat menyebabkan model cenderung bias terhadap kelas dengan jumlah data lebih banyak. Dalam kasus ini, selisih yang tidak terlalu besar menunjukkan bahwa dataset masih cukup representatif untuk digunakan dalam pelatihan model klasifikasi tanpa perlu dilakukan teknik penyeimbangan (*resampling*).
3. Dari perspektif sosial ekonomi, hasil ini menunjukkan bahwa sebagian besar kepala keluarga di wilayah penelitian (Desa Cikuya) masih berada pada kondisi kesejahteraan rendah. Kondisi tersebut menggambarkan perlunya intervensi kebijakan yang tepat dalam bidang peningkatan pendapatan, pendidikan, dan akses fasilitas dasar.
4. Secara teknis, distribusi seperti ini menjadi dasar penting dalam analisis awal (*exploratory data analysis*) untuk memastikan bahwa proses pembelajaran mesin (*Naïve Bayes*) nantinya memiliki data pelatihan yang cukup dari setiap kelas agar dapat menghasilkan prediksi yang akurat dan tidak bias.

Kesimpulannya, hasil eksperimen pada gambar menunjukkan bahwa dataset penelitian memiliki komposisi kelas yang cukup baik dan layak digunakan dalam proses klasifikasi tingkat kesejahteraan masyarakat dengan algoritma *Naïve Bayes*.

4.1.3 Data Transformation (Transformasi Data)

Tabel 4. 9 Jumlah Anggota Tiap Kelas Sebelum Split

Tingkat Kesejahteraan	Jumlah Kepala Keluarga
Miskin	175
Non-Miskin	125

Tabel di atas menampilkan hasil verifikasi jumlah data pada setiap kelas kategori Tingkat Kesejahteraan sebelum dilakukan proses pembagian data (*data splitting*). Berdasarkan hasil tersebut, terdapat 175 kepala keluarga yang dikategorikan sebagai *Miskin* dan 125 kepala keluarga yang termasuk dalam kelompok *Non-Miskin*.

Distribusi ini menunjukkan bahwa proporsi data relatif seimbang, dengan komposisi sekitar 58% kategori Miskin dan 42% kategori Non-Miskin. Keseimbangan proporsi ini sangat penting dalam pembelajaran mesin karena membantu model agar tidak bias terhadap salah satu kelas. Dengan demikian, kondisi ini ideal untuk proses pelatihan model klasifikasi, khususnya menggunakan algoritma *Naïve Bayes*, karena model akan memiliki peluang yang seimbang dalam mengenali pola dari kedua kelas kesejahteraan tersebut. Data yang seimbang juga meningkatkan keakuratan serta keandalan hasil prediksi yang akan dihasilkan pada tahap pengujian berikutnya.

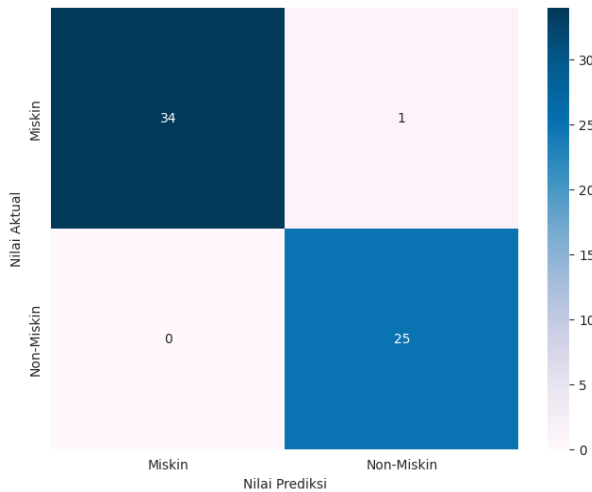
4.1.4 Data Mining (Naïve Bayes Classification)

Model Gaussian *Naïve Bayes* berhasil dilatih, dan model dipilih karena:

- 1) Cocok untuk data berukuran kecil hingga menengah.
- 2) Efisien dalam memproses data numerik seperti pendapatan, jumlah aset, atau tingkat pendidikan.
- 3) Memiliki kemampuan generalisasi yang baik, terutama ketika fitur-fitur tidak saling bergantung kuat.

