Journal of SAINTEKS
Vol 4 , No 1, November 2026, page 12-20
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

Urgensi Implementasi UU Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi pada Tingkat Kelurahan: Klasifikasi Persepsi Masyarakat Menggunakan Naive Bayes

Linda Wahyuni*, Nazwa Fitri Gufary

Teknik Informatika dan Komputer, Program Studi Sistem Informasi, Universitas Potensi Utama, Medan, Indonesia
Email: [1,*]lindawahyuni391@gmail.com, [2]nzwawaftri28@gmail.com
Email Penulis Korespondensi: lindawahyuni391@gmail.com

Abstrak-Kantor Lurah Pulo Brayan Kota mengelola data pribadi warga dalam pelayanan administrasi sehingga perlindungan data perlu menjadi bagian dari tata kelola layanan publik, terutama setelah berlakunya Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi. Penelitian ini mengklasifikasikan persepsi masyarakat terhadap keamanan dan privasi data digital menggunakan algoritma Naive Bayes pada RapidMiner. Data berasal dari kuesioner 200 responden dengan lima atribut: keamanan data, privasi data, kesadaran digital, kepercayaan digital, dan umur. Evaluasi menggunakan Cross Validation dengan 3, 5, dan 7 folds menghasilkan akurasi masing-masing 84,52%, 88,50%, dan 89,53%. Hasil tertinggi pada 7 folds menunjukkan kemampuan klasifikasi yang menjanjikan, tetapi selisihnya terhadap 5 folds hanya 1,03 poin persentase dan simpangan bakunya lebih besar; karena itu, hasil ini belum cukup untuk menyatakan keunggulan statistik. Secara substantif, dominasi kelas netral (62,5%) dan kecenderungan kelas negatif pada responden yang menilai keamanan, privasi, atau kesadaran digital secara rendah dapat menjadi sinyal awal bagi kelurahan untuk memperjelas prosedur pengelolaan data, membatasi akses, dan meningkatkan literasi privasi. Model ini merupakan alat bantu pemetaan persepsi, bukan audit kepatuhan UU PDP. Validitas dan reliabilitas instrumen serta korelasi antaratribut perlu diuji lebih lanjut sebelum hasil digunakan sebagai dasar keputusan kebijakan.

Kata Kunci: Naive Bayes; UU Perlindungan Data Pribadi; Kelurahan; Privasi Data; Persepsi Masyarakat

1. PENDAHULUAN

Kantor Lurah Pulo Brayan Kota merupakan unit pemerintahan yang berhadapan langsung dengan warga dalam pelayanan administrasi, termasuk proses yang membutuhkan pengumpulan dan pemrosesan data pribadi. Karena itu, perlindungan data bukan hanya persoalan teknis sistem informasi, melainkan bagian dari tanggung jawab pelayanan publik. Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi (UU PDP) memberi kerangka hukum bagi pemrosesan data pribadi dan menuntut pengendalian yang bertanggung jawab. Pada tingkat kelurahan, urgensinya terletak pada praktik sehari-hari: siapa yang dapat mengakses data, untuk tujuan apa data digunakan, bagaimana dokumen disimpan, dan bagaimana warga memperoleh informasi mengenai pemrosesan datanya.

Keamanan data berhubungan dengan perlindungan informasi dari akses, perubahan, pengungkapan, atau kehilangan yang tidak sah, sedangkan privasi berkaitan dengan hak dan kendali individu atas penggunaan informasi pribadinya (Anggen Suari & Sarjana, 2023). Keduanya saling berkaitan, tetapi tidak identik: sistem dapat dianggap aman secara teknis namun tetap menimbulkan persoalan privasi apabila tujuan pengumpulan, akses, atau penggunaan data tidak dijelaskan dengan memadai. Dalam konteks pelayanan kelurahan, persepsi warga menjadi salah satu masukan untuk menilai apakah praktik pengelolaan data dipahami dan dipercaya masyarakat. Persepsi tersebut tidak menggantikan pemeriksaan kepatuhan hukum atau audit keamanan, tetapi dapat membantu mengidentifikasi area komunikasi dan pelayanan yang perlu diperbaiki.

Persepsi masyarakat dapat dipengaruhi pengalaman berinteraksi dengan layanan, pengetahuan mengenai risiko digital, serta tingkat kepercayaan terhadap institusi. Oleh sebab itu, pemetaan persepsi warga relevan sebagai langkah diagnostik awal bagi kelurahan untuk menentukan prioritas edukasi dan perbaikan tata kelola. Namun, hasil survei harus ditafsirkan sesuai batasannya: jawaban responden menggambarkan penilaian subjektif, bukan bukti langsung bahwa suatu sistem telah aman atau melanggar UU PDP.

Penelitian ini menggunakan Naive Bayes bukan untuk menguji hubungan sebab-akibat, melainkan untuk mengklasifikasikan label persepsi (Positif, Netral, dan Negatif) dari atribut kategorikal yang tersedia. Metode ini dipilih karena menghitung probabilitas kelas berdasarkan pola atribut dan relatif sederhana diterapkan pada data survei berkategori. Akan tetapi, Naive Bayes mengasumsikan independensi bersyarat atribut terhadap kelas; asumsi tersebut tidak berarti bahwa korelasi antarvariabel harus selalu nol, dan keterkaitan antara keamanan, privasi, serta kepercayaan tetap berpotensi membuat estimasi probabilitas kurang tepat. Oleh karena itu, hasil klasifikasi perlu dibaca sebagai estimasi prediktif, bukan bukti bahwa atribut-atribut tersebut berdiri sendiri.

