

Hybrid sampling untuk meningkatkan akurasi deteksi kanker serviks pada data tidak seimbang: kajian komparatif

Slamet Widodo ^{a,1,*}, Samudi ^{a,2}, Herlambang Brawijaya ^{a,3}

^a Universitas Bina Sarana Informatika, Jln. Kramat Raya No. 98, Jakarta, Indonesia, 10450

¹ slamet.smd@bsi.ac.id*; ² samudi.smx@bsi.ac.id; ³ herlambang.hba@bsi.ac.id

* Penulis Koresponden

INFO ARTIKEL

Histori Artikel

Pengajuan 2025-09-02

Diperbaiki 2025-09-25

Diterima 2025-12-08

Kata Kunci

Borderline-SMOTE,

Deteksi Dini,

Kanker Serviks,

Ketidakeimbangan Kelas,

Resampling

ABSTRAK

Tingginya angka kematian akibat kanker serviks pada perempuan mendorong pemanfaatan deteksi dini berbasis *machine learning*. Namun, implementasinya terkendala masalah ketidakseimbangan kelas, seperti pada UCI *Cervical Cancer Behavior Risk dataset* yang hanya memiliki 5,8–7,3% kasus positif. Penelitian ini mengusulkan evaluasi teknik resampling termasuk *SMOTE*, *ADASYN*, *Random Undersampling*, dan *Borderline-SMOTE* yang dikombinasikan dengan algoritma *RF*, *XGBoost*, *LR*, *GNB*, dan *k-NN*. Dengan menggunakan *Stratified K-Fold Cross Validation* untuk menjaga distribusi kelas asli pada setiap fold dan menjamin proses resampling hanya dilaksanakan pada data pelatihan di setiap iterasi, hasil penelitian menunjukkan bahwa *Borderline-SMOTE* berhasil meningkatkan kinerja model secara signifikan, terutama pada *Random Forest* yang mencapai *Recall* 0,87 dan *AUC-ROC* 0,94. Temuan ini diharapkan dapat menjadi landasan bagi penelitian lanjutan yang berfokus pada optimasi metode sampling adaptif dan integrasi model *ensemble* untuk sistem deteksi dini kanker serviks.

ABSTRACT

Keyword

Borderline-SMOTE,

Cervical Cancer,

Class Imbalance,

Early Detection,

Resampling

Cervical Cancer has a high mortality rate among women, driving the adoption of early detection systems based on machine learning. However, their implementation is hindered by class imbalance issues, as seen in the UCI Cervical Cancer Behavior Risk Dataset, where positive cases constitute only 5.8–7.3% of the data. This study proposes an evaluation of resampling techniques—including SMOTE, ADASYN, Random Undersampling, and Borderline-SMOTE—combined with classification algorithms such as RF, XGBoost, LR, GNB, and k-NN. Using Stratified K-Fold Cross Validation to preserve the original class distribution in each fold and ensuring resampling is applied only to the training data in each iteration, the results demonstrate that Borderline-SMOTE significantly improved model performance. Specifically, the Random Forest model achieved a Recall of 0.87 and an AUC-ROC of 0.94. These findings are expected to provide a foundation for future research focused on optimizing adaptive sampling methods and integrating ensemble models for more reliable Cervical Cancer early detection system.

Ini adalah artikel akses terbuka di bawah lisensi [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/).

