Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT
Penerbit : Cv. Utility Project Solutio

Penerapan Algoritma Naive Bayes Untuk Klasifikasi Kesiapan Administrasi pada KUA Kecamatan Kapuas Barat

Rahma Meina Riwanda[1], Finki Dona Marleny[2], Windarsyah[3]

[1,2,3)]Informatika, Universitas Muhammadiyah Banjarmasin, Barito Kuala, Indonesia

Email: [1]rahma_meina_2255201110015@umbjm.ac.id, [2]finkidona@umbjm.ac.id, [3]windarsyah@umbjm.ac.id,
Email Penulis Korespondensi: [1]rahma_meina_2255201110015@umbjm.ac.id

Info Artikel

Kata Kunci:
Data Mining
Kesiapan Administrasi Pernikahan
Klasifikasi
KUA
Naive Bayes

Abstrak

Kantor Urusan Agama (KUA) merupakan instansi pemerintah yang bertanggung jawab menyelenggarakan pelayanan pencatatan serta pengelolaan administrasi pernikahan bagi masyarakat Muslim sesuai ketentuan yang berlaku. Pengelolaan data yang masih manual menyebabkan kesulitan dalam memverifikasi kesiapan administrasi calon pengantin secara cepat dan akurat. Penelitian terdahulu masih terbatas pada dataset kecil, kriteria tunggal berbasis usia, dan tanpa *cross-validation*, sehingga belum tersedia model klasifikasi kesiapan administrasi pernikahan yang komprehensif dan andal. Penelitian ini menerapkan algoritma Naive Bayes untuk mengklasifikasikan kesiapan administrasi pernikahan berdasarkan kombinasi usia, pendidikan, dan status pernikahan pada KUA Kecamatan Kapuas Barat, dengan memanfaatkan 1.010 *record* data periode 2021–2025 yang dipisahkan menjadi 70% data pelatihan dan 30% data pengujian, diimplementasikan melalui Python dan *scikit-learn*. Evaluasi dengan *Confusion Matrix* menghasilkan akurasi 92,67%, presisi 93,33%, *recall* 98,44%, dan *F1-Score* 95,82%, sementara *K-Fold Cross Validation* (k=10) menghasilkan rata-rata akurasi 90,10% dengan standar deviasi 2,77%. Hasil ini membuktikan bahwa Naive Bayes efektif untuk klasifikasi kesiapan administrasi pernikahan dengan akurasi melampaui penelitian sebelumnya (84%), serta berkontribusi sebagai model pendukung keputusan berbasis data mining bagi petugas KUA dalam verifikasi administrasi yang lebih cepat, konsisten, dan akurat.

Keywords:
Classification
Data Mining
KUA
Marriage Administrative Readiness
Naive Bayes

Abstract

The Office of Religious Affairs (Kantor Urusan Agama/KUA) is a government institution responsible for providing marriage registration services and administrative management for Muslim communities in accordance with applicable regulations. Manual data management has caused difficulties in verifying the administrative readiness of prospective couples quickly and accurately. Previous studies were limited to small datasets, single age-based criteria, and lacked cross-validation, leaving a gap in comprehensive and reliable classification models for marriage administrative readiness. This study applies the Naive Bayes algorithm to classify marriage administrative readiness based on a combination of age, education, and marital status at KUA Kapuas Barat District, utilizing 1,010 marriage records from the 2021–2025 period, which were divided into 70% training data and 30% testing data, implemented using Python and scikit-learn. Evaluation using a Confusion Matrix yielded an accuracy of 92.67%, precision of 93.33%, recall of 98.44%, and F1-Score of 95.82%, while K-Fold Cross Validation (k=10) produced an average accuracy of 90.10% with a standard deviation of 2.77%. These results demonstrate that Naive Bayes is effective for marriage administrative readiness classification, surpassing a previous study's accuracy of 84%, and contributes a data mining-based decision support model to help KUA officers verify marriage administrative readiness more quickly, consistently, and accurately.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

1. PENDAHULUAN

Kantor Urusan Agama (KUA) adalah unit pelaksana di tingkat kecamatan yang memiliki kewenangan resmi dalam pencatatan dan pengelolaan administrasi pernikahan bagi masyarakat Muslim. Lembaga ini memiliki peran strategis dalam memastikan setiap pernikahan memenuhi persyaratan hukum yang berlaku, termasuk melakukan verifikasi kesiapan administrasi pasangan calon pengantin sebelum akad nikah dilangsungkan. KUA Kecamatan Kapuas Barat, Kabupaten Kapuas, Kalimantan Tengah, tercatat menangani 1.010 peristiwa pernikahan selama periode 2021–2025. KUA ini dipilih sebagai lokasi penelitian karena merupakan salah satu KUA dengan volume pencatatan pernikahan tertinggi di Kabupaten Kapuas, sekaligus masih menjalankan proses verifikasi administrasi secara sepenuhnya manual hingga saat penelitian dilakukan. Besarnya volume data pernikahan tersebut menuntut adanya sistem verifikasi yang cepat, akurat, dan konsisten agar pelayanan administrasi dapat berjalan secara optimal [1]. Namun, verifikasi kesiapan administrasi pasangan calon pengantin yang menjadi inti dari proses pelayanan tersebut hingga saat ini masih dilakukan secara manual oleh petugas, sehingga memerlukan pendekatan yang lebih efisien dan sistematis.