Penelitian terdahulu menunjukkan penggunaan Naive Bayes pada konteks yang beragam. Farhan et al. (2023) menerapkannya untuk klasifikasi pasal UU ITE berbasis teks dan melaporkan akurasi 80%, sedangkan Yusuf dan Sari (2022) menggunakannya untuk mengklasifikasikan pemahaman mahasiswa terhadap program MBKM dengan rata-rata akurasi 86,16%. Siregar et al. (2025) mengkaji sentimen terhadap kampanye penggunaan kecerdasan buatan oleh pemerintah dan melaporkan akurasi 75%. Temuan tersebut menunjukkan bahwa Naive Bayes dapat digunakan pada tugas klasifikasi yang berbeda, tetapi angka akurasi antarpublikasi tidak dapat dibandingkan secara langsung karena sumber data, jumlah kelas, distribusi label, dan skema validasinya berbeda.

Copyright © 2026 the author, Page 12
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4 , No 1, November 2026, page 12-20
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

Kesenjangan yang dituju penelitian ini bukan sekadar penerapan algoritma pada objek baru, melainkan penggunaan hasil klasifikasi persepsi warga sebagai masukan awal bagi tata kelola data di unit pelayanan publik tingkat kelurahan dalam konteks UU PDP. Penelitian ini mengolah 200 respons kuesioner dengan atribut keamanan data, privasi data, kesadaran digital, kepercayaan digital, dan umur, lalu membandingkan kinerja pada 3, 5, dan 7 folds. Kontribusinya adalah (1) memberikan potret awal distribusi persepsi pada lokasi penelitian, (2) memperlihatkan pola atribut yang paling membedakan kelas berdasarkan likelihood, dan (3) merumuskan tindak lanjut pelayanan yang dapat dipertimbangkan kelurahan. Kebaruan ini bersifat kontekstual dan aplikatif; penelitian tidak mengklaim sebagai audit kepatuhan UU PDP.

Penelitian ini bertujuan mengklasifikasikan persepsi masyarakat terhadap keamanan dan privasi data digital di Kantor Lurah Pulo Brayan Kota menggunakan Naive Bayes serta membandingkan kinerja model pada Cross Validation 3, 5, dan 7 folds. Hasil diharapkan menjadi bahan pemetaan awal untuk menetapkan prioritas komunikasi dan pengelolaan data pribadi. Karena penelitian berfokus pada klasifikasi, kesimpulan kebijakan dibatasi pada rekomendasi awal dan tidak dimaksudkan sebagai penetapan kepatuhan atau ketidakpatuhan terhadap UU Nomor 27 Tahun 2022.

2. METODE PENELITIAN

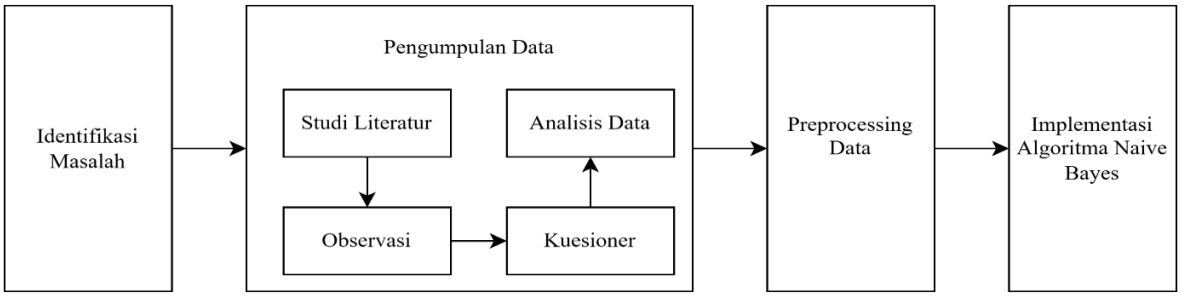
2.1 Kerangka Dasar Penelitian

Penelitian menggunakan pendekatan kuantitatif berbasis klasifikasi data mining. Data yang dianalisis berupa jawaban kuesioner dari 200 responden di lingkungan Kantor Lurah Pulo Brayan Kota. Unit analisis adalah respons individu, sedangkan keluaran model berupa kelas persepsi Positif, Netral, atau Negatif. Desain ini sesuai untuk mengevaluasi kemampuan prediksi label yang tersedia, tetapi tidak dapat digunakan untuk menyimpulkan hubungan kausal antara atribut atau menilai kepatuhan hukum instansi secara langsung.

Atribut prediktor terdiri atas Keamanan Data Digital (X1), Privasi Data (X2), Kesadaran Digital (X3), Kepercayaan Digital (X4), dan Umur (X5). Label target adalah Persepsi Masyarakat (Y), yang terdiri atas kelas Positif, Netral, dan Negatif. Naive Bayes dipilih karena dapat mengolah atribut kategorikal melalui probabilitas prior dan likelihood. Pemilihan metode ini tetap disertai kehati-hatian karena asumsi independensi bersyarat dapat kurang sesuai apabila jawaban tentang keamanan, privasi, dan kepercayaan memiliki keterkaitan konseptual.

2.2 Tahapan Penelitian

Adapun diagram tahapan penelitian yang dilakukan pada Kantor Lurah Pulo Brayan Kota dapat dilihat pada Gambar 1. Tahapan dimulai dari identifikasi masalah, pengumpulan data, *preprocessing* data, dan diakhiri dengan implementasi algoritma Naive Bayes.



Gambar 1. Diagram Tahapan Penelitian

2.2.1 Identifikasi Masalah

Tahap ini dilakukan untuk mengidentifikasi permasalahan yang berkaitan dengan keamanan dan privasi data digital di Kantor Lurah Pulo Brayan Kota, mencakup potensi risiko kebocoran data pribadi, rendahnya kesadaran digital masyarakat, serta kurangnya kepercayaan publik terhadap sistem digital yang digunakan dalam pelayanan administrasi. Hasil identifikasi ini menjadi dasar dalam menentukan lima variabel penelitian dan landasan dalam perancangan kuesioner yang disebarkan kepada 200 responden.