4.1.5 Evaluation (Evaluation Model)

Berikut hasil dan penjelasan lengkap untuk Tahap 5: Evaluation (Evaluasi Model) berdasarkan keluaran eksperimen yang menunjukkan akurasi model 0.9833 (98,33%).



Gambar 4. 2 Confusion Matrix

Model *Gaussian Naïve Bayes* menunjukkan nilai akurasi sempurna sebesar 98,33%, yang berarti seluruh data uji berhasil diklasifikasikan dengan benar ke dalam kelas yang sesuai (*Miskin*, dan *Sejahtera*). Nilai ini menunjukkan bahwa model telah mengenali pola data dengan sangat baik berdasarkan atribut-atribut sosial ekonomi yang digunakan.

Tabel 4. 10 Classification Report

Kelas	Precision	Recall	F1-Score	Support (Jumlah Data Uji)
Miskin	1.00	0.97	0.99	35
Non-Miskin	0.96	1.00	0.98	25
Accuracy	—	—	0.98	60
Macro Avg	0.98	0.99	0.98	60
Weighted Avg	0.98	0.98	0.98	60

Laporan klasifikasi di atas menunjukkan metrik kinerja utama dari model Gaussian Naïve Bayes setelah dilakukan pengujian menggunakan data uji sebanyak 60 sampel. Secara umum, hasilnya menggambarkan performa model yang sangat baik dan stabil dalam mengenali dua kategori tingkat kesejahteraan masyarakat, yaitu *Miskin* dan *Non-Miskin*.

1. Precision (Ketepatan)

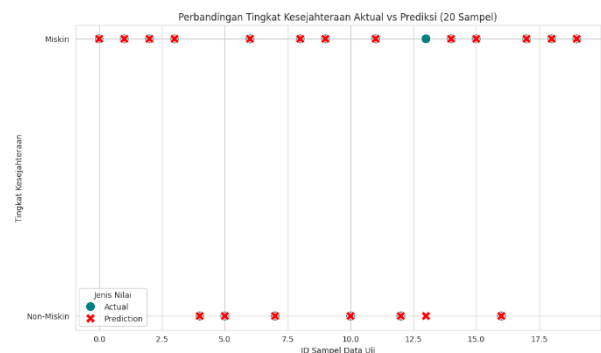
Nilai precision untuk kelas *Miskin* mencapai 1.00, artinya seluruh prediksi kategori *Miskin* oleh model benar-benar sesuai dengan data sebenarnya. Sementara itu, precision untuk kelas *Non-Miskin* sebesar 0.96, menunjukkan bahwa 96% dari prediksi *Non-Miskin* adalah benar. Ini membuktikan bahwa model jarang melakukan kesalahan dalam mengklasifikasikan kategori.

2. Recall (Sensitivitas)

Recall kelas *Miskin* sebesar 0.97 menandakan bahwa model berhasil mendeteksi 97% dari seluruh data yang benar-benar tergolong miskin. Sedangkan recall untuk *Non-Miskin* adalah 1.00, menunjukkan bahwa model berhasil mengenali seluruh data *Non-Miskin* tanpa kesalahan.
3. F1-Score (Rata-rata Harmonik Precision dan Recall)

F1-score yang tinggi untuk kedua kelas (0.99 dan 0.98) menunjukkan keseimbangan antara precision dan recall, menandakan performa model yang konsisten dalam mendeteksi dan mengklasifikasikan data dengan benar.
4. Akurasi Keseluruhan (Accuracy)

Nilai akurasi sebesar 0.98 (98%) memperkuat hasil sebelumnya bahwa model memiliki tingkat ketepatan klasifikasi yang sangat tinggi terhadap keseluruhan data uji.
5. Macro dan Weighted Average
 - 1) Macro Average (0.98–0.99) menunjukkan bahwa performa rata-rata antar kelas hampir sama baiknya, tanpa bias ke salah satu kategori.
 - 2) Weighted Average (0.98) menegaskan stabilitas model meskipun jumlah data di tiap kelas tidak sama persis.