Hingga saat ini, proses pengecekan kesiapan administrasi pernikahan di KUA Kecamatan Kapuas Barat masih sepenuhnya dikerjakan secara manual oleh petugas. Berdasarkan ketentuan UU No.16 Tahun 2019 tentang Perkawinan, kesiapan administrasi pasangan ditentukan berdasarkan kombinasi faktor usia, tingkat pendidikan, dan status pernikahan sebelumnya [2]. Penilaian secara manual terhadap ketiga faktor tersebut membutuhkan waktu yang tidak sedikit dan rentan terhadap inkonsistensi antar petugas, sehingga berpotensi menurunkan akurasi dan kualitas hasil verifikasi, khususnya saat volume pendaftaran meningkat Idealnya, proses verifikasi kesiapan administrasi pernikahan dapat dilakukan secara otomatis, cepat, dan objektif dengan memanfaatkan pendekatan berbasis data. Kemajuan teknologi *machine learning* memberikan peluang bagi instansi pemerintah untuk membangun sistem pengambilan keputusan otomatis berbasis pola dari data historis [3]. Penggunaan metode data mining pada data pernikahan terbukti dapat menghasilkan informasi bermanfaat dalam menunjang efisiensi pengelolaan administrasi pemerintah [4]. Teknik klasifikasi dalam data mining memungkinkan sistem memprediksi kategori suatu data secara otomatis berdasarkan atribut-atribut yang relevan, sehingga sangat berpotensi diterapkan dalam proses verifikasi kesiapan administrasi pernikahan di lingkungan KUA, sebagaimana telah dibuktikan efektivitasnya dalam penentuan kelayakan layanan administrasi publik lainnya seperti BPJS Kesehatan [5]. Kondisi ini semakin relevan mengingat pemerintah Indonesia tengah mendorong transformasi digital layanan publik di seluruh instansi, termasuk di tingkat kecamatan, sebagai bagian dari agenda modernisasi tata kelola pemerintahan.

Beberapa penelitian terdahulu telah mencoba menerapkan algoritma klasifikasi pada kasus administrasi pernikahan dan layanan publik sejenis. Hadid dan Eliyani menerapkan Naive Bayes pada data KUA Palmerah dengan 300 data dan menghasilkan akurasi 84%, namun hanya menggunakan kriteria usia tanpa mempertimbangkan faktor pendidikan dan status pernikahan [6]. Izzi, Setyanto, dan Hartanto menerapkan Naive Bayes dengan metode SMOTE pada data numerik dan membuktikan algoritma ini efektif menangani ketidakseimbangan kelas, namun tidak diterapkan pada domain administrasi pernikahan secara langsung [7]. Apriliyani dan Salim menganalisis performa Naive Bayes pada unbalanced dataset dan menegaskan pentingnya penggunaan metrik presisi, recall, dan F1-Score secara bersamaan, namun tidak menyertakan validasi K-Fold Cross Validation untuk mengukur kestabilan model [8]. Depari, Widiastiwi, dan Santoni membandingkan Naive Bayes, Decision Tree, dan Random Forest pada data medis dan menghasilkan akurasi 87,3%, membuktikan keandalan Naive Bayes pada berbagai domain data meskipun tidak spesifik pada domain pernikahan [9]. Fauziah, Tambunan, dan Susiani mengklasifikasikan kesiapan menikah pada dewasa muda menggunakan Naive Bayes dan memperoleh hasil cukup baik, namun tidak menggunakan data administrasi KUA secara langsung sehingga kurang representatif untuk konteks verifikasi dokumen pernikahan [10]. Dari kelima penelitian tersebut, terdapat gap konsisten: belum ada model klasifikasi kesiapan administrasi pernikahan yang dibangun dengan dataset besar, mengintegrasikan tiga faktor penentu sekaligus (usia, pendidikan, dan status pernikahan), serta dilengkapi validasi cross-validation untuk memastikan kestabilan model.

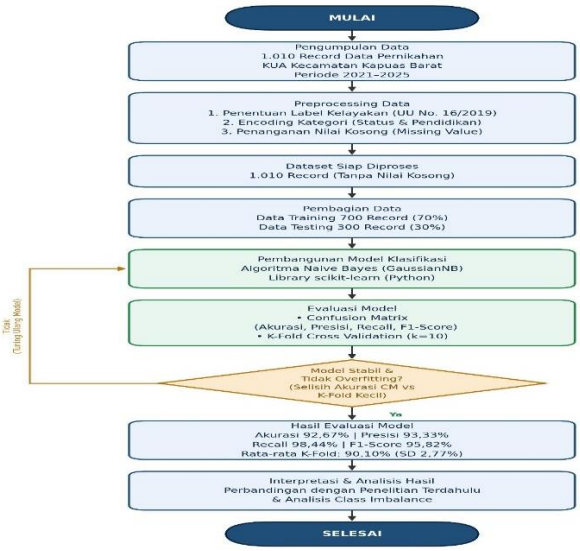
Algoritma Naive Bayes dipilih dalam penelitian ini karena terbukti efektif dalam menangani klasifikasi data kategorik pada konteks administratif, sebagaimana ditunjukkan oleh Irfayanti yang berhasil menerapkan Naive Bayes untuk mengklasifikasikan status pegawai pada perusahaan swasta menggunakan data kategorik yang karakteristiknya serupa dengan data administrasi instansi [11]. Selain itu, Naive Bayes memiliki kompleksitas komputasi yang rendah serta kemampuan adaptasi yang baik terhadap variasi ukuran dan distribusi dataset, termasuk pada kondisi data yang tidak seimbang, sebagaimana dijelaskan oleh Wang, Yan, Liu, dan Li[12]. Berdasarkan gap tersebut, penelitian ini bertujuan menerapkan algoritma Naive Bayes untuk mengklasifikasikan kesiapan administrasi pernikahan pada KUA Kecamatan Kapuas Barat menggunakan dataset 1.010 record periode 2021–2025, dengan kriteria yang lebih komprehensif mencakup tiga faktor sekaligus yaitu usia, pendidikan, dan status pernikahan, serta dilengkapi validasi menggunakan K-Fold Cross Validation (k=10) untuk menghasilkan model yang lebih andal dibandingkan penelitian sebelumnya.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

2. METODOLOGI PENELITIAN

Studi ini dilaksanakan melalui empat tahapan pokok, yaitu pengumpulan data, preprocessing, pembangunan model dengan algoritma Naive Bayes, serta evaluasi menggunakan Confusion Matrix dan K-Fold Cross Validation. Implementasi seluruh tahapan dilakukan menggunakan bahasa pemrograman Python dengan library scikit-learn. Tahapan penelitian secara keseluruhan digambarkan pada diagram alur berikut.