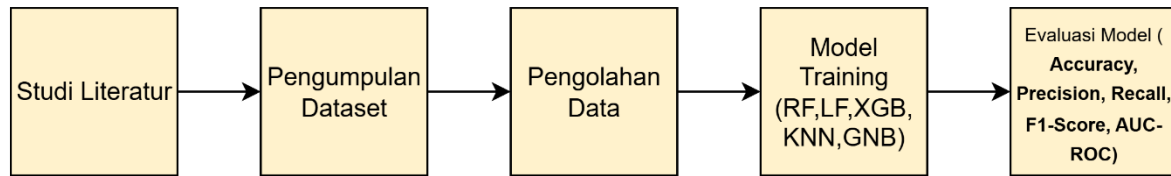


1. Pendahuluan

Kanker serviks menempati peringkat keempat sebagai kanker paling umum pada perempuan di dunia, dengan perkiraan 660.000 kasus baru dan 350.000 kematian pada tahun 2022. Deteksi dini berbasis *machine learning* (ML) berpotensi meningkatkan akurasi diagnosis, namun sering terhambat oleh masalah ketidakseimbangan kelas pada data latih [1]. Pada *dataset* UCI kanker serviks (*Behavior Risk*), proporsi variabel target biopsi sangat tidak seimbang, di mana kasus positif (terdiagnosis kanker) hanya berkisar antara 5,8% hingga 7,3% dari total data, sementara kasus negatif mendominasi dengan proporsi lebih dari 92% [2]–[4]. Ketidakseimbangan ini menyebabkan model ML cenderung bias terhadap kelas mayoritas. Akibat ketidakseimbangan kelas, model ML cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas [8]. Dampaknya, akurasi saja tidak cukup menggambarkan kinerja model, sebab bisa jadi model gagal mendeteksi kasus kanker yang justru lebih krusial untuk diidentifikasi [9], [10]. Beberapa peneliti yang menggunakan *dataset* ini telah membuktikan dampak negatif dari masalah ini. Salah satunya, studi [11] melaporkan bahwa model Random Forest tanpa penanganan *imbalance* mampu mencapai akurasi 94%, namun hanya berhasil mendeteksi 28% kasus kanker (nilai *Recall* yang sangat rendah). Temuan serupa terjadi pada model *Support Vector Machine* (SVM) yang kesulitan mendeteksi kelas minoritas akibat dominasi data negatif, sehingga hanya meraih F1-score 0,25 [8]. Hal ini mempertegas bahwa teknik konvensional seperti *Stratified K-Fold Cross Validation* (SKCV), meski sudah menjaga distribusi kelas di setiap *fold*, belum cukup mengatasi bias selama pelatihan model [12]. Untuk mengatasi masalah tersebut, sejumlah teknik resampling banyak dibahas dalam berbagai literatur. Salah satunya adalah *Synthetic Minority Over-sampling Technique* (SMOTE), yang menghasilkan data sintesis untuk kelas minoritas dengan melakukan interpolasi antara *instance* yang ada dan tetangga terdekatnya [12]. Teknik lain seperti *Adaptive Synthetic Sampling* (ADASYN) juga pernah diterapkan. ADASYN bersifat lebih adaptif dengan berfokus pada contoh minoritas yang paling sulit diklasifikasikan. Misalnya, yang berada di tengah *cluster* data mayoritas sehingga model mendapat bobot latih yang lebih proporsional [13], [14]. Efektivitas teknik-teknik ini telah terlihat dalam penelitian terkait kanker serviks. Salah satu laporan menyatakan bahwa nilai *Recall* untuk kasus kanker naik dari 30% menjadi 72% setelah penerapan SMOTE pada *dataset* yang sama [15]. Namun, hingga saat ini belum ada kesepakatan mengenai metode resampling mana yang paling optimal, terlebih ketika dikombinasikan dengan algoritma ML seperti XGBoost atau *Logistic Regression*, khususnya pada data yang berdimensi tinggi dan mengandung *noise* [16]. Berdasarkan uraian di atas, penelitian ini bertujuan melakukan evaluasi komprehensif dengan menguji berbagai teknik resampling seperti SMOTE, ADASYN, *Random Undersampling*, dan *Borderline-SMOTE* yang dikombinasikan dengan *Stratified K-Fold Cross Validation* (SKCV) dan beberapa algoritma ML populer (*RF*, *XGBoost*, *LR*, *GNB* dan *k-NN*) pada *dataset* UCI kanker serviks tersebut. Fokusnya adalah menemukan konfigurasi terbaik yang mampu memaksimalkan deteksi kasus positif (tinggi *Recall*) tanpa mengorbankan stabilitas model secara keseluruhan, sehingga dapat menjadi landasan untuk pengembangan sistem diagnosis yang lebih baik.

