



## Integrasi Naive Bayes dan ArcGIS untuk Penentuan Prioritas Mustahik Zakat di LAZISMU Sumatera Utara

**Mhd. Arkan Mufadhal<sup>1\*</sup>, Firahmi Rizky<sup>2</sup>**

<sup>1\*,2</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Email: <sup>1\*</sup>mafproject87@gmail.com, <sup>2</sup>firahmirizky@umsu.ac.id

### Abstrak

Pengelolaan distribusi zakat yang objektif dan tepat sasaran merupakan tantangan bagi lembaga amal zakat, termasuk LAZISMU Sumatera Utara, yang selama ini menentukan prioritas penerima zakat (mustahik) tanpa mekanisme klasifikasi berbasis data dan belum memanfaatkan analisis lokasi geografis. Penelitian ini bertujuan membangun sistem penentuan prioritas mustahik menggunakan algoritma Naive Bayes yang diintegrasikan dengan Sistem Informasi Geografis (ArcGIS). Data penelitian terdiri atas 199 data latih (gabungan data mustahik tahun 2024 dan 2025) serta 100 data uji independen (data tahun 2026), dengan lima atribut kategorikal: Status Asnaf, Kategori Pendapatan, Kategori Tanggungan, Kondisi Rumah, dan Status Rumah. Atribut numerik didiskretisasi dan Laplace Smoothing diterapkan untuk mengatasi zero-probability. Pengujian model terhadap 100 data uji menghasilkan akurasi 92,00%, precision 92,47%, recall 98,85%, dan F1-Score 95,56%, dengan atribut Status Asnaf dan Kondisi Rumah teridentifikasi sebagai atribut dengan daya pembeda terbesar. Hasil klasifikasi selanjutnya divisualisasikan dengan ArcGIS, mengidentifikasi Kota Medan, Kabupaten Deli Serdang, dan Kota Binjai sebagai klaster dengan konsentrasi mustahik prioritas tinggi terbesar, sementara Kabupaten Simalungun dan Kota Tebing Tinggi memiliki persentase prioritas tertinggi secara proporsional. Penelitian ini menunjukkan bahwa kombinasi Naive Bayes dan GIS mampu meningkatkan objektivitas dan efisiensi distribusi zakat berbasis kelayakan individu maupun sebaran wilayah.

**Kata Kunci:** Naive Bayes, ArcGIS, Mustahik Zakat, Klasifikasi, Sistem Informasi Geografis.

### PENDAHULUAN

Zakat merupakan salah satu rukun Islam yang memiliki dimensi sosial-ekonomi signifikan dalam upaya pengentasan kemiskinan (Chaerunnisa et al., 2025). Penyaluran zakat yang tepat sasaran kepada mustahik menjadi kunci efektivitas pengelolaan dana zakat, sebagaimana diatur dalam Undang-Undang No. 23 Tahun 2011 tentang Pengelolaan Zakat (Rasyad Al Fajar & Jannah, 2021). Namun, lembaga amal zakat seringkali menghadapi keterbatasan dalam mengidentifikasi dan memprioritaskan mustahik secara sistematis dan berbasis data (Grahesti et al., 2023).

LAZISMU (Lembaga Amil Zakat, Infaq, dan Shadaqah Muhammadiyah) Sumatera Utara merupakan salah satu lembaga amal zakat yang aktif beroperasi di Provinsi Sumatera Utara. Proses penentuan prioritas penerima zakat pada lembaga ini masih mengandalkan penilaian berdasarkan kategori dan data wilayah tanpa mekanisme klasifikasi yang objektif, serta belum memanfaatkan analisis lokasi geografis secara terintegrasi, sehingga berpotensi menimbulkan ketidaktepatan sasaran dan ketimpangan distribusi zakat antarwilayah (Nazila, 2024).

Algoritma Naive Bayes merupakan metode klasifikasi probabilistik berbasis Teorema Bayes yang dikenal efektif untuk data kategorikal berukuran kecil hingga menengah, dengan proses perhitungan yang cepat dan tingkat ketepatan yang cukup baik meskipun mengasumsikan independensi antar atribut (Firmansyah et al., 2023; Fatra et al., 2024). Metode ini telah diaplikasikan secara luas dalam berbagai domain klasifikasi berbasis data sosial-ekonomi. Untuk melengkapi analisis statistik, integrasi dengan Sistem Informasi Geografis (SIG) memberikan dimensi spasial yang kritis dalam pengambilan keputusan distribusi berbasis wilayah (Nurul & Rizky, 2026).

Penggunaan ArcGIS dalam konteks pemetaan data sosial telah terbukti mampu menghasilkan visualisasi persebaran yang informatif bagi pemangku kebijakan, misalnya dalam pemetaan sebaran stunting dan tingkat kemiskinan (Widodo & Herlambang, 2024). Peta kepadatan (kernel density) dan agregasi berbasis wilayah administratif memungkinkan lembaga amal zakat untuk memprioritaskan

penyaluran ke wilayah dengan konsentrasi mustahik tertinggi. Pendekatan serupa juga relevan dengan sistem pendukung keputusan berbasis GIS yang telah diterapkan pada sektor pelayanan publik lain (Nurul & Rizky, 2026).

Penelitian ini bertujuan untuk: (1) menganalisis proses penyaluran zakat di LAZISMU Sumatera Utara yang belum memiliki sistem penentuan prioritas yang objektif dan berbasis data; (2) membangun model klasifikasi Naive Bayes untuk menentukan prioritas mustahik zakat berdasarkan atribut kategori penerima; dan (3) mengintegrasikan hasil klasifikasi dengan ArcGIS untuk memetakan dan menganalisis sebaran penerima zakat berdasarkan lokasi geografisnya. Hasil penelitian ini diharapkan dapat membantu LAZISMU Sumatera Utara memperbaiki akurasi distribusi zakat agar lebih obyektif, transparan, dan tepat sasaran.