Gambar 1. Diagram Alur (Flowchart) Penelitian

Gambar 1 menunjukkan bahwa penelitian diawali dengan pengumpulan data pernikahan, dilanjutkan dengan preprocessing data yang mencakup penentuan label kelayakan, encoding kategori, dan penanganan nilai kosong, kemudian data dibagi menjadi data training dan testing sebelum masuk ke tahap pembangunan model Naive Bayes. Model yang dihasilkan selanjutnya dievaluasi menggunakan Confusion Matrix dan K-Fold Cross Validation untuk memastikan kestabilan performa sebelum diinterpretasikan lebih lanjut.

2.1 Pengumpulan Data

Data yang digunakan adalah data pernikahan KUA Kecamatan Kapuas Barat periode 2021–2025 dengan total 1.010 record. Data diperoleh dari arsip administrasi KUA melalui teknik dokumentasi dan wawancara dengan dua orang petugas administrasi yang bertanggung jawab langsung dalam pencatatan data pernikahan. Wawancara dilakukan dalam dua sesi, yaitu sesi pertama untuk memahami alur proses verifikasi kelayakan yang selama ini berjalan secara manual, dan sesi kedua untuk memvalidasi atribut data yang akan digunakan dalam penelitian. Data arsip yang diperoleh selanjutnya diverifikasi kesesuaiannya dengan buku register nikah untuk memastikan keakuratan dan kelengkapan setiap record. Kriteria kesiapan administrasi yang menjadi dasar penentuan label juga telah dikonsultasikan dan divalidasi bersama Kepala KUA Kecamatan Kapuas Barat pada sesi wawancara kedua, guna memastikan kesesuaiannya dengan praktik verifikasi yang berlaku secara operasional. Atribut yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Atribut Data Pernikahan

NO	Atribut	Tipe Data	Nilai
1.	Usia Pria	Numerik	21–55 tahun
2.	Usia Wanita	Numerik	21–55 tahun
3.	Status Pria	Kategorik	Jejaka / Duda
4.	Status Wanita	Kategorik	Perawan / Janda
5.	Pendidikan Pria	Kategorik	SD / MI / SMP / SMA / S1
6.	Pendidikan Wanita	Kategorik	SD / MI / SMP / SMA / S1
7.	Kelayakan (Label)	Kategorik	Layak / Tidak Layak

2.2 Preprocessing Data

Tahap *preprocessing* dilakukan untuk mempersiapkan data sebelum masuk ke proses klasifikasi, meliputi beberapa langkah berikut.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

A. Penentuan Label Kelayakan

Kriteria kelayakan ditetapkan berdasarkan kombinasi faktor usia, pendidikan, dan status pernikahan sesuai Undang-Undang Nomor 16 Tahun 2019 serta mempertimbangkan kesiapan sosial pasangan. Aturan penentuan label adalah sebagai berikut [2]:

1) **Tidak layak**, jika memenuhi salah satu kondisi:
   1. Usia pria < 27 tahun DAN pendidikan pria SD/MI/SMP
   2. Usia wanita < 23 tahun DAN pendidikan wanita SD/MI/SMP
   3. Status pria = Duda DAN usia pria < 27 tahun
   4. Status wanita = Janda DAN usia wanita < 23 tahun
2) **Layak**, jika tidak memenuhi satu pun kriteria di atas.

B. Encoding Kategori

Atribut kategori dikonversi menjadi nilai numerik agar dapat diproses oleh algoritma, dengan skema sebagai berikut: status pria (Jejaka=0, Duda=1), status wanita (Perawan=0, Janda=1), dan tingkat pendidikan (SD/MI=1, SMP=2, SMA=3, S1=4).

C. Penanganan Nilai Kosong

Pemeriksaan nilai kosong (*missing value*) dilakukan terhadap seluruh 1.010 *record* sebelum proses klasifikasi. Hasil pemeriksaan menunjukkan bahwa tidak ditemukan nilai kosong pada dataset ini, sehingga seluruh data dapat langsung digunakan tanpa perlu penghapusan baris. Kondisi ini dimungkinkan karena data bersumber langsung dari arsip buku register nikah KUA yang telah terisi secara lengkap dan terverifikasi.

2.3 Algoritma Naive Bayes

Naive Bayes merupakan metode klasifikasi yang berlandaskan teorema Bayes, dengan asumsi bahwa setiap atribut bersifat independen satu sama lain [10]. Secara matematis, teorema Bayes dinyatakan sebagai berikut :

$$P(C|X) = \frac{P(X|C)\times P(C)}{P(X)} \quad (1)$$

Keterangan : P (C | X) = probabilitas kelas C diberikan fitur X (posterior probability), P (X | C) = probabilitas fitur X diberikan kelas C (likelihood), P (C) = probabilitas awal kelas C (prior probability) , P (X) = probabilitas fitur X (evidence).

Langkah klasifikasinya meliputi: menghitung probabilitas prior setiap kelas, menghitung probabilitas kondisional setiap atribut, mengalikan keduanya, lalu memilih kelas dengan probabilitas tertinggi. Implementasi menggunakan *GaussianNB* dari *scikit-learn*, dipilih karena seluruh atribut telah direpresentasikan dalam bentuk numerik setelah proses *encoding* [10].

2.4 Pembagian Data dan Evaluasi Model

Sebelum tahap pemodelan, dataset terlebih dahulu dipisahkan menjadi data pelatihan (*training*) sebanyak 700 *record* dan data pengujian (testing) sebanyak 300 *record*, mengikuti rasio pembagian 70:30.

a. Confusion Matrix

Confusion Matrix digunakan untuk mengukur performa model dengan membandingkan nilai aktual dan prediksi [1], [2]. Metrik yang dihitung meliputi:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Presisi \frac{TP}{TP+FP} \quad (3)$$

$$Recall \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = \frac{2\times Precesion\times Recall}{Precesion+Recall} \quad (5)$$

Keterangan: TP= *True Positive,* TN= *True Negative*, FP= *False Positive*, FN= *False Negative*.

b. K-Fold Cross Validation

K-Fold Cross Validation mempartisi data menjadi k bagian (fold) dengan ukuran yang setara [15], [16]. Pelatihan model dilakukan sebanyak k iterasi, di mana setiap kali satu fold berperan sebagai data testing dan fold lainnya sebagai data training. Rata-rata akurasi dihitung dengan persamaan:

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

$$A = \frac{1}{k}\sum_{i=1}^{k} A_i \quad (6)$$

Keterangan: A¯= rata-rata akurasi, k= jumlah *fold*, Ai= akurasi pada *fold* ke-i.