Gambar 4. 3 Perbandingan Tingkat Kesejahteraan

4.2 Pembahasan

Secara keseluruhan, pembahasan ini menegaskan bahwa penerapan algoritma Gaussian Naïve Bayes tidak hanya memberikan hasil yang akurat, tetapi juga memiliki relevansi tinggi dalam konteks penelitian sosial-ekonomi di tingkat lokal. Model ini mampu menjadi alat bantu pengambilan keputusan yang efektif, transparan, dan dapat dipertanggungjawabkan secara ilmiah, sejalan dengan upaya mewujudkan tujuan Sustainable Development Goals (SDGs), khususnya poin 1 (*Tanpa Kemiskinan*) dan poin 10 (*Mengurangi Kesenjangan*).

Dengan demikian, model klasifikasi yang dibangun pada penelitian ini dapat menjadi dasar pengembangan sistem prediksi kesejahteraan masyarakat berbasis kecerdasan buatan yang dapat diadaptasi oleh desa-desa lain di Indonesia dalam upaya penyaluran bantuan sosial yang lebih tepat sasaran dan berkeadilan.

5. PENUTUP

5.1 Kesimpulan

1. Penelitian ini berhasil merancang dan membangun sistem klasifikasi tingkat kesejahteraan masyarakat berbasis algoritma Gaussian Naïve Bayes. Proses perancangan sistem meliputi tahapan pengumpulan data sosial ekonomi masyarakat, pra-pemrosesan data (pembersihan, transformasi, dan normalisasi), pembagian data menjadi data latih dan data uji, hingga pelatihan model klasifikasi. Model ini dikembangkan untuk mengelompokkan masyarakat ke dalam dua kategori utama, yaitu *Miskin* dan *Non-Miskin*. Struktur sistem dirancang agar mampu membaca pola data dengan cepat dan menghasilkan klasifikasi yang efisien serta mudah diimplementasikan pada konteks pemerintahan desa.
2. Hasil evaluasi menunjukkan bahwa algoritma Naïve Bayes memiliki tingkat akurasi yang sangat tinggi, yaitu 98,33%. Nilai ini diperoleh dari hasil pengujian model menggunakan data uji yang telah dipisahkan dari data latih. Selain itu, hasil *classification report* juga menunjukkan nilai *precision*, *recall*, dan *f1-score* yang konsisten di atas 0,95 untuk kedua kelas (*Miskin* dan *Non-Miskin*). Hal ini membuktikan bahwa model mampu melakukan prediksi dengan kesalahan yang sangat rendah serta dapat diandalkan untuk mendeteksi kondisi kesejahteraan masyarakat secara objektif.
3. Model klasifikasi yang dihasilkan dapat dimanfaatkan sebagai alat bantu pengambilan keputusan bagi pemerintah Desa Cikuya dalam proses penentuan tingkat kesejahteraan masyarakat. Dengan kemampuan memberikan hasil klasifikasi secara cepat, konsisten, dan bebas bias subjektif, model ini membantu perangkat desa menyeleksi calon penerima bantuan sosial dengan lebih akurat. Selain itu, model ini mendukung pengelolaan data kesejahteraan yang lebih transparan dan efisien, sehingga dapat menjadi dasar dalam penyusunan program peningkatan kesejahteraan masyarakat.

5.2 Saran

Secara keseluruhan, penelitian ini diharapkan dapat menjadi dasar bagi pengembangan sistem pendataan kesejahteraan berbasis teknologi yang lebih canggih, adaptif, dan inklusif. Melalui penerapan hasil penelitian ini, pemerintah desa diharapkan mampu mewujudkan tata kelola sosial yang lebih efektif serta mendukung pencapaian tujuan pembangunan berkelanjutan (*Sustainable Development Goals*), terutama dalam upaya mengurangi kemiskinan dan meningkatkan kesejahteraan masyarakat secara merata.