2.2.2 Pengumpulan Data

Pengumpulan data dilakukan melalui studi literatur, observasi, dan kuesioner. Studi literatur digunakan untuk menyusun landasan konseptual variabel; observasi memberikan konteks awal pelayanan administrasi; sedangkan kuesioner menghasilkan data persepsi dari 200 responden. Jawaban diperiksa dan direkapitulasi sebelum dikategorikan untuk proses klasifikasi.

a. Studi literatur, dilakukan dengan mengkaji referensi ilmiah yang berkaitan dengan keamanan data digital, privasi data, kesadaran digital, dan kepercayaan masyarakat terhadap sistem digital sebagai landasan teori dan acuan penyusunan variabel serta instrumen kuesioner.

Copyright © 2026 the author, Page 13
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4 , No 1, November 2026, page 12-20
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

b. Observasi, dilakukan dengan mengamati secara langsung kondisi pengelolaan data digital dan sistem pelayanan administrasi di Kantor Lurah Pulo Brayan Kota untuk memperoleh gambaran awal mengenai tingkat keamanan data, penerapan privasi digital, serta kesadaran aparatur dalam mengelola informasi masyarakat.
c. Kuesioner disebarkan kepada 200 responden dan mencakup lima atribut prediktor serta satu label persepsi sebagaimana dirangkum pada Tabel 1. Untuk menjaga ketepatan interpretasi, butir pertanyaan seharusnya ditelaah kesesuaian isi dan kejelasannya sebelum pengumpulan data. Namun, dokumen penelitian yang tersedia tidak memuat hasil uji validitas empiris, reliabilitas, maupun uji korelasi antaratribut. Karena hasil tersebut tidak tersedia, naskah ini tidak menyatakan bahwa instrumen telah terbukti valid dan reliabel; pengujian tersebut perlu dilakukan menggunakan data respons per butir sebelum artikel mengklaim mutu psikometrik instrumen.
d. Analisis data, dilakukan dengan memeriksa dan merekapitulasi seluruh jawaban dari 200 responden, kemudian mengelompokkan jawaban ke dalam sub variabel yang telah ditentukan sehingga menghasilkan *dataset* terstruktur yang siap digunakan pada tahap *preprocessing*.

Tabel 1. Kuesioner Penelitian

Variabel	Pertanyaan
X1. Keamanan Data Digital (Whitman & Mattord, 2018)	Menurut Anda, bagaimana tingkat keamanan data digital masyarakat yang dikelola oleh Kantor Lurah Pulo Brayan Kota?
X2. Privasi Data Digital (Westin, 1967)	Menurut Anda, bagaimana perlindungan privasi data pribadi masyarakat yang diterapkan oleh Kantor Lurah Pulo Brayan Kota?
X3. Kesadaran Masyarakat (Gilster, 1997)	Menurut Anda, seberapa penting kesadaran masyarakat dalam menjaga dan melindungi data pribadi di era digital?
X4. Kepercayaan Masyarakat (Mayer et al., 1995)	Menurut Anda, seberapa besar tingkat kepercayaan masyarakat terhadap kemampuan Kantor Lurah Pulo Brayan Kota dalam menjaga kerahasiaan dan keamanan data pribadi?
X5. Umur (Santrock, 2019)	Berapakah umur Anda saat ini?
Y. Persepsi Masyarakat (Robbins & Judge, 2017)	Bagaimana persepsi Anda secara keseluruhan terhadap keamanan dan privasi data digital di Kantor Lurah Pulo Brayan Kota?

2.2.3 Preprocessing Data

Preprocessing meliputi pemeriksaan kelengkapan jawaban, penanganan data tidak valid sesuai kriteria penelitian, serta pengubahan jawaban menjadi kategori yang digunakan model. Kategori dan label harus ditetapkan konsisten sebelum validasi silang. Implementasi Naive Bayes dilakukan pada RapidMiner terhadap 200 data hasil *preprocessing*. Label target diambil dari variabel Persepsi Masyarakat (Y), dengan kelas Positif, Netral, dan Negatif. Model menghitung probabilitas posterior setiap kelas berdasarkan probabilitas prior dan *likelihood* atribut. Pada asumsi Naive Bayes, setiap atribut dianggap independen secara bersyarat ketika kelas diketahui. Apabila atribut memiliki keterkaitan kuat, misalnya penilaian keamanan dengan privasi atau kepercayaan, perkalian *likelihood* dapat memberi bobot berlebih pada informasi yang saling tumpang tindih. Karena uji korelasi belum tersedia, potensi ini diperlakukan sebagai keterbatasan interpretasi, bukan dianggap telah terbukti menimbulkan bias.

Implementasi algoritma Naive Bayes dilakukan pada RapidMiner menggunakan 200 data responden hasil *preprocessing* dan operator *Cross Validation* dengan *number of folds* 3, 5, dan 7. Label klasifikasi diambil langsung dari persepsi masyarakat yang terdiri dari tiga kelas, yaitu Positif, Netral, dan Negatif. Algoritma kemudian mengklasifikasikan persepsi berdasarkan kelima variabel penelitian. Dasar perhitungan algoritma ini adalah teorema Bayes yang dinyatakan pada Persamaan (1), dengan X adalah data yang kelasnya belum diketahui, H adalah hipotesis bahwa data X termasuk kelas tertentu, P(H|X) adalah probabilitas hipotesis H berdasarkan kondisi X (*posterior probability*), P(H) adalah probabilitas hipotesis H (*prior probability*), P(X|H) adalah probabilitas X berdasarkan kondisi pada hipotesis H, dan P(X) adalah probabilitas X.

$$P(H|X) = \frac{P(X|H)\cdot P(H)}{P(X)} \quad (1)$$

Probabilitas prior setiap kelas dihitung menggunakan Persamaan (2), dengan K adalah jumlah data pada kelas Ci dan A adalah jumlah seluruh data. Dengan asumsi independensi antar atribut, probabilitas data X terhadap suatu kelas dihitung dengan mengalikan probabilitas setiap atribut seperti pada Persamaan (3), dan kelas hasil klasifikasi ditentukan dari nilai terbesar perkalian antara likelihood dan prior pada Persamaan (4).