## METODE

### Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder berupa rekam data mustahik LAZISMU Sumatera Utara periode 2024–2026. Data tahun 2024 (100 record) dan 2025 (99 record setelah pembersihan dari total 102 record awal) digabungkan sebagai data latih dengan total 199 record, sedangkan data tahun 2026 (100 record) digunakan sepenuhnya sebagai data uji independen. Pembagian ini dipilih agar pengujian model mencerminkan skenario nyata, yaitu model dilatih menggunakan data historis untuk memprediksi prioritas penerima zakat pada periode berikutnya. Atribut yang digunakan sebagai fitur prediktor meliputi: (1) Status Asnaf (Fakir, Miskin, Amil, Muallaf, Fisabilillah, Ibnu Sabil, Gharim); (2) Kategori Pendapatan (hasil diskretisasi Pendapatan Bulanan); (3) Kategori Tanggungan (hasil diskretisasi Jumlah Tanggungan); (4) Kondisi Rumah (Layak, Sederhana, Tidak Layak); dan (5) Status Rumah (Milik, Sewa, Menumpang). Label kelas target (Label\_Prediksi) terdiri dari dua kategori: 1 = Prioritas Tinggi dan 0 = Bukan Prioritas. Setiap record juga memiliki atribut Latitude dan Longitude yang menunjukkan lokasi domisili mustahik di Provinsi Sumatera Utara (rentang koordinat 2,88°–3,80° LU dan 98,33°–99,70° BT), digunakan pada tahap analisis spasial.

Tabel 1. Rekapitulasi Jumlah Data Penelitian

| Tahun Data | Jumlah Data Awal | Jumlah Data Setelah Pembersihan | Peran dalam Penelitian    |
|------------|------------------|---------------------------------|---------------------------|
| 2024       | 100              | 100                             | Data Latih (Training Set) |
| 2025       | 102              | 99                              | Data Latih (Training Set) |
| 2026       | 100              | 100                             | Data Uji (Testing Set)    |

Ditemukan satu baris data pada berkas tahun 2025 yang mengalami pergeseran kolom akibat kesalahan entri, serta satu baris dengan nilai label kosong (missing value); kedua baris tersebut dikeluarkan dari data penelitian, sehingga data tahun 2025 berkurang dari 102 menjadi 99 record.

### Pra-pemrosesan dan Discretisasi Data

Atribut Pendapatan Bulanan dan Jumlah Tanggungan yang bersifat numerik didiskretisasi menjadi kategori agar dapat diolah oleh Naive Bayes kategorikal. Pendapatan Bulanan dibagi menjadi empat kategori, sedangkan Jumlah Tanggungan dibagi menjadi tiga kategori, sebagaimana disajikan pada Tabel 2 di bawah ini.

Tabel 2. Discretisasi Atribut Numerik

| Atribut            | Kategori      | Rentang Nilai             |
|--------------------|---------------|---------------------------|
| Pendapatan Bulanan | Sangat Rendah | < Rp1.000.000             |
|                    | Rendah        | Rp1.000.000 – Rp1.999.999 |
|                    | Sedang        | Rp2.000.000 – Rp2.999.999 |
|                    | Tinggi        | ≥ Rp3.000.000             |
| Jumlah Tanggungan  | Sedikit       | ≤ 2 orang                 |
|                    | Sedang        | 3 – 4 orang               |
|                    | Banyak        | ≥ 5 orang                 |

Distribusi label pada masing-masing subset data (sebelum dan setelah pembagian latih-uji) disajikan pada Tabel 3, yang menunjukkan ketidakseimbangan kelas (class imbalance) dengan proporsi mustahik berlabel ‘Prioritas Tinggi’ yang jauh lebih besar dibandingkan ‘Bukan Prioritas’ pada kedua subset.

Tabel 3. Distribusi Label pada Data Latih dan Data Uji

| Subset Data | Jumlah Data | Label 1 (Prioritas Tinggi) | Label 0 (Bukan Prioritas) |
|-------------|-------------|----------------------------|---------------------------|
| Data Latih  | 199         | 177 (88,9%)                | 22 (11,1%)                |
| Data Uji    | 100         | 87 (87,0%)                 | 13 (13,0%)                |

### Algoritma Naive Bayes dan Laplace Smoothing

Naive Bayes adalah metode klasifikasi probabilistik berbasis Teorema Bayes dengan asumsi independensi antar atribut (Firmansyach et al., 2023). Klasifikasi kelas  $C$  untuk data dengan atribut  $X = (X_1, X_2, \dots, X_n)$  dihitung menggunakan persamaan:

$$P(C|X) = [P(C) \times P(X_1|C) \times P(X_2|C) \times \dots \times P(X_n|C)]/P(X)$$