2. Metode

Penelitian ini bertujuan menguji sejauh mana teknik resampling dan model *machine learning* dapat memprediksi kanker serviks berdasarkan faktor risikonya. Secara garis besar, alur penelitian diawali dengan tinjauan literatur dan pengumpulan *dataset*, lalu dilanjutkan dengan preprocessing data termasuk penanganan missing value dan standardisasi fitur. Selanjutnya, beberapa metode resampling akan diterapkan dalam skema validasi untuk mengatasi ketidakseimbangan kelas yang umum ditemui dalam data kesehatan. Berbagai model klasifikasi kemudian akan dilatih dan diuji menggunakan validasi silang. Tahap terakhir adalah menganalisis hasil evaluasi secara mendalam guna menemukan kombinasi teknik resampling dan model yang paling efektif. Seluruh tahapan ini dapat dilihat secara visual pada Gambar 1



Gambar 1. Diagram alir metode penelitian

Rancangan penelitian bersifat komparatif-eksperimental, bertujuan membandingkan langsung kinerja beberapa model dan teknik resampling pada *dataset* yang identik. Untuk menjamin konsistensi hasil dan mencegah data leakage, seluruh proses evaluasi diintegrasikan ke dalam sebuah *pipeline* algoritma yang terstruktur, sebuah praktik terpenting dalam *machine learning* [17]. Berikut penjelasan masing-masing proses dari tahapan tersebut.

a) Studi Literatur

Pada tahap awal, dimulai dengan melakukan studi literatur mendalam. Penelusuran berbagai jurnal dan sumber terkait untuk memahami masalah ketidakseimbangan kelas, berbagai teknik *resampling*, serta algoritma *machine learning* yang biasa diterapkan dalam deteksi kanker serviks dan domain kesehatan serupa. Kajian ini menjadi fondasi bagi pemilihan metode yang tepat, mulai dari algoritma klasifikasi (seperti *RF*, *LR*, *XGBoost*, *KNN*, dan *Naive Bayes*) hingga teknik penanganan *imbalance*, termasuk *SMOTE*, *ADASYN*, *Random Undersampling*, dan *Borderline-SMOTE*;

b) Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari UCI kanker serviks *Behavior Risk dataset*. Kumpulan data ini mencakup berbagai variabel faktor risiko kanker serviks, misalnya usia, perilaku seksual, riwayat penggunaan kontrasepsi, dan hasil pemeriksaan medis. Variabel targetnya adalah hasil biopsi, yang menunjukkan apakah seorang pasien positif atau negatif menderita kanker;

c) Pengolahan Data

Data yang diambil dari repositori tidak langsung bisa dipakai, melainkan harus melalui serangkaian tahap pengolahan terlebih dahulu, seperti:

Preprocessing data merupakan langkah krusial dalam mempersiapkan data untuk pemodelan, terlebih pada *dataset* faktor risiko kanker serviks yang sering kali mengandung banyak *missing value* khususnya pada variabel medis seperti riwayat penyakit menular seksual (STDs). Untuk menangani nilai yang hilang, kami melakukan imputasi. Variabel numerik, contohnya usia atau jumlah kehamilan, biasanya diisi dengan nilai median agar tidak terlalu terpengaruh *outlier*. Sementara untuk variabel kategorikal biner seperti riwayat STD, nilai yang hilang diganti dengan modus. Pendekatan seperti ini telah diterapkan dalam penelitian sebelumnya [18] dan terbukti mampu meningkatkan stabilitas model, khususnya pada algoritma yang bergantung pada perhitungan jarak [17]. Selain itu, normalisasi menggunakan *Min-Max* juga diterapkan untuk menyelaraskan skala antar variabel dan mencegah bias, seperti yang telah dilakukan dalam penelitian terkait [2]. Sebagian besar variabel dalam *dataset* ini sebenarnya sudah berbentuk biner (0 dan 1). Namun, jika terdapat variabel kategorikal dengan lebih dari dua kategori, kami menggunakan encoding seperti *One-Hot Encoding* agar data tersebut dapat diproses oleh model *machine learning*. Teknik ini merupakan langkah standar dalam alur kerja pemodelan *machine learning* [17]. Data yang digunakan dalam penelitian ini memiliki masalah ketidakseimbangan kelas yang signifikan pada variabel target. Dari total 858 sampel, hanya 55 sampel (6,4%) yang merupakan kasus positif kanker serviks berdasarkan hasil biopsi, sedangkan 803 sampel (93,6%) merupakan kasus negatif. Rasio ketidakseimbangan yang mencapai 1:14,6 menunjukkan bahwa untuk