$$P(C_i) = \frac{K}{A} \quad (2)$$

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \quad (3)$$

$$P(C_i|X) = P(X|C_i) \cdot P(C_i) \quad (4)$$

2.3 Pengujian dan Evaluasi Model

Evaluasi dilakukan dengan Cross Validation pada 3, 5, dan 7 folds. Dalam setiap iterasi, satu bagian digunakan sebagai data uji dan bagian lain sebagai data latih. Metrik yang dilaporkan meliputi akurasi, presisi, recall, serta simpangan baku

Copyright © 2026 the author, Page 14
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4 , No 1, November 2026, page 12-20
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

yang tersedia dari hasil RapidMiner. Perbandingan ini bersifat deskriptif. Untuk menyatakan bahwa selisih kinerja antarjumlah fold signifikan, diperlukan pengujian berpasangan atas hasil fold atau pengulangan validasi dengan pengaturan yang konsisten; pengujian signifikansi tersebut tidak tercantum dalam data penelitian yang tersedia.

Kinerja model diukur menggunakan *confusion matrix* untuk memperoleh nilai akurasi, presisi, dan *recall*. Akurasi dihitung dengan Persamaan (5), presisi kelas ke-i dengan Persamaan (6), dan *recall* kelas ke-i dengan Persamaan (7), dengan n^{ii} adalah jumlah data kelas ke-i yang diprediksi benar, N adalah jumlah seluruh data, TP adalah jumlah prediksi benar, FP adalah jumlah data kelas lain yang diprediksi sebagai kelas tersebut, dan FN adalah jumlah data kelas tersebut yang diprediksi sebagai kelas lain.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

3. HASIL DAN PEMBAHASAN

Hasil disajikan secara ringkas untuk menjawab tiga hal: distribusi kelas persepsi, pola atribut yang membedakan kelas, dan perbandingan kinerja model. Angka probabilitas atribut digunakan untuk membaca kecenderungan data, sedangkan hasil Cross Validation digunakan untuk mengevaluasi prediksi. Interpretasi dibatasi pada 200 responden dan tidak digeneralisasi sebagai gambaran seluruh warga atau sebagai bukti langsung kondisi keamanan sistem di kantor lurah.

3.1 Data Responden

Data responden setelah preprocessing diringkas pada Tabel 2. Tabel tersebut menampilkan contoh data, sedangkan keseluruhan analisis menggunakan 200 respons.

Tabel 2. Data Responden Hasil *Preprocessing*

Responden	Keamanan Data Digital	Privasi Data	Kesadaran Digital	Kepercayaan Digital	Umur	Persepsi
R1	Cukup Aman	Cukup Terlindungi	Tidak Sadar	Tidak Percaya	Tua	Negatif
R2	Tidak Aman	Terlindungi	Tidak Sadar	Percaya	Dewasa	Netral
R3	Tidak Aman	Tidak Terlindungi	Sadar	Percaya	Dewasa	Negatif
R4	Tidak Aman	Terlindungi	Tidak Sadar	Cukup Percaya	Dewasa	Netral
R5	Cukup Aman	Terlindungi	Tidak Sadar	Tidak Percaya	Dewasa	Netral
R6	Cukup Aman	Terlindungi	Sadar	Cukup Percaya	Muda	Positif
R7	Tidak Aman	Cukup Terlindungi	Sadar	Cukup Percaya	Muda	Netral
R8	Aman	Cukup Terlindungi	Tidak Sadar	Tidak Percaya	Muda	Netral
R9	Tidak Aman	Cukup Terlindungi	Cukup Sadar	Tidak Percaya	Dewasa	Negatif
R10	Tidak Aman	Terlindungi	Cukup Sadar	Tidak Percaya	Dewasa	Netral
...	...	...	...	...	...	...
R200	Aman	Cukup Terlindungi	Sadar	Cukup Percaya	Muda	Positif

Dari 200 respons, kelas Negatif berjumlah 34 (17,0%), Netral 125 (62,5%), dan Positif 41 (20,5%). Dominasi kelas Netral perlu diperhatikan saat membaca akurasi karena model dapat memperoleh nilai tinggi jika lebih sering mengenali kelas mayoritas. Oleh karena itu, presisi, recall, dan pola kesalahan per kelas tetap diperlukan untuk menilai kegunaan model.

3.2 Perhitungan Probabilitas

Probabilitas prior menunjukkan bahwa kelas Netral merupakan kelas mayoritas (62,5%), diikuti Positif (20,5%) dan Negatif (17,0%). Pola likelihood pada Tabel 3–7 menunjukkan perbedaan yang paling jelas pada atribut kesadaran digital dan umur. Karena kelima atribut ditampilkan dalam tabel masing-masing, uraian setelah setiap tabel dibatasi pada pola yang paling relevan agar tidak mengulang angka secara berlebihan.

Selanjutnya dihitung probabilitas bersyarat (likelihood) setiap kategori atribut terhadap masing-masing kelas, yaitu jumlah data pada kategori tersebut dibagi jumlah data pada kelas yang bersangkutan. Sebagai contoh, pada variabel

Copyright © 2026 the author, Page 15
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4 , No 1, November 2026, page 12-20
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

Keamanan Data Digital, P(Aman|Positif) = 27/41 = 0,6585, P(Aman|Netral) = 43/125 = 0,3440, dan P(Aman|Negatif) = 4/34 = 0,1176. Hasil perhitungan seluruh variabel ditunjukkan pada Tabel 3 sampai Tabel 7.