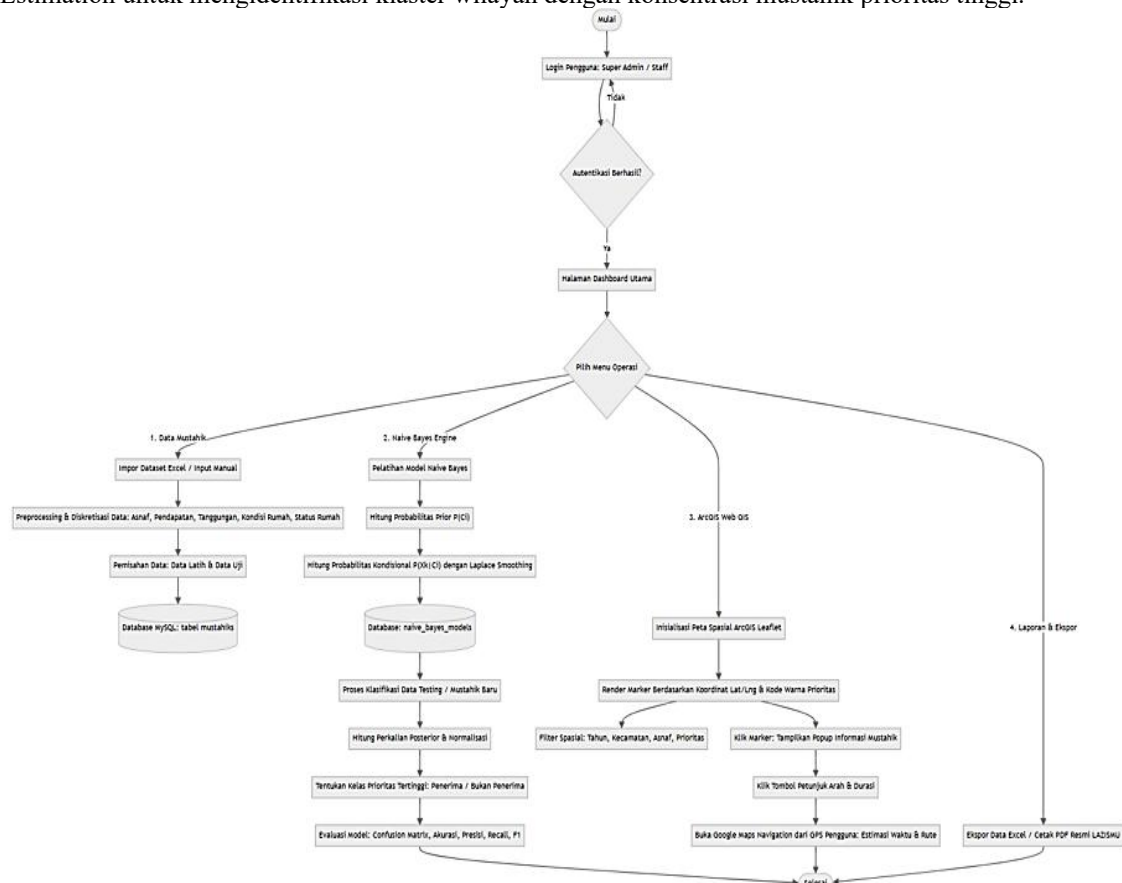
Karena nilai  $P(X)$  bersifat konstan untuk setiap kelas yang dibandingkan pada data yang sama, penentuan kelas dilakukan dengan mencari nilai maksimum dari pembilang  $P(C) \times \prod P(X_i|C)$ . Probabilitas prior  $P(C)$  dihitung berdasarkan proporsi kemunculan masing-masing kelas pada data latih:  $P(C) = \text{jumlah data kelas } C \div \text{jumlah seluruh data latih}$ . Untuk mengatasi masalah zero-probability pada kategori yang tidak muncul dalam data latih, diterapkan Laplace Smoothing (Fatra et al., 2024):

$$P(X_i = x_i|C) = (n(x_i, C) + 1)/(n(C) + k)$$

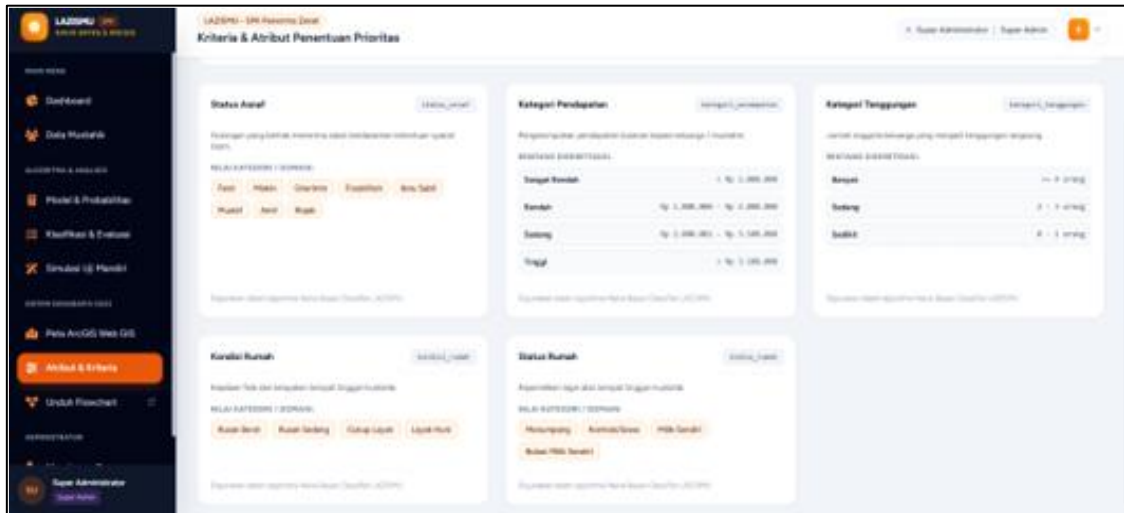
di mana  $n(x_i, C)$  adalah jumlah kemunculan nilai  $x_i$  pada kelas  $C$ ,  $n(C)$  adalah total data kelas  $C$ , dan  $k$  adalah banyaknya kategori berbeda pada atribut  $X_i$ .

### Evaluasi Model dan Integrasi ArcGIS

Model yang dibangun dari 199 data latih diuji terhadap 100 data uji independen (data tahun 2026). Evaluasi kinerja menggunakan confusion matrix yang menghasilkan empat metrik: Accuracy =  $(TP+TN)/(TP+TN+FP+FN)$ ; Precision =  $TP/(TP+FP)$ ; Recall =  $TP/(TP+FN)$ ; dan F1-Score =  $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$ . Hasil klasifikasi kemudian diintegrasikan ke dalam ArcGIS (ArcGIS Pro/ArcMap) melalui tahapan: (1) impor data Latitude, Longitude, dan Label\_Prediksi menggunakan fitur XY Table to Point dengan sistem referensi koordinat WGS 1984 (EPSG:4326); (2) simbolisasi titik berdasarkan Label\_Prediksi menggunakan metode Unique Values; (3) agregasi spasial jumlah dan proporsi mustahik prioritas tinggi per kabupaten/kota menggunakan fitur Spatial Join/Summarize Within; dan (4) pembuatan peta kepadatan (heat map) menggunakan Kernel Density Estimation untuk mengidentifikasi kluster wilayah dengan konsentrasi mustahik prioritas tinggi.