Penelitian ini menggunakan k=10 untuk menghasilkan estimasi akurasi yang stabil dan tidak bergantung pada satu pembagian data tertentu [15].

3. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil seluruh tahapan penelitian, mulai dari *preprocessing* data, analisis deskriptif, perhitungan Naive Bayes manual, hingga evaluasi model menggunakan *Confusion Matrix* dan *K-Fold Cross Validation*.

3.1 Hasil *preprocessing* Data

Setelah tahap preprocessing selesai dilakukan, data siap digunakan berjumlah 1.010 record tanpa nilai kosong. Proses ekstraksi usia dari format teks berhasil mengonversi seluruh nilai menjadi numerik, dan proses encoding kategorik menghasilkan representasi numerik yang siap diproses oleh algoritma Naive Bayes. Contoh data setelah preprocessing disajikan pada Tabel 2.

Tabel 2. Contoh Data Setelah *Preprocessing*

No.	Usia Pria	Usia Wanita	Status Pria	Status Wanita	Pend. Pria	Pend. Wanita	Kelayakan
1.	42	24	0	0	4	4	Layak
2.	48	47	0	0	1	4	Layak
3.	25	22	1	0	2	3	Tidak Layak
4.	41	22	0	0	4	1	Layak
5.	54	30	0	0	1	2	Layak

Keterangan *encoding*: Status (0=Jejaka/Perawan, 1=Duda/Janda); Pendidikan (1=SD/MI, 2=SMP, 3=SMA, 4=S1).

3.2 Analisis Deskriptif Data

a. Distribusi Label Kelayakan

Hasil penentuan label kelayakan terhadap 1.010 data pernikahan menghasilkan distribusi sebagaimana disajikan pada Tabel 3.

Tabel 3. Distribusi Data Berdasarkan Label Kelayakan

Kelas	Jumlah Data	Persentase
Layak	851	84.3%
Tidak Layak	159	15.7%
Total	1.010	100%

Gambar 2. Distribusi Kelayakan Pernikahan (Total: 1.010 Data)

Gambar 2 menampilkan hasil sebanyak 851 data (84,3%) terklasifikasi sebagai Layak dan 159 data (15,7%) terklasifikasi sebagai Tidak Layak. Distribusi kelas menunjukkan adanya ketidakseimbangan (*class imbalance*) dengan rasio yang cukup signifikan antara kedua kelas. Kondisi ini perlu diperhatikan dalam interpretasi metrik evaluasi, di mana nilai akurasi saja tidak cukup untuk menggambarkan performa model secara menyeluruh [8].

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

b. Distribusi Usia Pasangan

Distribusi usia pria menunjukkan rentang 21–55 tahun, sedangkan usia wanita berada pada rentang yang serupa, tanpa pasangan berusia di bawah batas minimum undang-undang (19 tahun) dalam dataset ini. Sebagian besar kasus Tidak Layak ditemukan pada pasangan pria berusia di bawah 27 tahun dengan tingkat pendidikan rendah (SD/MI/SMP), atau pria berstatus Duda dengan usia di bawah 27 tahun. Pola distribusi usia ini mencerminkan karakteristik demografis masyarakat Kapuas Barat yang mayoritas menikah pada usia produktif, dengan dominasi pasangan pria berusia 25–35 tahun. Namun, keberadaan kasus Tidak Layak pada kelompok pria muda berpendidikan rendah menunjukkan bahwa faktor pendidikan turut berperan penting dalam kesiapan pasangan, tidak semata ditentukan usia memperkuat relevansi penggunaan kombinasi tiga atribut sekaligus dalam penentuan label kelayakan pada penelitian ini.

3.3 Perhitungan Naive Bayes Manual

Verifikasi cara kerja model dilakukan pada satu data uji (Tabel 4): Usia Pria 25 tahun, Usia Wanita 22 tahun, Status Pria Duda (1), Status Wanita Perawan (0), Pendidikan Pria SMP (2), Pendidikan Wanita SMA (3), dengan label aktual Tidak Layak.

Tabel 4. Data Uji Perhitungan Manual

Atribut	Nilai	Encoding
Usia Pria	25 tahun	25
Usia Wanita	22 tahun	22
Status Pria	Duda	1
Status Wanita	Perawan	0
Pendidikan Pria	SMP	2
Pendidikan Wanita	SMA	3
Label Aktual	Tidak Layak	-

a. Langkah Perhitungan

Perhitungan diawali dengan menentukan probabilitas prior masing-masing kelas, yaitu P (Layak) = 0,8426 dan P (Tidak Layak) = 0,1574. Selanjutnya dilakukan pengecekan kriteria kelayakan, di mana data uji memenuhi kondisi Tidak Layak karena status pria adalah Duda dengan usia 25 tahun yang berada di bawah ambang batas 27 tahun. Probabilitas kondisional kemudian dihitung menggunakan fungsi distribusi Gaussian sesuai persamaan (7) untuk setiap atribut numerik. Setelah seluruh probabilitas kondisional dikalikan dengan probabilitas prior masing-masing kelas, model mengklasifikasikan data uji sebagai tidak layak, sesuai dengan label aktualnya, sehingga prediksi dinyatakan benar (*True Negative*).

$$P(x_i \mid C) = \frac{1}{\sqrt{2\pi\sigma_C^2}} \exp\left(-\frac{(x_i-\mu_C)^2}{2\sigma_C^2}\right) \qquad (7)$$

3.4 Hasil *Confusion Matrix*

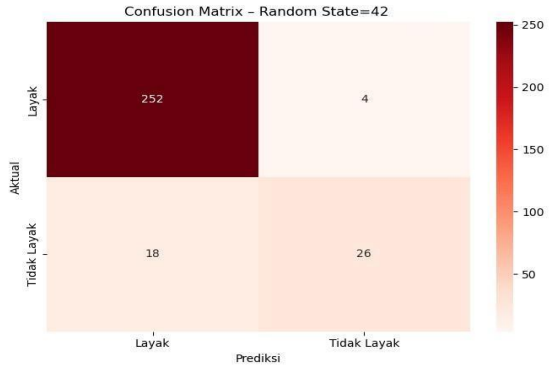
Evaluasi model pada 300 data testing dengan rasio pembagian 70:30 (*random_state*=42) menghasilkan *Confusion Matrix* sebagaimana disajikan pada Tabel 5.

