

Prediksi Tingkat Kemiskinan Jawa Barat Menggunakan Naïve Bayes: Studi Implementasi Data Mining

Zulaeha¹, Wafha Zahra Mulqiya², Tyanshi Firli Mahrani³

^{1,2,3} Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

¹zulaeha.312210575@mhs.pelitabangsa.ac.id

²wafha.312210577@mhs.pelitabangsa.ac.id

³tyanshi.312210581@mhs.pelitabangsa.ac.id

Abstrak— Kemiskinan masih menjadi masalah sosial dan ekonomi yang signifikan di banyak daerah di Indonesia, termasuk Provinsi Jawa Barat. Untuk menanggulangnya, pendekatan berbasis data diperlukan agar kebijakan yang dibuat lebih tepat sasaran. Tujuan penelitian ini adalah untuk membuat model prediksi tingkat kemiskinan dengan menggunakan algoritma Naive Bayes dan memasukkannya ke dalam aplikasi berbasis web yang interaktif. Data yang digunakan berasal dari data tahun 2024 dari Badan Pusat Statistik (BPS) Jawa Barat, yang mencakup persentase kemiskinan di 31 kabupaten/kota. Data dikategorikan ke dalam tiga kategori: rendah (kurang dari 5%), sedang (lebih dari 5%–10%) dan tinggi (lebih dari 10%). Model dibuat menggunakan platform Google Colab dan dilengkapi dengan pustaka Python seperti scikit-learn. Antarmuka aplikasinya dibuat menggunakan framework Streamlit. Untuk menilai model, metrik akurasi, presisi, dan recall digunakan. Hasil pengujian menunjukkan bahwa model memiliki akurasi sebesar 86%, presisi sebesar 93%, dan recall sebesar 86%. Ini menunjukkan bahwa model memiliki performa klasifikasi yang cukup baik. Pengguna non-teknis, seperti lembaga pemerintah, dapat melakukan prediksi secara real-time dengan aplikasi ini dan melihat data dengan grafik interaktif. Studi ini menunjukkan bahwa algoritma Naive Bayes dapat digunakan secara efektif dalam sistem pendukung keputusan berbasis data mining untuk mengurangi kemiskinan daerah.

Kata Kunci—Prediksi Kemiskinan, Naive Bayes, Streamlit, Visualisasi Data, Aplikasi Berbasis Web

I. PENDAHULUAN

Salah satu masalah sosial dan ekonomi yang signifikan adalah kemiskinan, terutama di wilayah dengan populasi yang padat seperti Provinsi Jawa Barat. Kualitas hidup dan kesejahteraan masyarakat dapat terganggu oleh rendahnya tingkat pendapatan, terbatasnya lapangan kerja, dan kurangnya pembangunan di daerah [1]. Akibatnya, upaya preventif dan prediktif diperlukan untuk mengetahui tingkat kemiskinan di daerah sejak dini. Ini akan digunakan sebagai acuan untuk membuat kebijakan dan perencanaan pembangunan daerah yang lebih tepat sasaran [2].

Peluang untuk mengembangkan sistem prediksi tingkat kemiskinan yang akurat semakin meningkat seiring dengan perkembangan ilmu data dan teknologi informasi [3]. Salah satu metode yang dapat digunakan untuk kebutuhan tersebut adalah algoritma Naive Bayes, yang dapat mengklasifikasi

data dengan cepat dan efisien berdasarkan pola nilai-nilai yang digunakan sebagai parameter [4].

Pengembangan aplikasi prediksi tingkat kemiskinan dapat dilakukan secara efektif dan interaktif dengan menggunakan framework Streamlit. Aplikasi ini tidak hanya memungkinkan pengguna untuk memprediksi tingkat kemiskinan suatu wilayah secara realtime, tetapi juga dapat menampilkan visualisasi data yang memudahkan interpretasi pola dan tren dari nilai-nilai yang digunakan. [5].

Dalam penelitian ini, metode Naive Bayes digunakan untuk membuat model prediksi tingkat kemiskinan yang dapat membantu dalam mengkategorikan wilayah menjadi kategori rendah, sedang, atau tinggi [6]. Penelitian ini tidak hanya berfokus pada membuat alat bantu visual, tetapi juga menguji seberapa baik model klasifikasi bekerja dengan data dari Badan Pusat Statistik (BPS) Jawa Barat [7].

Tujuan penelitian ini adalah untuk menerapkan dan mengevaluasi algoritma Naive Bayes dalam proses klasifikasi wilayah Provinsi Jawa Barat berdasarkan tingkat kemiskinan. Tujuan utama penelitian ini adalah untuk membuat model prediksi yang dapat menentukan status kemiskinan berdasarkan nilai persentase yang tercatat untuk setiap wilayah. Mengukur akurasi, presisi, dan recall model juga merupakan tujuan penelitian ini. Visualisasi hasil prediksi interaktif juga membantu pengguna dari berbagai latar belakang, termasuk pemangku kebijakan, memahami data.

Dalam bidang sosial ekonomi menggunakan metode Naive Bayes. Manfaat ilmiah dari penelitian ini terutama terletak pada proses mengklasifikasikan data kemiskinan berdasarkan satu ciri. Penelitian ini menunjukkan bahwa algoritma yang sederhana sekalipun dapat menghasilkan hasil klasifikasi yang akurat dan bermanfaat dengan data BPS dan pendekatan statistik berbasis distribusi Gaussian. Penelitian ini juga menghasilkan model yang dapat digunakan sebagai alat pengambilan keputusan dengan visualisasi data untuk menentukan wilayah mana yang paling membutuhkan intervensi sosial. Diharapkan bahwa aplikasi prediksi tingkat kemiskinan ini akan membantu pihak terkait dalam membuat kebijakan, menjadi langkah strategis untuk mempercepat pemerataan pembangunan, mengurangi tingkat kemiskinan di Provinsi Jawa Barat dan dapat membantu proses pengambilan keputusan berdasarkan data di sektor pemerintahan dan pembangunan sosial.