Gambar 1. Flowchart Sistem Penentuan Prioritas Mustahik



Gambar 2. Atribut yang Digunakan dalam Model Naive Bayes.

## HASIL DAN PEMBAHASAN

### Probabilitas Prior

Berdasarkan 199 data latih, diperoleh  $P(C=0) = 22/199 = 0,1106$  untuk kelas ‘Bukan Prioritas’ dan  $P(C=1) = 177/199 = 0,8894$  untuk kelas ‘Prioritas Tinggi’. Nilai ini menunjukkan sebaran kelas yang tidak seimbang (imbalanced), sehingga pada tahap evaluasi turut diperhatikan metrik precision dan recall di samping akurasi.

Tabel 4. Perhitungan Probabilitas Prior  $P(C)$

| Kelas (C) | Keterangan       | Jumlah Data | $P(C)$             |
|-----------|------------------|-------------|--------------------|
| 0         | Bukan Prioritas  | 22          | $22/199 = 0,1106$  |
| 1         | Prioritas Tinggi | 177         | $177/199 = 0,8894$ |

### Probabilitas Likelihood per Atribut

Probabilitas likelihood dihitung untuk setiap nilai kategori pada kelima atribut terhadap masing-masing kelas menggunakan Laplace Smoothing. Hasil perhitungan disajikan pada Tabel 5 hingga Tabel 9.

Tabel 5. Likelihood Atribut Status Asnaf ( $k=7$ )

| Kategori     | $n(C=0)$ | $P(X_i C=0)$            | $n(C=1)$ | $P(X_i C=1)$              |
|--------------|----------|-------------------------|----------|---------------------------|
| Amil         | 4        | $(4+1)/(22+7) = 0,1724$ | 4        | $(4+1)/(177+7) = 0,0272$  |
| Fakir        | 1        | $(1+1)/(22+7) = 0,0690$ | 61       | $(61+1)/(177+7) = 0,3370$ |
| Fisabilillah | 4        | $(4+1)/(22+7) = 0,1724$ | 5        | $(5+1)/(177+7) = 0,0326$  |
| Gharim       | 0        | $(0+1)/(22+7) = 0,0345$ | 2        | $(2+1)/(177+7) = 0,0163$  |
| Ibnu Sabil   | 3        | $(3+1)/(22+7) = 0,1379$ | 7        | $(7+1)/(177+7) = 0,0435$  |
| Miskin       | 4        | $(4+1)/(22+7) = 0,1724$ | 96       | $(96+1)/(177+7) = 0,5272$ |
| Muallaf      | 6        | $(6+1)/(22+7) = 0,2414$ | 2        | $(2+1)/(177+7) = 0,0163$  |

Tabel 6. Likelihood Atribut Kategori Pendapatan ( $k=4$ )

| Kategori      | $n(C=0)$ | $P(X_i C=0)$            | $n(C=1)$ | $P(X_i C=1)$               |
|---------------|----------|-------------------------|----------|----------------------------|
| Rendah        | 6        | $(6+1)/(22+4) = 0,2692$ | 101      | $(101+1)/(177+4) = 0,5635$ |
| Sangat Rendah | 0        | $(0+1)/(22+4) = 0,0385$ | 21       | $(21+1)/(177+4) = 0,1215$  |
| Sedang        | 7        | $(7+1)/(22+4) = 0,3077$ | 38       | $(38+1)/(177+4) = 0,2155$  |
| Tinggi        | 9        | $(9+1)/(22+4) = 0,3846$ | 17       | $(17+1)/(177+4) = 0,0994$  |

Tabel 7. Likelihood Atribut Kategori Jumlah Tanggungan ( $k=3$ )

| Kategori | $n(C=0)$ | $P(X_i C=0)$             | $n(C=1)$ | $P(X_i C=1)$              |
|----------|----------|--------------------------|----------|---------------------------|
| Banyak   | 0        | $(0+1)/(22+3) = 0,0400$  | 22       | $(22+1)/(177+3) = 0,1278$ |
| Sedang   | 2        | $(2+1)/(22+3) = 0,1200$  | 57       | $(57+1)/(177+3) = 0,3222$ |
| Sedikit  | 20       | $(20+1)/(22+3) = 0,8400$ | 98       | $(98+1)/(177+3) = 0,5500$ |

Tabel 8. Likelihood Atribut Kondisi Rumah (k=3)

| Kategori    | n(C=0) | P(Xi C=0)                | n(C=1) | P(Xi C=1)                 |
|-------------|--------|--------------------------|--------|---------------------------|
| Layak       | 1      | $(1+1)/(22+3) = 0,0800$  | 42     | $(42+1)/(177+3) = 0,2389$ |
| Sederhana   | 19     | $(19+1)/(22+3) = 0,8000$ | 86     | $(86+1)/(177+3) = 0,4833$ |
| Tidak Layak | 2      | $(2+1)/(22+3) = 0,1200$  | 49     | $(49+1)/(177+3) = 0,2778$ |

Tabel 9. Likelihood Atribut Status Rumah (k=3)

| Kategori  | n(C=0) | P(Xi C=0)                | n(C=1) | P(Xi C=1)                 |
|-----------|--------|--------------------------|--------|---------------------------|
| Menumpang | 0      | $(0+1)/(22+3) = 0,0400$  | 24     | $(24+1)/(177+3) = 0,1389$ |
| Milik     | 19     | $(19+1)/(22+3) = 0,8000$ | 90     | $(90+1)/(177+3) = 0,5056$ |
| Sewa      | 3      | $(3+1)/(22+3) = 0,1600$  | 63     | $(63+1)/(177+3) = 0,3556$ |

Sebagai catatan, kategori Gharim pada atribut Status Asnaf tidak muncul pada data latih kelas 0 (Bukan Prioritas), sehingga tanpa Laplace Smoothing probabilitasnya akan bernilai nol dan berpotensi menghilangkan kontribusi atribut lain pada perkalian probabilitas. Dengan Laplace Smoothing, probabilitas tersebut tetap bernilai kecil namun tidak nol, yaitu sebesar 0,0345.