Tabel 5. *Confusion Matrix* Naive Bayes

Aktual / Prediksi	Prediksi: Layak	Prediksi: Tidak Layak
Aktual: Layak	252 *(TP)*	4 (FN)
Aktual: Tidak Layak	18 (FP)	26 (TN)

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio



Gambar 3. *Confusion Matrix* Naive Bayes (Random State = 42)

Gambar 3 menampilkan hasil *Confusion Matrix* dari model Naive Bayes pada 300 data testing. Model berhasil memprediksi 252 data Layak dengan benar (*True Positive*) dan 26 data Tidak Layak dengan benar (*True Negative*). Terdapat 18 data yang salah diprediksi sebagai Layak padahal aktualnya Tidak Layak (*False Positive*), dan 4 data yang salah diprediksi sebagai Tidak Layak padahal aktualnya Layak (*False Negative*). Hasil ini menunjukkan model cukup baik dalam mengklasifikasikan kelas Layak, namun masih terdapat kesalahan pada kelas Tidak Layak yang perlu diperhatikan lebih lanjut.

a. Perhitungan Metrik Evaluasi

Berdasarkan nilai *Confusion Matrix* pada Tabel 5, metrik evaluasi dihitung menggunakan persamaan (8), (9), (10), dan (11):

$$Akurasi = \frac{252+26}{252+26+18+4} = \frac{278}{300} = 92{,}67\% \quad (8)$$

$$Presisi = \frac{252}{252+18} = \frac{252}{270} = 93{,}33\% \quad (9)$$

$$Recall = \frac{252}{252+4} = \frac{252}{256} = 98{,}44\% \quad (10)$$

$$F1-Score = \frac{2\times 0{,}9333\times 0{,}9844}{0{,}9333+0{,}9844} = \frac{1{,}8379}{1{,}9177} = 95{,}82\% \quad (11)$$

Tabel 6. Hasil Evaluasi Model Naive Bayes

Metrik	Akurasi	Presisi	Recall	F1-Score
Nilai	92.67%	93.33%	98.44%	95.82%

Nilai *recall* sebesar 98,44% menunjukkan bahwa model sangat baik dalam mengenali pasangan yang Layak. Nilai *F1-Score* 95,82% mengkonfirmasi keseimbangan antara presisi dan *recall* pada kelas mayoritas.

3.5 Hasil *K-Fold Cross Validation*

Untuk memvalidasi kestabilan dan generalisasi model, dilakukan *K-Fold Cross Validation* dengan k=10terhadap keseluruhan 1.010 data. Hasil akurasi setiap *fold* disajikan pada Tabel 7.

Tabel 7. Hasil *K-Fold Cross Validation* (k=10)

Fold Ke	1	2	3	4	5	6	7	8	9	10	Rata-rata
Akurasi	93.00%	96.00%	85.00%	90.00%	88.00%	91.00%	90.00%	90.00%	89.00%	89.00%	90.10%

Hasil *K-Fold Cross Validation* menunjukkan rata-rata akurasi sebesar 90.10% dengan standar deviasi 2.77%. Nilai standar deviasi yang relatif kecil mengindikasikan bahwa performa model konsisten di seluruh *fold* dan tidak bergantung pada satu pembagian data tertentu. Nilai *fold* tertinggi dicapai pada *Fold* 2 (96,00%) dan terendah pada *Fold* 3 (85,00%), dengan selisih 11% yang masih berada dalam batas toleransi kestabilan model.

a. Perbandingan Akurasi *Confusion Matrix* dan *Cross Validation*

Perbandingan antara akurasi *Confusion Matrix* (92,67%) dan rata-rata *K-Fold Cross Validation* (90,10%) menunjukkan selisih sebesar 2,57%. Selisih yang kecil ini mengkonfirmasi bahwa model tidak mengalami *overfitting*

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

terhadap data *training*. Distribusi akurasi antar *fold* yang divisualisasikan menunjukkan pola yang stabil, sehingga model dinilai layak untuk diterapkan pada data pernikahan baru di luar periode penelitian.

3.6 Analisis Performa Model Naive Bayes

Algoritma Naive Bayes yang diimplementasikan menggunakan *GaussianNB* pada data pernikahan KUA Kecamatan Kapuas Barat berhasil mencapai akurasi sebesar 92,67% dengan pembagian data 70:30. Hasil ini menunjukkan bahwa algoritma Naive Bayes efektif diterapkan pada data administrasi pernikahan yang memiliki kombinasi atribut numerik dan kategorik. Capaian ini melampaui penelitian rujukan oleh Hadid dan Eliyani yang menghasilkan akurasi 84% pada KUA Palmerah dengan 300 data [6]. Peningkatan performa ini didukung oleh jumlah data yang lebih besar (1.010 *record*) serta kriteria kelayakan yang lebih komprehensif yang mempertimbangkan kombinasi tiga faktor sekaligus, yaitu usia, pendidikan, dan status pernikahan.