Tabel 3. Probabilitas Keamanan Data Digital (X1)

Keamanan Data Digital	Jumlah Data per Kelas			Probabilitas		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Aman	27	43	4	0,6585	0,3440	0,1176
Cukup Aman	12	47	10	0,2927	0,3760	0,2941
Tidak Aman	2	35	20	0,0488	0,2800	0,5882
Total	41	125	34	1,0000	1,0000	1,0000

Pada atribut keamanan data, kategori "Tidak Aman" lebih sering ditemukan pada kelas Negatif (20 dari 34 responden; likelihood 0,5882), sedangkan kategori "Aman" lebih sering pada kelas Positif (27 dari 41; 0,6585). Pola ini menggambarkan kesesuaian antara penilaian keamanan dan label persepsi, tetapi tidak membuktikan kondisi keamanan teknis aktual.

Tabel 4. Probabilitas Privasi Data (X2)

Privasi Data	Jumlah Data per Kelas			Probabilitas		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Terlindungi	26	43	4	0,6341	0,3440	0,1176
Cukup Terlindungi	13	43	8	0,3171	0,3440	0,2353
Tidak Terlindungi	2	39	22	0,0488	0,3120	0,6471
Total	41	125	34	1,0000	1,0000	1,0000

Pada atribut privasi, kategori "Tidak Terlindungi" lebih dominan pada kelas Negatif (22 dari 34; 0,6471), sedangkan "Terlindungi" lebih dominan pada kelas Positif (26 dari 41; 0,6341). Hasil ini mengisyaratkan bahwa komunikasi mengenai cara data warga dikumpulkan, diakses, dan disimpan dapat menjadi area tindak lanjut yang relevan.

Tabel 5. Probabilitas Kesadaran Digital (X3)

Kesadaran Digital	Jumlah Data per Kelas			Probabilitas		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Sadar	29	36	1	0,7073	0,2880	0,0294
Cukup Sadar	11	50	7	0,2683	0,4000	0,2059
Tidak Sadar	1	39	26	0,0244	0,3120	0,7647
Total	41	125	34	1,0000	1,0000	1,0000

Kesadaran digital memperlihatkan pemisahan likelihood yang kuat: kategori "Tidak Sadar" memiliki likelihood 0,7647 pada kelas Negatif, sedangkan kategori "Sadar" memiliki likelihood 0,7073 pada kelas Positif. Karena label persepsi dan penilaian kesadaran sama-sama berasal dari kuesioner, pola ini harus dipahami sebagai keterkaitan dalam respons survei, bukan hubungan sebab-akibat.

Tabel 6. Probabilitas Kepercayaan Digital (X4)

Kepercayaan Digital	Jumlah Data per Kelas			Probabilitas		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Percaya	26	30	2	0,6341	0,2400	0,0588
Cukup Percaya	13	43	8	0,3171	0,3440	0,2353
Tidak Percaya	2	52	24	0,0488	0,4160	0,7059
Total	41	125	34	1,0000	1,0000	1,0000

Pada atribut kepercayaan, kategori "Tidak Percaya" paling sering muncul pada kelas Negatif (likelihood 0,7059), sementara kategori "Percaya" lebih banyak pada kelas Positif (0,6341). Namun, kategori "Tidak Percaya" juga muncul pada kelas Netral (likelihood 0,4160), sehingga kepercayaan tidak sepenuhnya memisahkan kelas secara tunggal.

Tabel 7. Probabilitas Umur (X5)

Umur	Jumlah Data per Kelas			Probabilitas		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Muda	38	9	5	0,9268	0,0720	0,1471
Dewasa	2	106	4	0,0488	0,8480	0,1176
Tua	1	10	25	0,0244	0,0800	0,7353
Total	41	125	34	1,0000	1,0000	1,0000

Atribut umur menunjukkan perbedaan likelihood yang menonjol: kategori Muda dominan pada kelas Positif (0,9268), Dewasa pada kelas Netral (0,8480), dan Tua pada kelas Negatif (0,7353). Temuan ini merupakan pola pada

Copyright © 2026 the author, Page 16
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4 , No 1, November 2026, page 12-20
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

sampel penelitian dan tidak boleh ditafsirkan sebagai karakteristik tetap semua kelompok umur. Faktor lain yang tidak diukur dapat turut menjelaskan perbedaan persepsi.

3.3 Perhitungan Probabilitas Posterior

Data hasil kuesioner dari 200 responden diolah dan diklasifikasikan menggunakan algoritma Naive Bayes berdasarkan probabilitas prior dan likelihood yang telah diperoleh. Sebagai contoh, perhitungan dilakukan pada data responden ke-1 dengan Keamanan Data Digital = Cukup Aman, Privasi Data = Cukup Terlindungi, Kesadaran Digital = Tidak Sadar, Kepercayaan Digital = Tidak Percaya, dan Umur = Tua, sebagaimana ditunjukkan pada Tabel 8.

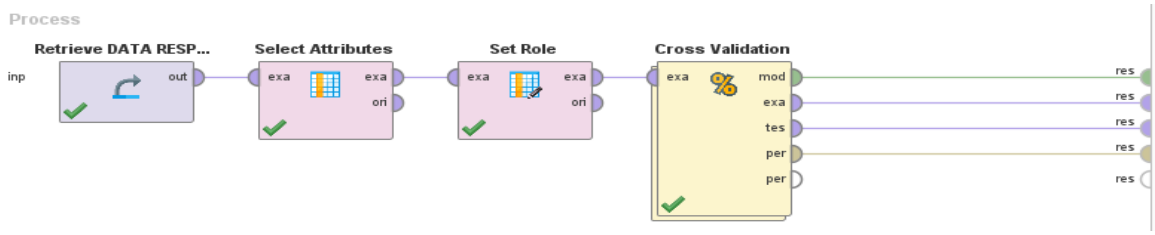
Tabel 8. Probabilitas Posterior Data Responden Ke-1