II. TINJAUAN PUSTAKA

Penelitian yang dilakukan oleh Fuadi Ahmad, Citra Suarna, dan Gusti Dwilestari menunjukkan dalam jurnal "Klasifikasi Data Kemiskinan Menggunakan Metode Naive Bayes untuk Mengetahui Tingkat Kemiskinan: Studi Kasus Desa Karangasem, Majalengka" bahwa algoritma Naive Bayes dengan efektif menggunakan usia, pendidikan, pekerjaan, dan tanggungan keluarga untuk mengklasifikasikan tingkat kemiskinan. Dengan akurasi sebesar 92 persen dan recall sebesar 86 persen, model penelitian menunjukkan kemampuan yang luar biasa untuk menemukan kategori kemiskinan. Penelitian ini menunjukkan bahwa metode Naive Bayes berfungsi dengan baik dengan data yang kompleks tetapi terstruktur, terutama dalam konteks lokal seperti desa atau kecamatan [8].

Selain itu, efektivitas metode ini didukung oleh penelitian yang ditulis oleh F. Fathoni dkk. dalam jurnal yang berjudul "Penerapan Naive Bayes pada Klasifikasi Tingkat Kemiskinan di Indonesia Menggunakan Data Sosial-Ekonomi". Dengan menggunakan dataset berskala nasional, mereka menemukan akurasi 92,4%. Meskipun hasilnya sangat menjanjikan, mereka menekankan bahwa model masih memiliki kelemahan dalam mengidentifikasi kelompok minoritas, seperti kelompok yang "ekstrem miskin". Hal ini menggarisbawahi pentingnya pengolahan data yang seimbang dan kebutuhan akan pengembangan model tambahan untuk lebih menyadari kelas minoritas [9].

Penelitian yang diterbitkan dalam jurnal berjudul "Prediksi Angka Kemiskinan pada Desa Kemang Bejalu Menggunakan Metode Naive Bayes" oleh Arda Damayanti dkk. berfokus pada penggunaan algoritma ini pada tingkat mikro, yaitu desa. Mereka menunjukkan bahwa Naive Bayes masih mampu memberikan hasil klasifikasi yang sangat akurat (86%–90%) meskipun data berskala kecil dan terbatas. Penelitian ini menunjukkan bahwa metode ini dapat berfungsi dengan baik di berbagai skala, mulai dari desa hingga nasional, dengan preprocessing data yang baik [10].

Dari ketiga penelitian tersebut, dapat disimpulkan bahwa metode Naive Bayes memiliki keunggulan dalam hal efisiensi komputasi, kemudahan implementasi, dan akurasi yang memadai untuk berbagai konteks pengklasifikasian tingkat kemiskinan. Oleh karena itu, metode ini dapat digunakan untuk mengembangkan aplikasi prediksi kemiskinan seperti yang dilakukan dalam penelitian ini. Selain itu, hasil prediksi dapat disajikan dengan lebih informatif dan interaktif dengan bantuan antarmuka pengguna seperti Streamlit. Ini membuatnya lebih mungkin digunakan di tempat kerja di lembaga sosial dan pemerintah.

III. METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif, menggunakan metode klasifikasi berbasis algoritma Naive Bayes pada data kemiskinan wilayah Provinsi Jawa Barat. Berdasarkan data persentase kemiskinan dari Badan Pusat Statistik (BPS) tahun 2024, penelitian ini

berfokus pada pembuatan model prediksi untuk mengkategorikan daerah ke dalam tiga kategori kemiskinan: rendah, sedang, dan tinggi. Pembangunan dan pengujian model dilakukan menggunakan platform pemrograman Google Colab, yang menyediakan lingkungan berbasis cloud dan mendukung pustaka Python seperti scikit-learn, pandas, dan matplotlib. Sementara itu, framework Streamlit digunakan untuk mengembangkan aplikasi prediksi berbasis web yang interaktif untuk visualisasi dan pengembangan antarmuka pengguna. Ini memungkinkan pengguna non-teknis melihat hasil klasifikasi secara real-time.

A. Dataset

Dataset yang digunakan dalam pengembangan aplikasi ini diambil dari publikasi resmi Badan Pusat Statistik (BPS) Jawa Barat tahun 2024 dengan atribut sebagai berikut:

- Kota / Kabupaten: Nama daerah di Provinsi Jawa Barat
- Persentase Kemiskinan: Nilai persentase tingkat kemiskinan daerah

Atribut Persentase Kemiskinan digunakan sebagai fitur numerik tunggal, sedangkan variabel target terdiri dari tiga kategori:

- Rendah: Persentase $\leq 5\%$
- Sedang: Persentase $> 5\text{--}10\%$
- Tinggi: Persentase $> 10\%$

Data ini digunakan untuk melatih dan menguji model klasifikasi Naive Bayes guna memprediksi status tingkat kemiskinan daerah dengan tingkat akurasi yang dapat diandalkan.