Nilai *Recall* 98,44% menunjukkan model sangat sensitif mendeteksi pasangan Layak, sehingga kecil kemungkinan pasangan Layak terlewat (FN) penting karena kesalahan menolak pasangan layak berdampak langsung pada pelayanan publik. *F1-Score* 95,82% mengonfirmasi keseimbangan presisi-*recall* pada kelas mayoritas [17]. Hasil ini juga mengindikasikan bahwa penentuan label kelayakan berbasis kombinasi tiga faktor berkontribusi signifikan terhadap kualitas model, label yang jelas dan konsisten memungkinkan algoritma mempelajari pola distribusi data lebih baik, menghasilkan probabilitas posterior lebih akurat. Penggunaan *GaussianNB* terbukti sesuai untuk data atribut numerik hasil *encoding* karena mengestimasi parameter distribusi *Gaussian* secara otomatis tanpa konfigurasi kompleks kecocokan karakteristik data dengan jenis model menjadi faktor penentu utama keberhasilan implementasi.

a. Analisis Kesalahan Prediksi

Dari 300 data testing, 22 data diprediksi keliru (18 FP, 4 FN). Kasus FP—model memprediksi Layak padahal Tidak Layak—umumnya terjadi pada pasangan yang hanya memenuhi satu dari tiga kriteria ketidaklayakan, misalnya pria Duda berusia tepat di ambang 27 tahun dengan pendidikan SMA, menempatkan data pada area perbatasan (*borderline*) antar kelas sehingga model sulit membedakannya [18].

Kesalahan ini juga dapat disebabkan asumsi independensi Naive Bayes yang tidak selalu terpenuhi: atribut usia dan pendidikan berpotensi berkorelasi, di mana pasangan lebih muda cenderung berpendidikan lebih rendah [2]. Pelanggaran asumsi independensi ini dikenal sebagai salah satu keterbatasan utama algoritma Naive Bayes pada data berkorelasi [19]. Meskipun demikian, tingkat kesalahan 7,33% masih dalam batas dapat diterima untuk sistem pendukung keputusan administrasi. Hasil prediksi bukan keputusan final, melainkan rekomendasi awal yang tetap dikonfirmasi petugas KUA berdasarkan dokumen fisik dan pertimbangan hukum, sehingga dampak negatif kesalahan dapat diminimalisir melalui verifikasi berlapis. Ke depannya, mekanisme *explainability* prediksi dapat membantu petugas memahami alasan di balik rekomendasi model.

b. Analisis *Class Imbalance*

Dataset mengandung ketidakseimbangan kelas (84,3% : 15,7%), sehingga akurasi saja tidak cukup menggambarkan performa model tantangan umum yang dapat memengaruhi reliabilitas evaluasi [20], sejalan dengan Apriliyani dan Salim bahwa presisi, *recall*, dan *F1-Score* harus digunakan bersama untuk evaluasi lebih akurat pada data tidak seimbang [8].

Presisi kelas Tidak Layak yang lebih rendah mengindikasikan model masih perlu ditingkatkan dalam mendeteksi kasus minoritas. Izzi, Setyanto, dan Hartanto membuktikan SMOTE secara signifikan meningkatkan kemampuan Naive Bayes mengenali kelas minoritas [7]; Faulina, Nisa, dan Warsono menunjukkan kombinasi SMOTE dan Tomek Links efektif meningkatkan sensitivitas model pada data tidak seimbang keduanya direkomendasikan untuk penelitian lanjutan [13],[14]. Meskipun terdapat *imbalance*, model tetap menunjukkan performa baik berkat metrik evaluasi yang beragam; penggunaan *F1-Score* di samping akurasi memungkinkan penilaian lebih objektif. Kondisi *imbalance* ini sebenarnya mencerminkan kondisi nyata di lapangan, di mana pasangan layak memang jauh lebih banyak dibanding yang tidak layak, sehingga model tetap relevan dan representatif untuk konteks administrasi pernikahan sesungguhnya.

c. Analisis Stabilitas Model

K-Fold (k=10) menghasilkan rata-rata akurasi 90,10% dengan standar deviasi 2,77%. Konsistensi antara akurasi *Confusion Matrix* (92,67%) dan rata-rata *cross validation* (90,10%), selisih hanya 2,57%, mengonfirmasi model tidak *overfitting* [15]. Lumumba et al. menekankan pentingnya *K-Fold* untuk estimasi akurasi yang stabil dan tidak bergantung pada satu pembagian data, dengan tiap fold mempertahankan proporsi kelas sesuai distribusi data asli [16]. Variasi akurasi antar *fold* (85,00%–96,00%, rentang 11%) menunjukkan model cukup konsisten pada berbagai kondisi pembagian data, dan rentang ini masih dapat diterima mengingat *class imbalance* yang dapat memengaruhi hasil klasifikasi pada *fold* tertentu [17]. Perbandingan performa penelitian ini dengan penelitian-penelitian sebelumnya yang memiliki domain sejenis disajikan pada Tabel 8.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

Tabel 8. Perbandingan dengan Penelitian Terdahulu

No.	Penelitian	Metode	Domain	Akurasi
1.	Hadid & Eliyani [6]	Naive Bayes	Administrasi pernikahan KUA	84.00%
2.	Fauziah dkk. [10]	Naive Bayes	Kesiapan menikah dewasa muda	85.00%
3.	Nurjannah dkk.[18]	*Weighted* Naive Bayes	Klasifikasi kelayakan administrasi publik	89.00%
4.	Penelitian Ini	Naive Bayes (Python)	Data administrasi pernikahan KUA	92.67%

Perbandingan dengan penelitian terdahulu (Tabel 8) menunjukkan penelitian ini menghasilkan akurasi tertinggi (92,67%) dibanding Hadid & Eliyani (84,00%, administrasi pernikahan KUA), Fauziah dkk. (85,00%, kesiapan menikah dewasa muda), dan Nurjannah dkk (89,00%, Weighted Naive Bayes pada kelayakan administrasi publik) [18]. Data pembanding dipilih dari domain setara agar perbandingan lebih representatif. Keunggulan ini didukung tiga faktor utama: jumlah data lebih besar (1.010 record), kriteria kesiapan administrasi lebih komprehensif (tiga faktor sekaligus), serta implementasi *library* scikit-learn yang teroptimasi. Temuan ini memperkuat posisi Naive Bayes sebagai algoritma andal untuk klasifikasi data administrasi pernikahan di instansi pemerintahan, sejalan dengan penelitian sebelumnya mengenai efektivitas Naive Bayes dalam pengalokasian layanan publik berbasis data di pemerintah daerah [19].

d. Implikasi Praktis

Hasil penelitian memiliki implikasi praktis bagi pengelolaan administrasi pernikahan di KUA Kapuas Barat maupun KUA lain. Model berpotensi diintegrasikan ke sistem informasi administrasi pernikahan sebagai modul pendukung keputusan yang memberikan rekomendasi kesiapan administrasi pasangan secara otomatis, mengurangi waktu verifikasi awal secara signifikan terutama pada volume pendaftaran tinggi [4].