Variabel	Nilai Atribut	Probabilitas Kelas		
		Positif	Netral	Negatif
Keamanan Data Digital	Cukup Aman	0,2927	0,3760	0,2941
Privasi Data	Cukup Terlindungi	0,3171	0,3440	0,2353
Kesadaran Digital	Tidak Sadar	0,0244	0,3120	0,7647
Kepercayaan Digital	Tidak Percaya	0,0488	0,4160	0,7059
Umur	Tua	0,0244	0,0800	0,7353
P(X|Ci)		0,00000270	0,00134303	0,02746729
P(X|Ci) × P(Ci)		0,00000055	0,00083939	0,00466944

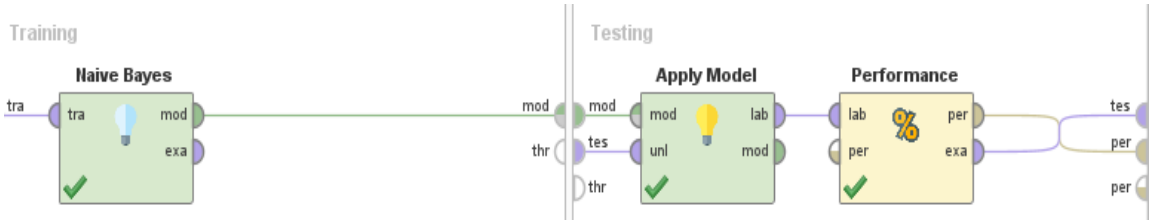
Untuk contoh responden pertama, hasil perkalian prior dan likelihood menghasilkan nilai tertinggi pada kelas Negatif (0,00466944), dibandingkan Netral (0,00083939) dan Positif (0,00000055). Model dengan demikian memprediksi kelas Negatif, sesuai label yang dicatat pada dataset. Contoh satu responden menjelaskan mekanisme perhitungan, tetapi bukan bukti kinerja model secara keseluruhan.

3.4 Pengujian Algoritma Naive Bayes pada RapidMiner

Pengujian algoritma Naive Bayes dilakukan menggunakan *tools* RapidMiner dengan susunan operator yang ditunjukkan pada Gambar 2 dan Gambar 3. Pengujian akurasi menggunakan operator *Cross Validation* dengan *number of folds* 3, 5, dan 7, dengan ukuran kinerja berupa *accuracy*, *precision*, *recall*, dan *confusion matrix*.



Gambar 2. Operator *Process* RapidMiner



Gambar 3. Operator *Training* dan *Testing* Process RapidMiner

Hasil evaluasi menunjukkan akurasi 89,53% (±5,58%) pada 7 folds, 88,50% (±3,35%) pada 5 folds, dan 84,52% (±3,01%) pada 3 folds. Ringkasan perbandingan disajikan pada Tabel 12 agar pembaca dapat menilai seluruh skema tanpa mengulang uraian setiap tabel.

accuracy: 88.50% +/- 3.35% (micro average: 88.50%)

	true Negatif	true Netral	true Positif	class precision
pred. Negatif	33	8	0	80.49%
pred. Netral	1	107	4	95.54%
pred. Positif	0	10	37	78.72%
class recall	97.06%	85.60%	90.24%	

Gambar 4. *Confusion Matrix* dengan *Number of Folds* 7

Copyright © 2026 the author, Page 17
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4 , No 1, November 2026, page 12-20
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

accuracy: 84.52% +/- 3.01% (micro average: 84.50%)

	true Negatif	true Netral	true Positif	class precision
pred. Negatif	30	17	0	63.83%
pred. Netral	4	100	2	94.34%
pred. Positif	0	8	39	82.98%
class recall	88.24%	80.00%	95.12%	

Gambar 5. *Confusion Matrix* dengan *Number of Folds* 5

accuracy: 89.53% +/- 5.58% (micro average: 89.50%)

	true Negatif	true Netral	true Positif	class precision
pred. Negatif	32	8	0	80.00%
pred. Netral	2	110	4	94.83%
pred. Positif	0	7	37	84.09%
class recall	94.12%	88.00%	90.24%	

Gambar 6. *Confusion Matrix* dengan *Number of Folds* 3

Tabel 12. Perbandingan Kinerja Model pada *Number of Folds* 3, 5, dan 7

Number of Folds	Akurasi (± SD)	Micro Average	Rata-rata Presisi	Rata-rata Recall
3	84,52% ± 3,01%	84,50%	80,38%	87,79%
5	88,50% ± 3,35%	88,50%	84,92%	90,97%
7	89,53% ± 5,58%	89,50%	86,31%	90,79%

Secara deskriptif, skema 7 folds memperoleh akurasi tertinggi (89,53%), disusul 5 folds (88,50%) dan 3 folds (84,52%). Namun, selisih antara 7 dan 5 folds hanya 1,03 poin persentase, sementara simpangan baku 7 folds lebih besar. Oleh sebab itu, 7 folds dapat disebut menghasilkan akurasi teramati tertinggi pada pengujian ini, tetapi belum dapat dinyatakan secara statistik lebih unggul daripada 5 folds. Perbandingan yang adil juga mensyaratkan pembagian data, stratifikasi kelas, dan pengaturan *random seed* yang terdokumentasi.

3.5 Pembahasan

Perbedaan akurasi antar-fold tidak otomatis menunjukkan bahwa jumlah fold yang lebih besar menghasilkan model yang secara intrinsik lebih baik. Jumlah data latih per iterasi memang berbeda, tetapi variasi komposisi data pada setiap fold juga memengaruhi hasil. Karena selisih akurasi 7 dan 5 folds kecil dan simpangan baku pada 7 folds lebih besar, temuan ini lebih tepat dibaca sebagai indikasi bahwa kedua skema memberi performa serupa pada dataset ini. Tanpa pengujian statistik berpasangan atau pengulangan validasi, klaim keunggulan harus dibatasi.