B. Processing Data

Sebelum data digunakan untuk pelatihan model, beberapa tahap preprocessing dilakukan untuk memastikan bahwa data bersih dan siap digunakan. Tahap-tahap ini meliputi:

- Pemeriksaan Duplikasi dan Nilai Kosong (Missing Values): Untuk menjaga kualitas dataset, baris duplikat atau nilai kosong di kolom Persentase Kemiskinan dicek. Jika ditemukan, data dihapus.
- Pembersihan Data: Agar model dapat memprosesnya, format nilai persentase harus dalam tipe numerik (float) dan tidak mengandung simbol seperti "%".

Labeling Target: kolom target diklasifikasikan menjadi tiga kategori:

- Rendah: $\leq 5\%$
- Sedang: $> 5\% - 10\%$
- Tinggi: $> 10\%$

Kategori ini digunakan untuk mengubah nilai numerik menjadi label klasifikasi.

Dataset yang telah dibersihkan digunakan sebagai dasar untuk pelatihan dan pengujian model klasifikasi.

C. Pembagian Dataset

Setelah proses preprocessing selesai, langkah selanjutnya adalah membagi dataset menjadi data pelatihan (training set) dan data pengujian (testing set). Tujuan dari pembagian ini adalah agar model dapat dilatih dengan sedikit data sebelum diuji dengan data yang belum pernah dilihat sebelumnya.

Proses pembagian dilakukan menggunakan fungsi `train_test_split` dari Pustaka `scikit-learn`, dengan konfigurasi sebagai berikut:

- 80% data digunakan sebagai training set, yaitu data yang digunakan untuk melatih model agar dapat mengenali pola dan membuat prediksi.
- 20% data digunakan sebagai testing set, yaitu data yang digunakan untuk menguji beberapa baik model dapat menggeneralisasi atau memprediksi data baru yang belum dilihat.

Proses pembagian konsisten dan dapat diulang dengan menggunakan parameter `random_state=42`. Angka-angka ini dianggap acak dan "dikunci", sehingga pembagian data tidak berubah setiap kali program berjalan.

Contoh code pembagian dataset :

```
X = data_clean[["Persentase"]]  
y = data_clean["Kategori"]  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
```

Gbr. 1 Code Pembagian Dataset

Interpretasi:

- X adalah fitur (nilai presentase kemiskinan).
- Y adalah label target (kategori kemiskinan: Rendah, Sedang, Tinggi).

Selain itu, karena jumlah data dari masing-masing kategori cukup seimbang (seimbang) pada saat ini, metode stratifikasi (`stratify=y`) tidak digunakan. Namun, jika dataset berikutnya menunjukkan ketidakseimbangan jumlah antara kategori, seperti terlalu banyak data kategori "Sedang" dan sedikit data kategori "Tinggi", maka stratifikasi diperlukan untuk memastikan proporsi kategori tetap konsisten dengan data pelatihan dan pengujian. Dengan pembagian seperti ini, diharapkan model dapat mengidentifikasi pola dalam data pelatihan dan memprediksi dengan akurat data baru.

D. Pelatihan dan Evaluasi Model

Setelah pembagian dataset selesai, langkah berikutnya adalah menginstruksikan model klasifikasi untuk memprediksi tingkat kemiskinan. Salah satu algoritma klasifikasi berbasis probabilistik yang paling umum digunakan untuk data numerik yang diasumsikan mengikuti distribusi normal (*Gaussian*) adalah Gaussian Naive Bayes, model yang digunakan dalam proyek ini.

Proses pelatihan menggunakan data latih (`X_train` dan `y_train`) yang diperoleh dari proses `train_test_split`. Model dilatih untuk menemukan pola hubungan antara nilai persentase kemiskinan, fitur input, dan kategori tingkat kemiskinan, target output, yang terdiri dari rendah, sedang, atau tinggi.

Setelah model dilatih, proses evaluasi dilakukan untuk mengetahui seberapa baik kinerjanya dalam klasifikasi. Di

antara metrik standar yang ditemukan dalam pustaka `sklearn.metrics` yang digunakan untuk melakukan evaluasi adalah:

- Accuracy Score adalah nilai yang digunakan untuk menghitung proporsi prediksi yang benar dibandingkan dengan total data uji.
- Precision Score adalah nilai yang digunakan untuk mengukur ketepatan prediksi untuk setiap kategori.
- Recall Score adalah ukuran seberapa baik model dapat mengumpulkan semua data dari kategori sebenarnya.

Tabel. 1 Hasil Evaluasi Model Naïve Bayes

| Mterik Evaluasi | Nilai |
|-----------------|------------|
| Akurasi | 0.86 (86%) |
| Presisi | 0.93 (93%) |
| Recall | 0.86 (86%) |

Hasil evaluasi menunjukkan bahwa model memiliki nilai akurasi sebesar 86%, yang menunjukkan bahwa 86% dari semua prediksi berada dalam kategori yang benar; nilai ketepatan sebesar 93%, yang menunjukkan bahwa prediksi yang dibuat oleh model sangat tepat dan jarang salah; dan nilai recall sebesar 86%, yang menunjukkan bahwa model cukup andal untuk menemukan dan mengklasifikasikan semua kategori tingkat kemiskinan yang ada dalam dataset.

Dengan menggabungkan ketiga metrik tersebut, dapat disimpulkan bahwa model Gaussian Naive Bayes yang digunakan berkinerja seimbang dan layak digunakan dalam sistem klasifikasi kemiskinan yang berbasis web.

E. Implementasi Aplikasi

Aplikasi prediksi tingkat kemiskinan ini dibuat dengan framework `Streamlit`, sebuah pustaka Python yang dirancang untuk membangun antarmuka web berbasis data yang cepat dan interaktif. Implementasi aplikasi ini terdiri dari beberapa komponen utama yang masing-masing menangani input, prediksi, visualisasi, dan ekspor hasil.