#### Contoh Perhitungan Klasifikasi Manual

Sebagai demonstrasi, diambil satu data riil dari data uji 2026, yaitu mustahik ID\_Mustahik = MSH0002 asal Kabupaten Deli Serdang, dengan atribut Status Asnaf = Fakir, Kategori Pendapatan = Rendah, Kategori Tanggungan = Sedang, Kondisi Rumah = Sederhana, dan Status Rumah = Milik.

$$\begin{aligned} SkorkelasC = 0(BukanPrioritas): & P(0) \times P(Fakir|0) \times P(Rendah|0) \times P(Sedang|0) \\ & \times P(Sederhana|0) \times P(Milik|0) \\ & = 0,1106 \times 0,0690 \times 0,2692 \times 0,1200 \times 0,8000 \times 0,8000 = 0,0001576 \end{aligned}$$

$$\begin{aligned} SkorkelasC = 1(PrioritasTinggi): & P(1) \times P(Fakir|1) \times P(Rendah|1) \times P(Sedang|1) \\ & \times P(Sederhana|1) \times P(Milik|1) \\ & = 0,8894 \times 0,3370 \times 0,5635 \times 0,3222 \times 0,4833 \times 0,5056 = 0,013298 \end{aligned}$$

Setelah normalisasi:  $P(C=0|X) = 1,17\%$  dan  $P(C=1|X) = 98,83\%$ . Karena skor kelas C=1 jauh lebih besar, mustahik MSH0002 diklasifikasikan ke dalam kelas Prioritas Tinggi (1), sesuai dengan label aktualnya, sehingga prediksi dinyatakan benar (true positive).

Tabel 10. Contoh Hasil Prediksi Model pada Data Uji (Sebagian)

| No | ID_Mustahik | Kabupaten/Kota  | Status Asnaf | Label Aktual | Label Prediksi | Ket.  |
|----|-------------|-----------------|--------------|--------------|----------------|-------|
| 1  | MSH0002     | Deli Serdang    | Fakir        | 1            | 1              | Benar |
| 2  | MSH0008     | Asahan          | Ibnu Sabil   | 0            | 1              | Salah |
| 3  | MSH0012     | Tebing Tinggi   | Fakir        | 1            | 1              | Benar |
| 4  | MSH0051     | Asahan          | Gharim       | 0            | 1              | Salah |
| 5  | MSH0054     | Serdang Bedagai | Fisabilillah | 0            | 0              | Benar |
| 6  | MSH0116     | Binjai          | Miskin       | 1            | 1              | Benar |
| 7  | MSH0140     | Binjai          | Muallaf      | 0            | 0              | Benar |
| 8  | MSH0183     | Binjai          | Miskin       | 0            | 1              | Salah |

#### Evaluasi Model: Confusion Matrix dan Metrik Kinerja

Pengujian terhadap keseluruhan 100 data uji menghasilkan confusion matrix pada Tabel 11. Model berhasil mengklasifikasikan 86 dari 87 sampel 'Prioritas Tinggi' dengan benar (hanya 1 false negative) dan 6 dari 13 sampel 'Bukan Prioritas' dengan benar (7 false positive).

Tabel 11. Confusion Matrix Hasil Pengujian Model

|                              | Prediksi: Bukan Prioritas (0) | Prediksi: Prioritas Tinggi (1) |
|------------------------------|-------------------------------|--------------------------------|
| Aktual: Bukan Prioritas (0)  | TN = 6                        | FP = 7                         |
| Aktual: Prioritas Tinggi (1) | FN = 1                        | TP = 86                        |

Berdasarkan nilai confusion matrix (TP=86, TN=6, FP=7, FN=1), diperoleh metrik evaluasi sebagai berikut:

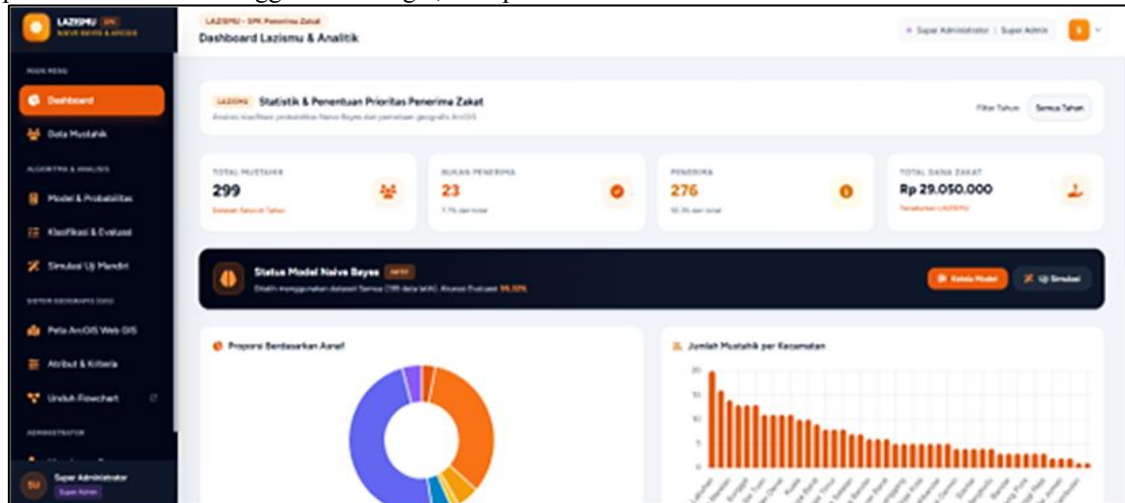
Tabel 12. Ringkasan Metrik Evaluasi Model Naive Bayes

| Metrik    | Perhitungan                                           | Nilai  |
|-----------|-------------------------------------------------------|--------|
| Accuracy  | $(86+6)/(86+6+7+1) = 92/100$                          | 92,00% |
| Precision | $86/(86+7) = 86/93$                                   | 92,47% |
| Recall    | $86/(86+1) = 86/87$                                   | 98,85% |
| F1-Score  | $2 \times (0,9247 \times 0,9885) / (0,9247 + 0,9885)$ | 95,56% |