Hal ini sejalan dengan prinsip pelayanan publik yang transparan dan akuntabel, di mana keputusan dapat ditelusuri berdasarkan data dan logika klasifikasi yang terukur. Dengan akurasi 92,67% dan stabilitas terkonfirmasi *K-Fold*, model layak dipertimbangkan sebagai komponen sistem informasi administrasi pernikahan berbasis *data mining* di instansi pemerintah [13]. Secara operasional, petugas menginputkan data calon pengantin, model menghasilkan rekomendasi beserta tingkat kepercayaan sebagai bahan pertimbangan awal sebelum verifikasi dokumen fisik, sehingga mempercepat dan menstandardisasi tahap pra-verifikasi tanpa menggantikan peran petugas [14].

Beberapa keterbatasan perlu diperhatikan: asumsi independensi antar atribut tidak selalu terpenuhi karena usia dan pendidikan berpotensi berkorelasi [21]. Untuk pengembangan ke depan, disarankan menerapkan SMOTE guna mengatasi *class imbalance* [22], membandingkan *Naive Bayes* dengan algoritma lain seperti *Decision Tree atau Random Forest* yang terbukti efektif pada domain administrasi publik [23], serta menambahkan atribut seperti pekerjaan dan wilayah asal pasangan untuk memperkaya analisis dan meningkatkan akurasi [24]. Model ini diharapkan menjadi landasan pengembangan sistem administrasi pernikahan terpadu di tingkat kabupaten/kota, sebagai langkah konkret menuju transformasi digital pelayanan publik yang lebih efisien dan berbasis bukti [25]. Luaran penelitian ini merupakan model klasifikasi berbasis data, bukan prototipe perangkat atau sistem informasi yang siap diimplementasikan secara langsung.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Naive Bayes menggunakan *library scikit-learn* berbasis Python untuk mengklasifikasikan kesiapan administrasi pernikahan pada 1.010 data pernikahan KUA Kecamatan Kapuas Barat periode 2021–2025. Kriteria kesiapan administrasi yang ditetapkan berdasarkan kombinasi tiga faktor sekaligus, yaitu usia, pendidikan, dan status pernikahan sesuai Undang-Undang Nomor 16 Tahun 2019, terbukti menghasilkan label klasifikasi yang representatif dengan distribusi 851 data Layak (84,3%) dan 159 data Tidak Layak (15,7%). Model yang dibangun mencapai akurasi sebesar 92,67% pada data testing dengan pembagian 70:30, disertai nilai presisi 93,33%, *recall* 98,44%, dan *F1-Score* 95,82%. Validasi menggunakan *K-Fold Cross Validation* (k=10) menghasilkan rata-rata akurasi 90,10% dengan standar deviasi 2,77%, yang mengkonfirmasi bahwa model stabil dan tidak mengalami *overfitting*. Performa ini melampaui penelitian sebelumnya pada domain administrasi pernikahan oleh Hadid dan Eliyani yang mencapai akurasi 84% dengan 300 data dan kriteria tunggal berbasis usia, serta memperkuat posisi Naive Bayes sebagai algoritma yang andal untuk klasifikasi kesiapan administrasi pernikahan di instansi pemerintah .Keterbatasan penelitian ini terletak pada adanya ketidakseimbangan distribusi kelas *(class imbalance*) dengan rasio 84,3% berbanding 15,7%, yang berpotensi memengaruhi kemampuan model dalam mendeteksi kelas minoritas secara optimal. Ke depannya, penelitian ini diharapkan dapat menjadi landasan dalam mengembangkan sistem pendukung keputusan administrasi pernikahan yang lebih lengkap dan terintegrasi.

UCAPAN TERIMAKASIH

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

Penulis mengucapkan terima kasih kepada Kepala dan seluruh staf KUA Kecamatan Kapuas Barat, Kabupaten Kapuas, Kalimantan Tengah, yang telah memberikan izin akses dan menyediakan data pernikahan periode 2021–2025 sehingga penelitian ini dapat terlaksana. Penulis juga menyampaikan terima kasih kepada Program Studi S1 Informatika, Fakultas Teknik, Universitas Muhammadiyah Banjarmasin atas dukungan akademik selama proses penelitian berlangsung.

REFERENCES

[1] A. Nurhuda, F. Firmansyah, and M. S. H. Napis, "Analisis Kualitas Pelayanan Publik Di Bidang Pencatatan Nikah Pada Kantor Urusan Agama," *Journal of Governance and Public Administration*, vol. 1, no. 1, pp. 76–89, Dec. 2023, doi: 10.59407/jogapa.v1i1.330.

[2] Syarifah Lisa Andriati, Mutiara Sari, and Windha Wulandari, "Implementasi Perubahan Batas Usia Perkawinan Menurut UU No. 16 Tahun 2019 tentang Perubahan Atas UU No. 1 Tahun 1974 tentang Perkawinan," *Binamulia Hukum*, vol. 11, no. 1, pp. 59–68, Mar. 2023, doi: 10.37893/jbh.v11i1.306.

[3] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[4] N. H. Nufus and S. Sriani, "Penerapan Metode K-Means Untuk Pengelompokkan Data Pelaporan Di Kantor Urusan Agama," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 2085–2093, Dec. 2024, doi: 10.47065/bits.v6i3.6316.

[5] A. M. Irfan and K. Kusrini, "Analisis Perbandingan Metode *Naïve Bayes* dan K-NN dalam Penentuan Lokasi Layanan Administrasi BPJS Kesehatan di Provinsi Maluku.," *Journal Computer Science and Information Systems : J-Cosys*, vol. 4, no. 2, Aug. 2024, doi: 10.53514/jco.v4i2.535.