1. Input Data

Pengguna menggunakan kolom input untuk memasukkan nilai persentase kemiskinan suatu daerah. Nilai-nilai ini digunakan dalam model Gaussian Naive Bayes untuk melakukan klasifikasi. Untuk memastikan bahwa nilai yang dimasukkan bertipe numerik, `Streamlit` menyediakan antarmuka input berbentuk `st.number_input`.

2. Proses Prediksi

Sistem akan menjalankan fungsi klasifikasi menggunakan model Naïve Bayes yang telah dilatih sebelumnya setelah pengguna memberikan nilai. Model menggunakan distribusi Gaussian untuk menghitung probabilitas dari tiga kelas: Rendah, Sedang, dan Tinggi. Kelas dengan probabilitas tertinggi dipilih sebagai hasil prediksi. Label status dan indikator warna menunjukkan hasil prediksi

langsung. Label status menunjukkan tingkat rendah, kuning, dan tinggi.

3. Visualisasi Data

Untuk membantu analisis, aplikasi menyediakan dua jenis grafik utama:

- Diagram lingkaran (Pie Chart), menunjukkan distribusi klasifikasi kemiskinan di setiap wilayah dalam dataset.
- Diagram Batang (Bar Chart), mengurutkan wilayah berdasarkan tingkat kemiskinan tertinggi. Untuk membuat tampilan yang informatif dan menarik, pustaka matplotlib dan seaborn digunakan, yang diintegrasikan dengan Streamlit.

4. Table Data dan Pencarian

Dengan menggunakan `st.dataframe()`, tabel yang berisi nama daerah, nilai persentase kemiskinan, dan hasil klasifikasi ditampilkan. Untuk melakukan analisis lebih lanjut, pengguna dapat melakukan pencarian atau pengurutan data untuk melihat wilayah tertentu.

5. Ekspor Hasil ke CSV

Selain itu, aplikasi memiliki tombol "Download CSV", yang memungkinkan pengguna mengunduh data hasil klasifikasi dalam format.csv. Fitur ini berguna untuk dokumentasi internal atau pelaporan resmi untuk dinas atau lembaga terkait.

6. Struktur Modular Kode

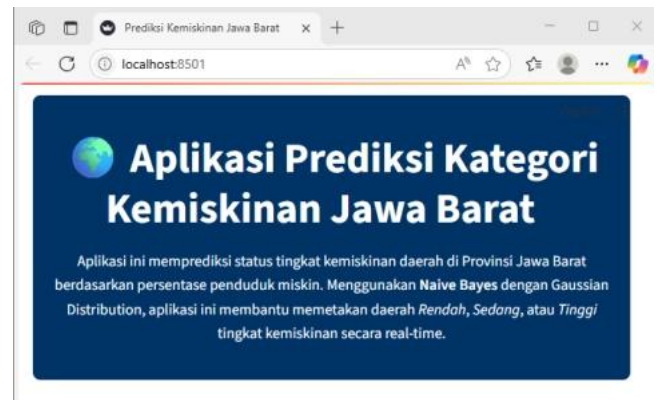
Kode program disusun secara modular untuk membedakan logika prediksi, preprocessing, visualisasi, dan antarmuka. Ini memudahkan pemeliharaan dan pengembangan lebih lanjut, seperti penambahan fitur atau penggantian model klasifikasi di masa depan.

Aplikasi ini dapat digunakan oleh pengguna yang tidak terlalu teknis, seperti aparat pemerintah daerah atau petugas lapangan, untuk membantu menemukan wilayah dengan tingkat kemiskinan yang memerlukan intervensi karena antarmuka yang mudah dipahami dan proses yang berjalan secara real-time.

IV. HASIL DAN PEMBAHASAN

A. Tampilan Aplikasi

Aplikasi prediksi tingkat kemiskinan yang dikembangkan menggunakan Streamlit dan algoritma *Gaussian Naive Bayes* ini memiliki tampilan antarmuka (UI) yang interaktif, informatif, dan mudah digunakan. Pengguna hanya perlu memasukkan nilai persentase kemiskinan dari suatu daerah, dan sistem akan secara otomatis memberikan klasifikasi status kemiskinan: Rendah, Sedang, atau Tinggi. Adapun gambar-gambar tampilan aplikasinya sebagai berikut:



Gbr. 2 Tampilan about



Gbr. 3 Tampilan Input dan Prediksi Status Kemiskinan

Pada **Gambar 3**. Adalah gambar ini menampilkan halaman awal aplikasi, di mana pengguna akan menemukan kolom input untuk memasukkan **nilai persentase kemiskinan** daerah. Setelah pengguna menekan tombol "Prediksi Status", sistem akan langsung memproses data dan mengklasifikasikan daerah tersebut berdasarkan model Gaussian Naïve Bayes. Hasil klasifikasi ditampilkan dalam bentuk **label status** beserta **warna kategori** (misalnya: hijau untuk "Rendah", kuning untuk "Sedang", merah untuk "Tinggi"). Sebagai contoh, jika pengguna memasukkan angka **4,01%**, maka sistem akan menampilkan kategori **Rendah**, karena nilai tersebut berada di bawah ambang batas 5%. Tampilan ini sangat membantu dalam skenario nyata, di mana petugas cukup memasukkan data persentase dan langsung mendapatkan hasil klasifikasi tanpa perlu menghitung atau membuka spreadsheet manual.