Nilai recall sebesar 98,85% menunjukkan model sangat andal dalam mengenali mustahik yang benar-benar berhak menerima prioritas tinggi, dengan hanya 1 kesalahan false negative dari 87 data prioritas tinggi, sehingga risiko mustahik prioritas terlewat menjadi sangat kecil — hal yang krusial dalam konteks distribusi zakat. Precision sebesar 92,47% menunjukkan bahwa dari seluruh data yang diprediksi sebagai prioritas tinggi, terdapat 7 data yang sebenarnya bukan prioritas (false positive), yang sebaiknya tetap diverifikasi ulang secara manual oleh amil sebelum penyaluran dana dilakukan. Akurasi 92,00% dan F1-Score 95,56% mengonfirmasi bahwa model Naive Bayes efektif digunakan pada data kategorikal berukuran relatif kecil hingga menengah (Firmansyach et al., 2023; Fatra et al., 2024).

Sebanyak 8 dari 100 data uji mengalami kesalahan klasifikasi (7 false positive dan 1 false negative). Kasus false positive umumnya terjadi pada mustahik dengan kombinasi atribut yang secara nilai tunggal (misalnya Kondisi Rumah = Sederhana atau Status Rumah = Milik) cenderung mengarah pada kategori prioritas tinggi berdasarkan pola mayoritas data latih, meskipun kombinasi atribut lain (misalnya Status Asnaf = Ibnu Sabil atau Gharim yang jarang muncul pada data latih) semestinya menurunkan skor probabilitasnya. Kondisi ini merupakan konsekuensi dari asumsi independensi atribut (naive) yang menjadi dasar algoritma, sehingga interaksi antar atribut tidak sepenuhnya dapat ditangkap oleh model.

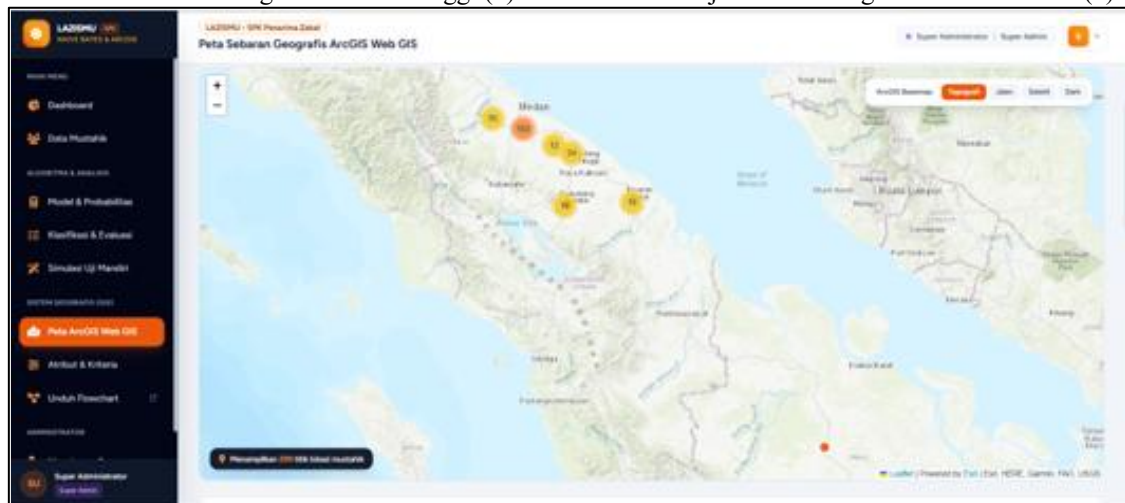
Atribut Status Asnaf dan Kondisi Rumah teridentifikasi sebagai atribut dengan daya pembeda (discriminative power) paling besar terhadap hasil klasifikasi, terlihat dari selisih probabilitas likelihood antar kelas yang cukup signifikan, misalnya kategori Miskin pada Status Asnaf memiliki likelihood 0,5272 pada kelas Prioritas Tinggi berbanding 0,1724 pada kelas Bukan Prioritas.



Gambar 3. Antarmuka Sistem Klasifikasi Mustahik

### Visualisasi Spasial dengan ArcGIS

Setiap record memiliki atribut Latitude dan Longitude yang menunjukkan lokasi domisili mustahik di Provinsi Sumatera Utara (rentang koordinat 2,88°–3,80° LU dan 98,33°–99,70° BT). Data ini diimpor ke ArcGIS menggunakan fitur XY Table to Point dengan sistem referensi koordinat WGS 1984 (EPSG:4326), kemudian disimbolisasi berdasarkan Label\_Prediksi menggunakan metode Unique Values warna merah untuk kategori Prioritas Tinggi (1) dan warna biru/hijau untuk kategori Bukan Prioritas (0).



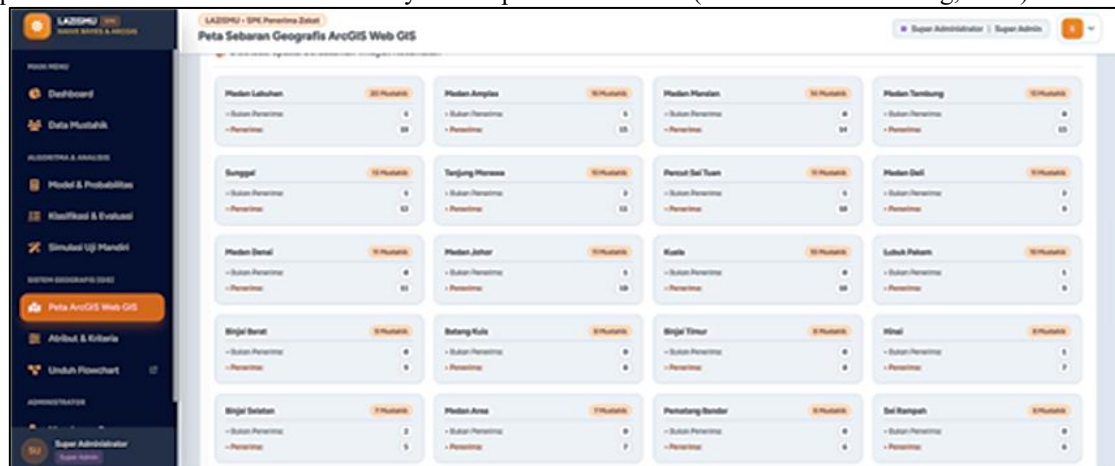
Gambar 4. Peta Persebaran Mustahik di Provinsi Sumatera Utara Berdasarkan Hasil Klasifikasi