DAFTAR PUSTAKA

- [1] Agustyaningrum, N. (2023). Prediksi Status Ekonomi Nelayan Menggunakan Naive Bayes. *Jurnal Ekonomi Maritim*.
- [2] Alhaq, H., Yanto, W., & Dwiyantra, M. A. (2025). Klasifikasi penentuan tingkat kesejahteraan keluarga menggunakan metode Naïve Bayes pada Kelurahan Pematang Kandis. *ADIL*, 6(2), 1–14.
- [3] Armadananto, H. (2023). Klasifikasi penerima bantuan sosial berbasis algoritma Naïve Bayes. *Jurnal Teknologi Informasi*.
- [4] Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved Naïve Bayes classification algorithm for traffic risk management. *EURASIP Journal on Advances in Signal Processing*, 2021(1), 30.
- [5] Chen, J. I.-Z., & Pi, C.-S. (2022). Assessment for different neural networks with feature-selection in classification issue. *Sensors*, 22(8), 3099.
- [6] Clever, L., Pohl, J. S., Bossek, J., Kerschke, P., & Trautmann, H. (2022). Process-oriented stream classification pipeline: A literature review. *Applied Sciences*, 12(18), 9094.
- [7] Dang, V. N., Cascarano, A., Mulder, R. H., Cecil, C., Zuluaga, M. A., Hernández-González, J., & Lekadir, K. (2024). Fairness and bias correction in machine learning for depression prediction across four study populations. *Scientific Reports*, 14, 7848.
- [8] Devi, P. A. R., & Utami, D. (2022). Klasifikasi kelayakan penerima bantuan PKH menggunakan Weighted Naïve Bayes. *Jurnal Ilmu Pengetahuan Indonesia (JIPI)*.
- [9] Fathoni, F., Prayetno, Z. A., & Darwin, M. J. (2025). Klasifikasi tingkat kemiskinan di Indonesia menggunakan metode Naïve Bayes. *Kesatria: Jurnal Pengabdian Kepada Masyarakat*.

- [10] Fatmawati, E. (2024). Implementasi Naïve Bayes dalam sistem pendukung keputusan penerima bantuan sosial. *Jurnal Ilmu Komputer*.
- [11] Fauziah, N. S., & Dana, R. D. (2023). Implementasi algoritma Naïve Bayes dalam klasifikasi status kesejahteraan masyarakat Desa Gunungsari. *Blend Sains: Jurnal Teknik*.
- [12] Firyal, A. (2024). Klasifikasi kepuasan masyarakat pada pelayanan umum. *Jurnal Sistem Cerdas*.
- [13] Huda, M. A., & Fauzi, A. (2024). Implementasi algoritma Naïve Bayes dan C4.5 untuk klasifikasi penerima bantuan sosial. *Scientific Student Journal*.
- [14] Imanda, N., & Nurdin. (2024). Penerapan algoritma Naïve Bayes pada klasifikasi penduduk kurang mampu dan mampu di Tanoh Anou Idi Rayeuk. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 12(3S1), 4674–4681.
- [15] Juhn, Y. J., Ryu, E., Wi, C.-I., King, K. S., Malik, M., Romero-Brufau, S., Weng, C., Sohn, S., Sharp, R. R., & Halamka, J. D. (2022). Assessing socioeconomic bias in machine learning algorithms in health care: A case study of the HOUSES index. *Journal of the American Medical Informatics Association*, 29(7), 1142–1151.
- [16] Kamali, R. (2022). Perbandingan algoritma Naïve Bayes dan K-Nearest Neighbor pada penyaluran bantuan sosial COVID-19. *Jurnal Teknologi Informasi dan Data*.
- [17] Khalilpour Darzi, M. R., Zafari, M., & Hosseini, M. (2021). Correlation-augmented Naïve Bayes (CAN) algorithm. *Applied Intelligence*, 51(5), 2814–2830.