Daftar Daerah dan Kategorinya

| | Kota | Persentase | Kategori |
|---|---------------------|------------|----------|
| 0 | Provinsi Jawa Barat | 7.46 | Sedang |
| 1 | Bogor | 7.05 | Sedang |
| 2 | Sukabumi | 6.87 | Sedang |
| 3 | Cianjur | 10.14 | Tinggi |
| 4 | Bandung | 6.19 | Sedang |
| 5 | Garut | 9.68 | Sedang |
| 6 | Tasikmalaya | 10.23 | Tinggi |
| 7 | Ciamis | 7.39 | Sedang |
| 8 | Kuningan | 11.88 | Tinggi |
| 9 | Cirebon | 11 | Tinggi |

Gbr.4 Tabel Data Wilayah dan Hasil Prediksi

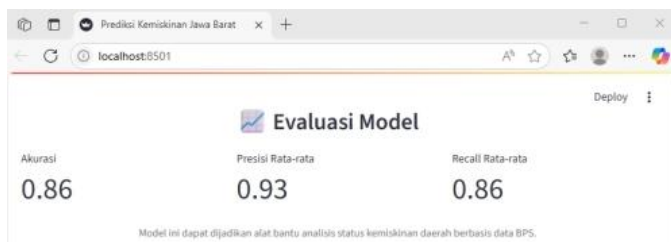
Pada **Gambar 4**. Adalah tampilan tabel ini menyajikan



daftar lengkap nama-nama kabupaten/kota di Provinsi Jawa Barat, disertai dengan **nilai persentase kemiskinan** masing-masing serta hasil prediksi status kemiskinan berdasarkan klasifikasi model. Fitur pencarian dan pengurutan (sort & search) disediakan agar pengguna dapat dengan cepat menemukan data tertentu. Tabel ini menjadi bagian penting dalam aplikasi karena memungkinkan pemantauan dan perbandingan kondisi antar wilayah. Dengan demikian, pengguna dapat mengenali daerah-daerah yang memerlukan perhatian lebih lanjut, seperti wilayah dengan persentase tinggi tetapi belum menerima intervensi yang sesuai.

Gbr. 5 Fitur Ekspor Data

Berikut **Gambar 5**. Adalah salah satu fitur unggulan dalam aplikasi adalah kemampuan untuk mengunduh hasil prediksi dalam format CSV. Setelah pengguna menjalankan klasifikasi untuk semua daerah, mereka dapat mengklik tombol “Download CSV” untuk menyimpan data yang berisi:



- Nama wilayah
- Nilai persentase kemiskinan
- Status klasifikasi

Fitur ini mendukung kebutuhan pelaporan instansi seperti Dinas Sosial, Bappeda, atau pemerintahan daerah yang membutuhkan dokumentasi hasil dalam format digital yang bisa dibuka di Excel atau Google Sheets.

B. Evaluasi Model

Model klasifikasi yang digunakan adalah **Gaussian Naïve Bayes**. Berdasarkan hasil pelatihan dan pengujian pada dataset BPS 2024, diperoleh metrik performa sebagai berikut:

Adapun gambar evaluasi Model pada gambar 6.

Gbr. 6 Evaluasi Model

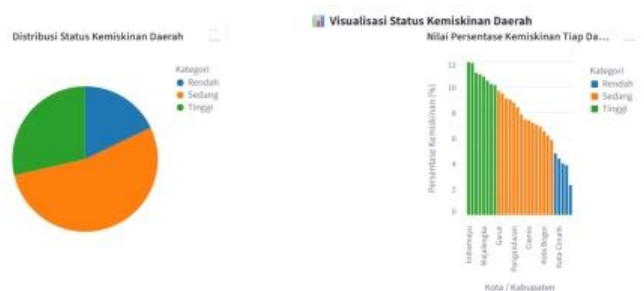
Pada **Gambar 6**. adalah Penjelasan mengenai evaluasi model:

- Akurasi 0,86 menunjukkan bahwa model berhasil mengklasifikasikan 86% wilayah dengan tepat sesuai kategorinya.
- Presisi tinggi (0,93) menandakan bahwa sebagian besar prediksi yang dilakukan model adalah benar (*false positive* rendah).
- Recall 0,86 menunjukkan bahwa model juga cukup andal dalam menangkap semua kasus nyata, walaupun masih terdapat beberapa *false negative*—khususnya pada kategori “Tinggi”.

Hasil ini menunjukkan bahwa model hanya menggunakan satu fitur utama, yaitu persentase kemiskinan, dan bahwa model dapat mengklasifikasikan status kemiskinan dengan cukup akurat. Nilai precision yang tinggi menunjukkan bahwa model jarang memberikan prediksi yang salah, dan nilai recall yang tinggi menunjukkan bahwa model mampu mengumpulkan sebagian besar data yang benar dari setiap kelas.

C. Visualisasi Prediksi

Adapun gambar Visualisasi Prediksi sebagai berikut pada gambar 7.



Visualisasi ini memberikan gambaran pola tingkat kemiskinan daerah sehingga dapat digunakan sebagai bahan pertimbangan untuk pengant...

Gbr. 7 Visualisasi Prediksi

Pada **Gambar 7**. Adalah aplikasi menyajikan hasil klasifikasi data melalui visualisasi interaktif, yaitu diagram batang (bar chart) dan diagram lingkaran (pie chart). Tujuan dari visualisasi ini adalah untuk membantu pengguna

memahami penyebaran status kemiskinan dan mengidentifikasi area yang memerlukan perhatian khusus.