Agregasi spasial menggunakan fitur Spatial Join/Summarize Within menghitung jumlah dan proporsi mustahik prioritas tinggi pada tiap wilayah administrasi (kabupaten/kota), sedangkan Kernel Density Estimation menghasilkan peta kepadatan (heat map) untuk mengidentifikasi kluster wilayah dengan konsentrasi mustahik prioritas tinggi. Rekapitulasi hasil agregasi (gabungan seluruh data 2024–2026) disajikan pada Tabel 13.

Tabel 13. Distribusi Geografis Mustahik Prioritas Tinggi per Kabupaten/Kota

| Kabupaten/Kota  | Jumlah Data | Prioritas Tinggi (Label=1) | Persentase Prioritas |
|-----------------|-------------|----------------------------|----------------------|
| Medan           | 103         | 96                         | 93,2%                |
| Deli Serdang    | 55          | 49                         | 89,1%                |
| Binjai          | 36          | 28                         | 77,8%                |
| Langkat         | 30          | 26                         | 86,7%                |
| Serdang Bedagai | 22          | 18                         | 81,8%                |
| Simalungun      | 19          | 18                         | 94,7%                |
| Asahan          | 19          | 15                         | 78,9%                |
| Tebing Tinggi   | 15          | 14                         | 93,3%                |

Berdasarkan Tabel 13, Kota Medan memiliki jumlah mustahik prioritas tinggi terbanyak secara absolut (96 dari 103 data), sejalan dengan statusnya sebagai wilayah dengan populasi dan jumlah penerima zakat terbesar. Namun, apabila ditinjau dari sisi persentase, Kabupaten Simalungun (94,7%) dan Kota Tebing Tinggi (93,3%) menunjukkan proporsi mustahik prioritas tinggi yang lebih tinggi dibandingkan Kota Medan (93,2%), yang mengindikasikan tingkat kerentanan ekonomi yang secara relatif lebih besar di wilayah tersebut meskipun jumlah data absolutnya lebih sedikit. Visualisasi ArcGIS pada peta kepadatan menampilkan kluster kepadatan tertinggi (hotspot) di sekitar wilayah Medan, Deli Serdang, dan Binjai yang secara geografis saling berdekatan (Kawasan Mebidangro), sehingga dapat dijadikan dasar penentuan prioritas distribusi zakat berbasis wilayah oleh pihak LAZISMU (Widodo & Herlambang, 2024).



Gambar 5. Peta Kepadatan (Kernel Density) dan Distribusi Geografis Mustahik Prioritas Tinggi per Kabupaten/Kota

## Pembahasan

Hasil penelitian menunjukkan bahwa kombinasi algoritma Naive Bayes dan ArcGIS mampu menjawab dua kebutuhan sekaligus dalam pengambilan keputusan penyaluran zakat, yaitu aspek klasifikasi kelayakan individu dan aspek spasial lokasi mustahik. Algoritma Naive Bayes terbukti efektif digunakan pada data kategorikal berukuran relatif kecil hingga menengah (199 data latih) dengan menghasilkan kinerja yang tinggi pada 100 data uji independen dari tahun berbeda, konsisten dengan temuan Firmansyach et al. (2023) dan Fatra et al. (2024) mengenai efektivitas Naive Bayes pada klasifikasi kategorikal.

Integrasi dengan ArcGIS memberikan nilai tambah berupa kemampuan analisis spasial yang tidak dimiliki oleh algoritma klasifikasi semata, yaitu identifikasi kluster wilayah dengan konsentrasi mustahik prioritas tinggi. Informasi ini dapat dimanfaatkan oleh LAZISMU Sumatera Utara untuk merancang strategi distribusi zakat yang lebih efisien, misalnya penentuan titik pengumpulan dan penyaluran yang lebih dekat dengan kluster wilayah prioritas tinggi, sehingga biaya operasional distribusi dapat ditekan tanpa mengurangi ketepatan sasaran. Pendekatan integrasi machine learning dengan GIS semacam ini juga relevan dengan sistem pendukung keputusan berbasis spasial pada sektor pelayanan publik lainnya (Nurul & Rizky, 2026).

Secara keseluruhan, temuan ini mendukung hipotesis bahwa penerapan Naive Bayes yang dipadukan dengan ArcGIS dapat meningkatkan efektivitas dan objektivitas proses penentuan prioritas penerima zakat, baik dari sisi kelayakan individu (atribut sosial-ekonomi) maupun dari sisi sebaran wilayah (atribut geografis), dibandingkan dengan metode penentuan prioritas yang dilakukan secara manual (Grahesti et al., 2023).

## KESIMPULAN

Algoritma Naive Bayes dapat diterapkan untuk mengklasifikasikan prioritas penerima zakat pada LAZISNU Sumatera Utara menggunakan lima atribut kategorikal, yaitu Status Asnaf, Kategori Pendapatan, Kategori Tanggungan, Kondisi Rumah, dan Status Rumah. Model dibangun menggunakan 199 data latih dengan probabilitas prior  $P(C=0) = 0,1106$  dan  $P(C=1) = 0,8894$ , serta probabilitas likelihood tiap atribut yang dihitung menggunakan Laplace Smoothing untuk mengatasi nilai probabilitas nol.