Confusion matrix menunjukkan kesalahan terutama terjadi antara kelas Netral dan kelas Positif atau Negatif. Kelas Netral sebagai kelas mayoritas memiliki presisi relatif tinggi, sedangkan sebagian respons Netral diprediksi sebagai Negatif. Konsekuensinya, model berpotensi menandai sebagian warga yang sebenarnya netral sebagai kelompok dengan persepsi negatif. Untuk penggunaan praktis, prediksi negatif sebaiknya menjadi sinyal untuk evaluasi komunikasi dan layanan, bukan dasar untuk memberi label atau perlakuan berbeda kepada individu.

Pola umur menunjukkan responden kategori Muda lebih sering berada pada kelas Positif, sedangkan kategori Tua lebih sering berada pada kelas Negatif. Pola ini dapat menjadi dasar untuk merancang komunikasi perlindungan data dengan bahasa dan media yang mudah diakses berbagai kelompok umur. Akan tetapi, penelitian ini tidak menguji mekanisme penyebabnya, sehingga tidak tepat menyimpulkan bahwa usia secara langsung menentukan persepsi atau kemampuan digital seseorang.

Sebagai pembanding deskriptif, jika semua responden selalu diprediksi sebagai kelas mayoritas Netral, akurasi yang diperoleh adalah 125/200 atau 62,5%. Akurasi Naive Bayes pada tiga skema (84,52%–89,53%) berada di atas baseline sederhana tersebut. Perbandingan ini menunjukkan bahwa model memanfaatkan pola atribut lebih baik daripada strategi yang hanya menebak kelas mayoritas, tetapi belum menggantikan kebutuhan evaluasi validitas instrumen dan pengujian model yang lebih kuat.

Pada kelas Positif, recall yang dilaporkan adalah 90,24% pada 5 dan 7 folds serta 95,12% pada 3 folds. Presisi yang lebih rendah dibandingkan recall menunjukkan bahwa sebagian respons dari kelas lain ikut diprediksi sebagai Positif. Perbedaan presisi dan recall ini menegaskan bahwa akurasi tunggal tidak cukup untuk menilai kesesuaian model bagi kebutuhan pelayanan.

Berdasarkan likelihood, atribut umur dan kesadaran digital tampak memiliki daya pembeda deskriptif yang menonjol. Namun, nilai likelihood bukan uji kepentingan atribut dan tidak menunjukkan kontribusi independen setelah pengaruh atribut lain dikendalikan. Karena korelasi antaratribut belum diuji, kesimpulan mengenai peran atribut harus diperlakukan sebagai indikasi awal yang perlu diverifikasi.

Copyright © 2026 the author, Page 18
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Akurasi penelitian ini tampak lebih tinggi daripada angka yang dilaporkan pada beberapa studi terdahulu, tetapi perbandingan tersebut tidak dapat dijadikan klaim bahwa model ini lebih unggul secara umum. Studi terdahulu menggunakan jenis data, label, jumlah sampel, dan prosedur evaluasi yang berbeda. Nilai utama penelitian ini terletak pada penerapan klasifikasi terhadap persepsi warga pada konteks layanan kelurahan dan penerjemahan hasilnya menjadi masukan perbaikan layanan.

Keterbatasan penelitian meliputi cakupan responden yang hanya berasal dari satu lokasi, dominasi kelas Netral, belum tersedianya laporan uji validitas dan reliabilitas kuesioner, serta belum dilakukannya pengujian korelasi antaratribut. Asumsi independensi bersyarat Naive Bayes juga berpotensi kurang sesuai untuk atribut keamanan, privasi, dan kepercayaan yang secara konseptual berkaitan. Selain itu, perbandingan 3, 5, dan 7 folds masih bersifat deskriptif dan hanya menggunakan satu algoritma. Penelitian lanjutan perlu menguji instrumen, menganalisis keterkaitan atribut, mendokumentasikan pengaturan validasi, mengulang pengujian, dan membandingkan model dengan algoritma alternatif.

4. KESIMPULAN

Penelitian ini mengklasifikasikan persepsi 200 responden terhadap keamanan dan privasi data digital di Kantor Lurah Pulo Brayan Kota menggunakan Naive Bayes pada RapidMiner. Kelas Netral merupakan kelompok terbesar (62,5%), disusul Positif (20,5%) dan Negatif (17,0%). Akurasi teramati tertinggi diperoleh pada 7 folds sebesar 89,53%, dibandingkan 88,50% pada 5 folds dan 84,52% pada 3 folds. Karena selisih antara 7 dan 5 folds kecil dan simpangan baku pada 7 folds lebih besar, hasil ini belum membuktikan keunggulan statistik salah satu skema. Pola likelihood memperlihatkan bahwa kesadaran digital, penilaian keamanan dan privasi, serta umur berkaitan dengan distribusi kelas pada sampel, tetapi tidak dapat digunakan untuk menyimpulkan hubungan sebab-akibat atau kondisi keamanan teknis yang sebenarnya. Bagi Kantor Lurah Pulo Brayan Kota, hasil ini dapat digunakan sebagai masukan awal untuk: (1) menyusun atau meninjau SOP pengumpulan, penggunaan, penyimpanan, retensi, dan pemusnahan data pribadi; (2) membatasi akses data sesuai tugas petugas dan meninjau pencatatan akses; (3) memberikan penjelasan yang mudah dipahami kepada warga mengenai tujuan pemrosesan data dan kanal pengaduan; serta (4) menyelenggarakan edukasi berkala bagi aparatur dan warga tentang perlindungan data. Rekomendasi tersebut merupakan langkah tata kelola awal dan bukan pernyataan bahwa kantor telah atau belum mematuhi UU Nomor 27 Tahun 2022. Sebelum model dijadikan dasar keputusan, penelitian perlu melengkapi uji validitas dan reliabilitas instrumen, analisis korelasi antaratribut, serta evaluasi model berulang dan/atau uji statistik perbandingan. Penelitian berikutnya juga perlu memperluas sampel dan membandingkan Naive Bayes dengan algoritma lain.