➤ Pie Chart – Distribusi Status Kemiskinan

Diagram lingkaran menampilkan komposisi klasifikasi dari 31 wilayah di Provinsi Jawa Barat berdasarkan kategori status kemiskinan:

- Kategori Rendah ($\leq 5\%$) : 7 wilayah
- Kategori Sedang ($> 5\% - 10\%$) : 15 wilayah
- Kategori Tinggi ($> 10\%$) : 9 wilayah

Dari komposisi tersebut, terlihat bahwa sebagian besar wilayah masih berada dalam kategori Sedang, namun terdapat 9 wilayah yang telah masuk kategori Tinggi, yang memerlukan perhatian dan intervensi lebih serius dari pemerintah daerah.

➤ Bar Chart – Wilayah dengan Tingkat Kemiskinan Tertinggi

Diagram batang menyajikan urutan daerah berdasarkan persentase populasi miskin dari yang paling banyak hingga yang paling sedikit. Grafik ini membantu pengguna dalam mengidentifikasi area fokus untuk penanganan kemiskinan.

Contoh nyata dari hasil visualisasi ini:

- Kabupaten Indramayu menempati posisi teratas dengan angka kemiskinan 11,93%, menjadikannya wilayah dengan tingkat kemiskinan tertinggi di Jawa Barat berdasarkan data tahun 2024.
- Karena nilainya berada di atas 10%, maka kabupaten ini secara otomatis diklasifikasikan ke dalam kategori “Tinggi”.

➤ Manfaat Visualisasi

Kombinasi grafik lingkaran dan grafik batang menawarkan manfaat yang signifikan, antara lain:

- Mempermudah pengguna, termasuk yang tanpa latar belakang teknis, dalam memahami data dengan cepat.
- Menentukan area prioritas secara visual, sehingga memudahkan dalam merencanakan strategi intervensi sosial dan ekonomi.
- Menyandingkan keadaan antar wilayah, baik dalam satu kategori maupun di antara kategori yang berbeda.

Dengan adanya visualisasi ini, aplikasi tidak hanya berfungsi sebagai alat klasifikasi, tetapi juga sebagai alat bantu analisis visual yang dapat diterapkan langsung dalam forum kebijakan, evaluasi pembangunan wilayah, dan pelaporan resmi.

D. Koefisien Model

Model Naive Bayes yang diterapkan dalam aplikasi ini beroperasi dengan prinsip probabilitas, dengan mengasumsikan bahwa data dalam setiap kategori (kelas) mengikuti distribusi Gaussian (normal). Model ini menentukan kemungkinan suatu input (persentase kemiskinan) tergolong ke dalam salah satu kategori (Rendah, Sedang, atau Tinggi) menggunakan parameter distribusi statistik.

Tiga parameter utama yang digunakan dalam model Gaussian Naive Bayes adalah:

- **Mean (μ):** Rata-rata nilai persentase kemiskinan dalam masing-masing kategori.
- **Variance (σ^2):** Ukuran sebaran atau keragaman data pada setiap kategori.
- **Prior Probability ($P(C)$):** Peluang awal suatu data termasuk ke dalam kategori tertentu, dihitung dari proporsi data pada masing-masing kelas.

Tabel. 2 Koefisien Model

| Kategori | Mean (μ) | Varian (σ^2) |
|----------|----------------|-----------------------|
| Rendah | 3,7167 | 1,0516 |
| Sedang | 7,8136 | 1,4863 |
| Tinggi | 10,5725 | 0,1610 |

Interpretasi Data:

➤ Mean (μ):

- Setiap rata-rata kelas terletak di tengah-tengah rentang nilai kategorinya, menunjukkan bahwa model terlatih secara seimbang.
- Contoh: Rata-rata kelas “Tinggi” sebesar 10,5725 sangat dekat dengan batas bawah kategori ($>10\%$).

➤ Varian (σ^2):

- Kelas Tinggi memiliki varian terkecil (0,1610) → artinya, nilai-nilai dalam kategori ini sangat terkonsentrasi, sehingga lebih mudah dipisahkan dari kelas lain.
- Kelas Sedang memiliki varian terbesar (1,4863) → mencerminkan bahwa data di kelas ini lebih beragam, karena mencakup wilayah dengan kemiskinan dari $>5\%$ hingga 10% .
- Kelas Rendah memiliki varian menengah (1,0516) yang masih stabil.

Rumus Perhitungan Probabilitas:

Naive Bayes menghitung probabilitas suatu nilai masuk ke dalam sebuah kategori menggunakan rumus distribusi normal berikut:

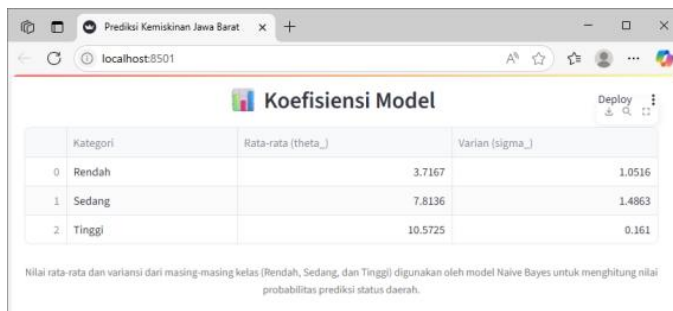
$$P(x|C) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Keterangan:

- **$P(x|C)$:** Peluang nilai x (misalnya, persentase kemiskinan suatu wilayah) berada pada kelas C (Rendah/Sedang/Tinggi)
- **μ :** Rata-rata kelas
- **σ^2 :** Variansi kelas
- **e :** Bilangan eksponensial ($\approx 2,718$)

Model akan menghitung probabilitas untuk setiap kelas, dan memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi.