[6] M. Hadid and E. Eliyani, "Penerapan Klasifikasi Kelayakan Pernikahan Pada KUA Palmerah Dengan Menggunakan Algoritma Naïve Bayes," *JRKT (Jurnal Rekayasa Komputasi Terapan)*, vol. 1, no. 02, Jun. 2021, doi: 10.30998/jrkt.v1i02.4089.

[7] Hizbul Izzi, Arief Setyanto, and Anggit Dwi Hartanto, "Optimalisasi Akurasi Algoritma Naïve Bayes Dengan Metode Syntetic Minority Oversampling Technique (Smote) Pada Data Numerik," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 1, pp. 217–227, Jan. 2025, doi: 10.29408/jit.v8i1.28340.

[8] Ericha Apriliyani and Y. Salim, "Analisis performa metode klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset," *Indonesian Journal of Data and Science*, vol. 3, no. 2, pp. 47–54, Jul. 2022, doi: 10.56705/ijodas.v3i2.45.

[9] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Informatik : Jurnal Ilmu Komputer*, vol. 18, no. 3, p. 239, Dec. 2022, doi: 10.52958/iftk.v18i3.4694.

[10] R. Fauziah, H. S. Tambunan, and S. Susiani, "Data Classification of Marriage Readiness in Young Adults Using the Naïve Bayes Algorithm," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 4, pp. 285–294, Dec. 2022, doi: https://doi.org/10.55123/jomlai.v1i4.1665.

[11] Y. Irfayanti, "Penerapan Metode *Naive Bayes* untuk Klasifikasi Status Pegawai pada Perusahaan Swasta," *Syntax Literate: Jurnal Ilmiah Indonesia*, vol. 7, no. 10, pp. 17935–17941, Oct. 2022, doi: https://doi.org/10.36418/syntax-literate.v7i10.13230.

[12] W. Wang, L. Yan, F. Liu, and Y. Li, "Improving Gaussian *Naive Bayes* classification on imbalanced data through coordinate-based minority feature mining," *PeerJ Comput. Sci.*, vol. 11, p. e3003, Jul. 2025, doi: 10.7717/peerj-cs.3003.

[13] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.

[14] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J. Big Data*, vol. 12, no. 1, p. 268, Dec. 2025, doi: 10.1186/s40537-025-01313-4.

[15] J. Kaliappan, A. R. Bagepalli, S. Almal, R. Mishra, Y.-C. Hu, and K. Srinivasan, "Impact of Cross-Validation on Machine Learning Models for Early Detection of Intrauterine Fetal Demise," *Diagnostics*, vol. 13, no. 10, p. 1692, May 2023, doi: 10.3390/diagnostics13101692.

[16] V. Lumumba, D. Kiprotich, M. Mpaine, N. Makena, and M. Kavita, "Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models," *American Journal of Theoretical and Applied Statistics*, vol. 13, no. 5, pp. 127–137, Oct. 2024, doi: 10.11648/j.ajtas.20241305.13.

[17] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," *Frontiers in Nanotechnology*, vol. 4, Aug. 2022, doi: 10.3389/fnano.2022.972421.

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 5 No 1.1 Spesial Issue Agustus 2026 | Page 17-27
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solutio

[18] N. Nurjannah, H. Cipta, and R. Aprilia, "Eligibility Classification of Aid Recipients Hope Family Program in Cinta Rakyat Village Using the Method *Weighted* Naive Bayes with Laplace Smoothing," *Jurnal Matematika, Statistika dan Komputasi*, vol. 20, no. 2, pp. 440–454, Dec. 2023, doi: 10.20956/j.v20i2.32069.

[19] M. I. Adha, F. D. Marleny, and A. A. Ningrum, "Optimalisasi Pengalokasian Bantuan Sosial di Kota Banjarmasin Menggunakan Pendekatan Naive Bayes*," Digital Transformation Technology*, vol. 5, no. 1, pp. 240–249, Jul. 2025, doi: 10.47709/digitech.v5i1.6001.

[20] F. P. Muhammad, Maimunah, and S. Nugroho, "Optimasi Penanganan Ketidakseimbangan Data pada Klasifikasi Pengaduan Masyarakat Menggunakan Metode Naïve Bayes," *Jurnal Tekno Kompak*, vol. 20, no. 2, pp. 503–514, May 2026, doi: https://doi.org/10.33365/jtk.v20i2.939.

[21] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Constrained Naïve Bayes with application to unbalanced data classification," *Cent. Eur. J. Oper. Res.*, vol. 30, no. 4, pp. 1403–1425, Dec. 2022, doi: 10.1007/s10100-021-00782-1.

[22] N. Faulina, K. Nisa, and W. Warsono, "Enhancing Tuberculosis Diagnosis: Effective *Naive Bayes* Classification using SMOTE and Tomek Links for Imbalanced Data," *InPrime: Indonesian Journal of Pure and Applied Mathematics*, vol. 6, no. 2, pp. 98–111, Oct. 2024, doi: 10.15408/inprime.v6i2.41463.

[23] A. Aqli, F. D. Marleny, and W. Windarsyah, "Klasifikasi Sekolah Potensial untuk Promosi Kampus UMBJM Menggunakan Random Forest," *Jurnal Media Informatika*, vol. 6, no. 3, pp. 1913–1919, May 2025, doi: https://doi.org/10.55338/jumin.v6i3.6122.

[24] C. A. Ferrer and E. Aragón, "Note on 'A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance,'" *Inf. Sci. (N. Y).*, vol. 630, pp. 322–324, Jun. 2023, doi: 10.1016/j.ins.2022.10.005.

[25] M. S. J. Sangaji and J. Irianto, "Transformasi Inovasi Pelayanan Publik menuju Pemerintahan Digital," *Jejaring Administrasi Publik*, vol. 17, no. 1, pp. 54–70, Jun. 2025, doi: 10.20473/jap.v17i1.72708

Submitted: **29/05/2026**; Accepted: **20/06/2026**; Published:**08/08/2026**