REFERENCES

Akbar, Y., & Qibthiyah, M. (2025). Penerapan Metode Naïve Bayes untuk Klasifikasi Produk Berdasarkan Kategori Penjualan di Toko Artemist. Jurnal Indonesia: Manajemen Informatika dan Komunikasi, 6(3), 1921–1930. https://doi.org/10.63447/jimik.v6i3.1603

Alvember, M., & Asri, D. P. B. (2025). Analisis Keamanan Data Pribadi Konsumen terhadap Regulasi Pencantuman Nomor Induk Kependudukan untuk Registrasi Nomor Telepon. *Arus Jurnal Sosial dan Humaniora*, *5*(2), 3017–3026. https://doi.org/10.57250/ajsh.v5i2.1590

Anggen Suari, K. R., & Sarjana, I. M. (2023). Menjaga Privasi di Era Digital: Perlindungan Data Pribadi di Indonesia. *Jurnal Analisis Hukum*, *6*(1), 132–142. https://doi.org/10.38043/jah.v6i1.4484

Belaga, N. A., Udju, H. R., & Ermalinda, J. (2025). Tugas dan fungsi lurah dalam pelayanan publik di Kelurahan Weri, Kecamatan Larantuka, Kabupaten Flores Timur. Petitum Law Journal, 3(1), 84–102. https://doi.org/10.35508/pelana.v3i1.21301

Farhan, Triase, & Harahap, A. M. (2023). Penggunaan Algoritma Naive Bayes Dalam Text Mining Untuk Klasifikasi Pasal UU ITE. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, *6*(2), 314. https://doi.org/10.53513/jsk.v6i2.7896

Fauziah, & Ramadhantya, A. S. (2024). Penggunaan Rapidminer untuk memprediksi kelulusan mahasiswa dengan algorithm Naive Bayes. Jurnal Teknologi Informasi, 10(1), 52–60. https://doi.org/10.52643/jti.v10i1.4671

Lesmana, F., Rizkimu, A. D., Nurrulloh, M. R., Fathurrahman, M. F., Hasibuan, A. H., & Fansyuri, M. (2025). Analisis kinerja algoritma Naive Bayes dalam klasifikasi data kategorikal prediksi keputusan bermain tenis berdasarkan cuaca. PT Jurnal Cendekia Indonesia, 1(1), 67–72. https://journal.jci.co.id/jitie/id/article/download/146/32/429

Mufid, S. F., & Alkadry, A. U. (2026). Keamanan digital di era transformasi teknologi: Peran literasi dan perilaku dalam perlindungan diri. Qolamuna: Keislaman, Pendidikan, Literasi dan Humaniora, 3(1), 251–264. https://jurnal.qolamuna.id/index.php/JQ/article/view/406

Mulyana, D. I., & Fadhil, A. M. (2024). Penerapan Data Mining untuk Menentukan Produk Terlaris pada PT Joynare Mitra Teknologi menggunakan Metode Naive Bayes. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, *5*(2), 1206–1222. https://doi.org/10.35870/jimik.v5i2.638

Oktaviana Dewi, I., Wahyudi, I., Setiawan, N., & Uyun, J. (2023). Fraud Ditinjau dari Etika Profesi dan Etika Bisnis Kasus PT Garuda Indonesia. *Media Komunikasi Ilmu Ekonomi*, *40*(1), 41–53. https://doi.org/10.58906/melati.v40i1.101

Ongkowiguno, C. M., & Marsal, I. (2025). Politik hukum perlindungan data pribadi dalam layanan publik digital di Indonesia. Judge: Jurnal Hukum, 6(4), 1321–1326. https://doi.org/10.54209/judge.v6i04.1851

Putri Sumarsono, S. I., & Nurmiati, E. (2026). Etika teknologi dalam perspektif Islam: Tantangan privasi dan keamanan data di era digital. IHSAN: Jurnal Penelitian dan Pendidikan Islam, 2(1), 23–28. https://doi.org/10.66031/ihsan.v2i1.189

Roy Hamdan, & Zaehol fatah. (2026). Analisis Opini Publik di YouTube terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes. *Jurnal Mahasiswa Teknik Informatika*, *5*(1), 259–265. https://doi.org/10.35473/jamastika.v5i1.4662

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Saádi, A. (2025). Pengumpulan data yang efisien pada penelitian tindakan kelas: Teknik, alat, dan tantangan. Al-Amin: Jurnal Ilmu Pendidikan dan Sosial Humaniora, 2(2), 90–108. https://doi.org/10.53398/alamin.v2i2.377

Siregar, K. A., Nasution, S., & Nabawy, P. (2025). Analisis sentimen netizen Indonesia terhadap kampanye penggunaan kecerdasan buatan oleh pemerintah menggunakan algoritma Naive Bayes. Jurnal Ilmu Komputer dan Informatika, 1(4), 112–121. https://garuda.kemdiktisaintek.go.id/documents/detail/5252824

Suriyani BB, Elwan, L. O. M., Iwan P, & Usman. (2025). Pelaksanaan Tertib Administrasi Dalam Meningkatkan Kualitas Pelayanan Publik Pada Kantor Lurah Nambo Kecamatan Nambo Kota Kendari. *Journal Publicuho*, *8*(3), 1811–1824. https://doi.org/10.35817/publicuho.v8i3.903

Yusuf, A. Y. P., & Sari, R. (2022). Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Pemahaman Program MBKM Bagi Mahasiswa. *Journal of Informatic and Information Security*, *3*(2), 171–180. https://doi.org/10.31599/jiforty.v3i2.1713

Copyright © 2026 the author, Page 20
This Journal is licensed under a Creative Commons Attribution 4.0 International License