Adapun gambar Koefisiense model pada Gambar 8.



| Kategori | Rata-rata (θ_{j_i}) | Varian ($\sigma_{j_i}^2$) |
|----------|------------------------------|-----------------------------|
| 0 Rendah | 3.7167 | 1.0516 |
| 1 Sedang | 7.8136 | 1.4863 |
| 2 Tinggi | 10.5725 | 0.161 |

Nilai rata-rata dan variansi dari masing-masing kelas (Rendah, Sedang, dan Tinggi) digunakan oleh model Naive Bayes untuk menghitung nilai probabilitas prediksi status daerah.

Gbr. 8 Koefisien Model

Pada **Gambar 8**, terdapat tabel koefisiensi model yang menunjukkan nilai rata-rata dan variansi dari setiap kategori. Tabel ini memperlihatkan cara model mengidentifikasi pola distribusi data untuk mengklasifikasikan input baru dengan tepat. Nilai-nilai ini diambil langsung dari dataset asli yang mencakup 31 daerah di Jawa Barat. Dengan memahami koefisien dari model ini, kita dapat menyimpulkan bahwa model Naive Bayes tidak hanya bergantung pada nilai input, tetapi juga mempertimbangkan pola distribusi serta sebaran data dalam setiap kelas untuk menghasilkan prediksi yang paling mungkin.

E. Pembahasan

Hasil penelitian ini menunjukkan bahwa model Gaussian Naive Bayes berhasil mengklasifikasikan tingkat kemiskinan setiap wilayah berdasarkan satu metrik, yaitu persentase kemiskinan. Hal ini menunjukkan bahwa metode yang digunakan dapat menghasilkan hasil yang signifikan dan bermanfaat untuk kebutuhan prediksi berbasis data, meskipun dianggap sederhana.

Keunggulan utama metode ini adalah efisiensi dan kecepatan proses klasifikasi, serta kemudahan penggunaan dalam aplikasi web menggunakan Streamlit. Selain itu, data dapat divisualisasikan dalam bentuk diagram lingkaran dan batang. Ini membantu pengguna, terutama pemangku kebijakan, untuk lebih memahami pola distribusi kemiskinan. Ini sangat penting karena banyak pengguna aplikasi berasal dari latar belakang non-teknis yang membutuhkan antarmuka yang informatif tetapi sederhana.

Penelitian ini memiliki beberapa keterbatasan, meskipun hasil evaluasi model cukup memuaskan. Sebagai contoh, jumlah fitur yang digunakan masih sangat terbatas. Meskipun pemodelan lebih mudah dilakukan dengan menggunakan satu fitur (persentase kemiskinan), itu mungkin menghilangkan konteks sosial-ekonomi yang lebih luas. Oleh karena itu, agar model memiliki konteks yang lebih kaya dan akurat, pengembangan selanjutnya harus memasukkan fitur tambahan seperti indeks pembangunan manusia (IPM), tingkat pendidikan, angka pengangguran, dan jumlah penduduk miskin absolut.

Selain itu, perlu mempertimbangkan distribusi data dalam kategori kemiskinan Rendah, Sedang, dan Tinggi. Jika ada ketidakseimbangan dalam jumlah data di masing-masing kelas, proses klasifikasi akan terpengaruh. Akibatnya, metode

penyeimbangan data seperti SMOTE atau undersampling mungkin dapat membantu penelitian berikutnya.

Secara keseluruhan, aplikasi yang telah dikembangkan sudah memenuhi tujuan utamanya: membuat perkiraan dan menunjukkan tingkat kemiskinan. Selain itu, implementasi yang dilakukan dalam penelitian ini menunjukkan bahwa penggabungan algoritma pengolahan data dengan antarmuka pengguna berbasis web dapat benar-benar membantu pengambilan keputusan strategis di bidang sosial ekonomi. Aplikasi ini dapat berkembang menjadi sistem keputusan yang lengkap untuk lembaga sosial dan pemerintah daerah.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil perancangan, implementasi, dan pengujian aplikasi untuk memprediksi tingkat kemiskinan di Provinsi Jawa Barat yang menggunakan algoritma Naive Bayes dan framework Streamlit dapat disimpulkan sebagai berikut:

1. Aplikasi yang dibuat sukses memproses dataset, melatih model klasifikasi, menunjukkan visualisasi interaktif, serta menghasilkan prediksi status kemiskinan wilayah secara realtime. Aplikasi ini juga menyediakan fitur untuk mengeksport hasil klasifikasi dalam format CSV, sehingga mendukung kebutuhan pelaporan instansi dengan baik.
2. Model Gaussian Naive Bayes yang diterapkan dapat menghasilkan klasifikasi dengan akurasi mencapai 86%, presisi rata-rata sebesar 93%, dan recall rata-rata sebesar 86%. Walaupun hanya mengandalkan satu fitur numerik (persentase kemiskinan), kinerja model tergolong baik dan konsisten, serta efisien dalam mengklasifikasikan tingkat kemiskinan.
3. Pemakaian Streamlit sebagai antarmuka berbasis web memungkinkan pengguna non-teknis seperti pegawai pemerintah atau petugas lapangan untuk dengan mudah menggunakan sistem hanya dengan memasukkan data dan mendapatkan keluaran klasifikasi dalam bentuk warna dan label kategori.
4. Penyajian dalam bentuk grafik lingkaran dan batang mempermudah pengguna untuk memahami distribusi status kemiskinan antarwilayah. Penggabungan visualisasi ini juga mempermudah penentuan area dengan tingkat kemiskinan tertinggi seperti Kabupaten Indramayu (11,93%) yang secara otomatis dikelompokkan dalam kategori "Tinggi".
5. Aplikasi ini bisa berfungsi sebagai alat pendukung awal bagi pemerintah daerah atau lembaga terkait dalam membuat kebijakan berbasis data, mempercepat langkah identifikasi daerah prioritas, dan merancang strategi penanggulangan kemiskinan yang lebih akurat.