Hasil pengujian model terhadap 100 data uji independen (data tahun 2026) menunjukkan kinerja klasifikasi yang tinggi, dengan nilai akurasi 92,00%, precision 92,47%, recall 98,85%, dan F1-Score 95,56%. Atribut Status Asnaf dan Kondisi Rumah terbukti menjadi atribut dengan daya pembeda paling besar dalam menentukan prioritas penerima zakat.

Integrasi hasil klasifikasi dengan ArcGIS berhasil memvisualisasikan sebaran geografis mustahik prioritas tinggi di Provinsi Sumatera Utara. Kota Medan, Kabupaten Deli Serdang, dan Kota Binjai teridentifikasi sebagai kluster dengan konsentrasi mustahik prioritas tinggi terbesar, sedangkan Kabupaten Simalungun dan Kota Tebing Tinggi memiliki persentase prioritas tinggi tertinggi ditinjau dari proporsi terhadap jumlah data pada masing-masing wilayah. Kombinasi algoritma Naive Bayes dan ArcGIS terbukti mampu meningkatkan objektivitas dan efisiensi proses penentuan prioritas penerima zakat, sehingga dapat menjadi alat bantu pengambilan keputusan bagi pengurus LAZISNU Sumatera Utara.

Hasil klasifikasi model sebaiknya tidak dijadikan satu-satunya dasar pengambilan keputusan, melainkan digunakan sebagai alat bantu pendukung keputusan yang tetap disertai verifikasi lapangan oleh amil, khususnya untuk kategori Status Asnaf yang jarang muncul pada data historis (Amil, Muallaf, Ibnu Sabil, Gharim). Penelitian selanjutnya disarankan untuk menambah jumlah data latih dari periode yang lebih panjang, membandingkan kinerja Naive Bayes dengan algoritma lain seperti Decision Tree atau Random Forest, serta mengembangkan sistem terintegrasi berbasis web yang menggabungkan klasifikasi otomatis dengan visualisasi ArcGIS Online secara real-time.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada LAZISNU Sumatera Utara yang telah memberikan izin dan akses data mustahik yang diperlukan dalam penelitian ini. Penulis juga mengucapkan terima kasih kepada Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, atas dukungan fasilitas dan bimbingan akademik selama proses penelitian berlangsung.

## DAFTAR PUSTAKA

- Chaerunnisa, R., Azizah Muchlis, N., & Nur Eliza, W. (2025). Zakat dalam Islam: Pengertian, Hukum, Jenis, Syarat, dan Ketentuan Lengkap. *Tadkhkirah: Jurnal Terapan Hukum Islam dan Kajian Filsafat Syariah*, 2(3), 233–242.
- Fatmawati, Misbahuddin, & Sanusi, T. N. M. (2024). Analisis Zakat Fitrah dan Zakat Mal dalam Islam. *Jurnal Penelitian Ilmu-Ilmu Sosial*, 1(6), 52–55.
- Fatra, E., Manguma, T. T. F., & Mastura. (2024). Analisis Sentimen Media Sosial Menggunakan Metode Naive Bayes. *Culture Education and Technology Research (CETERA)*, 2(3), 456–466. <https://doi.org/10.31004/ctr.v2i2.166>
- Firmansyach, W. A., Hayati, U., & Arie Wijaya, Y. (2023). Analisa Terjadinya Overfitting dan Underfitting pada Algoritma Naive Bayes dan Decision Tree dengan Teknik Cross Validation. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 262–269. <https://doi.org/10.36040/jati.v7i1.6329>
- Grahesti, A., Hutami, A. S., Sari, N. N., & Rohmah, J. M. (2023). Mengurai Permasalahan Pendistribusian Zakat dengan Analisis SWOT: Studi Kasus di Lembaga Amil Zakat Daarut Tauhid Solo. *Jurnal Ilmiah Ekonomi Islam*, 9(1), 1411–1420.
- Lailatul Zannah, Maslahah, M., Maylinda, M., Saputra, M. R., & Ridwan, M. (2024). Analisis Macam-Macam Zakat dan Wakaf dalam Manajemen Filantropi. *Jurnal Manajemen dan Pendidikan Agama Islam*, 2(5), 140–153. <https://doi.org/10.61132/jmpai.v2i5.507>

- Nazila, J. (2024). Pemanfaatan Teknologi dalam Pengelolaan Zakat di Era Digital. *JSE: Jurnal Sharia Economica*, 3(2), 116–125. <https://doi.org/10.46773/jse.v3i2.1369>
- Nurul, A., & Rizky, F. (2026). Sistem Pendukung Keputusan Berbasis GIS untuk Pemetaan Wilayah Prioritas Pelayanan Publik. *Indonesian Journal of Multidisciplinary Scientific Studies (IJOMSS)*, 4(6), 486–502. <https://doi.org/10.33151/ijomss.v4i6.1136>
- Ramadania, S., Nadila, S., Saharani, C., Mukhtar, K., Chandra, A. A., Pratama, F. H., & Risja, R. A. (2025). Konsep Syarat dan Rukun dalam Fiqih Ibadah: Studi. *Amanah*, 1(1), 12.
- Rasyad Al Fajar, M., & Jannah, M. (2021). Implementasi Undang-Undang Zakat No. 23 Tahun 2011 tentang Pengelolaan Zakat (Studi Kasus LAZISMU Kota Bima). *J-ESA (Jurnal Ekonomi Syariah)*, 4(2), 127–140. <https://doi.org/10.52266/jesa.v4i2.750>
- Widodo, A., & Herlambang, B. A. (2024). Sistem Informasi Geografis Sebaran Stunting dan Tingkat Kemiskinan di Provinsi Jawa Tengah Tahun 2023. *Jurnal Penelitian Sistem Informasi (JPSI)*, 2(1), 55–63. <https://doi.org/10.54066/jpsi.v1i4.1457>