B. Saran

1. Penambahan Variabel Fitur

Direkomendasikan untuk menyertakan fitur tambahan yang berkaitan dengan keadaan sosial ekonomi wilayah, seperti angka pengangguran, indeks pembangunan manusia (IPM), pendapatan rata-rata rumah tangga, atau jumlah anggota keluarga yang ditanggung. Hal ini akan mendukung model untuk menghasilkan klasifikasi yang lebih tepat.

2. Penyempurnaan Model

Untuk meningkatkan kinerja model, dapat dilakukan optimasi lebih lanjut seperti:

- Pemanfaatan penytelan hyperparameter
- Validasi silang (cross-validation)
- Membandingkan dengan algoritma klasifikasi lainnya seperti Decision Tree, Random Forest, atau XGBoost, untuk mengetahui model yang paling efektif berdasarkan data yang ada.

3. Pengembangan Fitur Aplikasi

Fitur tambahan yang dapat dikembangkan antara lain:

- Penyimpanan catatan prediksi ke dalam database.
- Integrasi akunnya pengguna untuk keamanan akses
- Penyusunan laporan otomatis (PDF atau grafik) untuk mendukung pelaporan resmi dinas.

4. Perluasan Dataset

Model dapat dikembangkan lebih lanjut dengan:

- Model bisa ditingkatkan lebih jauh dengan:
- Menambah data dari berbagai tahun (sejarah) supaya model dapat memahami tren.
- Mengintegrasikan data dari provinsi lain guna memperbaiki generalisasi model.

Dengan memperhatikan beberapa rekomendasi di atas, diharapkan aplikasi ini dapat terus ditingkatkan menjadi sistem dukungan keputusan yang lebih tepat, berguna, dan responsif dalam menghadapi masalah kemiskinan secara nyata.

Pangan Non Tunai Menggunakan Algoritma Naïve Bayes,” *JURIKOM J. Ris. Komput.*, vol. 9, no. 4, p. 1035, Aug. 2022, doi: 10.30865/jurikom.v9i4.4678.

- [5] D. L. Dede, Mochamad Arief Mardiansah, and Esthi Adityarini, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes Di Desa Jati Mulya,” *MULTINETICS*, vol. 10, no. 1, pp. 93–101, Jul. 2024, doi: 10.32722/multinetics.v10i1.6682.
- [6] “Klasifikasi penduduk miskin penerima PKH menggunakan metode naïve bayes dan KNN.”
- [7] P. Aji *et al.*, “Poverty Level Prediction Based on E-Commerce Data Using Naïve Bayes Algorithm and Similarity-Based Feature Selection,” *IJAIT Int. J. Appl. Inf. Technol.*, vol. 7, no. 02, p. 114, Oct. 2023, doi: 10.25124/ijait.v7i02.5374.
- [8] C. Fuadi Ahmad, N. Suarna, and G. Dwilestari, “Klasifikasi Data Kemiskinan Menggunakan Metode Naïve Bayes Untuk Mengetahui Tingkat Kemiskinan Studi Kasus: Desa Karangasem Kecamatan Leuwimunding Majalengka,” *J. Inform. Dan Teknol. Inf.*, vol. 2, no. 2, pp. 203–208, Nov. 2023, doi: 10.56854/jt.v2i2.190.
- [9] Z. A. Prayetno, M. J. Darwin, and L. Athalya, “Klasifikasi Tingkat Kemiskinan di Indonesia Menggunakan Metode Naïve Bayes,” vol. 6, no. 2, 2025.
- [10] H. Harliana and F. N. Putra, “Klasifikasi Tingkat Rumah Tangga Miskin Saat Pandemi Dengan Naïve Bayes Classifier,” *J. Sains Dan Inform.*, vol. 7, no. 2, pp. 165–173, Dec. 2021, doi: 10.34128/jsi.v7i2.339.

REFERENSI

- [1] E. S. Palupi, “CLASSIFICATION OF THE POOR IN SUMATERA AND JAVA ISLAND USING NAIVE BAYES ALGORITHM AND NAIVE BAYES ALGORITHM BASED ON PSO,” vol. 4, no. 3, 2022.
- [2] W. P. Nurmayanti, “Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak,” *Geodika J. Kaji. Ilmu Dan Pendidik. Geogr.*, vol. 5, no. 1, pp. 123–132, Jun. 2021, doi: 10.29408/geodika.v5i1.3430.
- [3] R. Riliandhita, I. Maulana, and P. Purwanto, “KLASIFIKASI PENENTUAN STATUS KEMISKINAN PENDUDUK KELURAHAN KARANGPAWITAN KARAWANG MENGGUNAKAN METODE C4.5,” *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1791–1796, Apr. 2024, doi: 10.36040/jati.v8i2.9219.
- [4] N. Nurahman, M. M. Alfitri, and E. Mashamy, “Klasifikasi Data Penduduk Untuk Menerima Bantuan