setiap satu kasus kanker, terdapat hampir lima belas kasus non-kanker. Kondisi tidak seimbang ini menyebabkan model *machine learning* seperti LR atau RF cenderung bias terhadap kelas mayoritas. Akibatnya, meskipun model dapat mencapai akurasi tinggi, model justru gagal mendeteksi kasus kanker yang menjadi fokus utama. Dalam kondisi terburuk, model tanpa penanganan *imbalance* bahkan dapat menghasilkan *Recall* 0% untuk kelas minoritas, yang berarti semua kasus kanker salah diprediksi sebagai sehat [18]. Untuk mengatasi masalah ini, teknik *resampling* seperti *SMOTE* dapat diterapkan secara khusus pada variabel biopsi. *SMOTE* menghasilkan sampel sintetis baru untuk kelas minoritas (55 kasus positif) melalui interpolasi linier antara suatu titik data minoritas dan tetangga terdekatnya. Dengan parameter default ($k=5$), teknik ini dapat menyeimbangkan jumlah sampel positif hingga setara dengan kelas mayoritas, misalnya menjadi sekitar 800 sampel sintetis. Penelitian sebelumnya [19] menunjukkan bahwa penerapan *SMOTE* pada *dataset* ini berhasil meningkatkan *Recall* untuk kelas kanker serviks dari 28% menjadi 75%, tanpa mengorbankan akurasi secara signifikan. Di sisi lain, teknik *Random Undersampling* (RUS) mengatasi ketidakseimbangan dengan mengurangi sampel kelas mayoritas secara acak. Pada *dataset* ini, RUS akan menurunkan jumlah sampel negatif dari 803 menjadi 55 sampel yang dipilih secara acak. Namun, pendekatan ini berisiko menghilangkan informasi penting dan pola-pola kunci dari kelas mayoritas, yang justru dapat menurunkan kemampuan generalisasi model. Oleh karena itu, pendekatan berbasis sintetis seperti *ADASYN* dan *Borderline-SMOTE* umumnya lebih direkomendasikan untuk *dataset* kanker serviks. *ADASYN* tidak hanya menghasilkan sampel sintetis, tetapi juga berfokus pada kasus-kasus minoritas yang paling sulit diklasifikasikan, seperti pasien dengan faktor risiko tinggi yang sebelumnya mengalami misklasifikasi.

d) Model Training

- **Random Forest**

Pada model *Random Forest* (RF), SK-Fold diterapkan dengan melatih pohon keputusan pada data latih dari tiap lipatan (*fold*), lalu hasil prediksi diuji pada lipatan validasi. RF membangun sejumlah pohon dengan sampel data acak (*bootstrapping*) dan melakukan prediksi dengan mayoritas suara (*majority voting*) [20]. Rumus prediksi RF dapat dituliskan dengan persamaan 1:

$$y \text{ mode}\{T1(x), T2(x), \dots, TB(x)\} \quad (1)$$

dengan $ht(x)$ adalah pohon ke- t dan T jumlah total pohon.

- **XGBoost (Extreme Gradient boosting)**

Pada *Extreme Gradient Boosting* (XGBoost), SK-Fold memastikan bahwa model dievaluasi secara stabil pada tiap lipatan, sementara *resampling* digunakan untuk mengurangi bias akibat ketidakseimbangan kelas. XGBoost bekerja dengan membangun model aditif dari pohon keputusan menggunakan optimasi gradient boosting [21]. Rumus dasarnya ditulis pada persamaan 2.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (2)$$

dengan F adalah ruang fungsi pohon, dan model diperbarui secara bertahap berdasarkan gradien dari fungsi kerugian.

- **Logistic Regression (LR)**

Model *Logistic Regression* (LR) menggunakan SK-Fold untuk membagi data latih dan uji, sementara *resampling* menjaga agar distribusi kelas tetap proporsional. LR memodelkan probabilitas keanggotaan kelas dengan fungsi logistik [22], ditunjukkan persamaan 3.

$$P = (y = 1|x) = \frac{1}{1+e^{(\beta_0+\beta_1+\beta_n x)}} \quad (3)$$

dengan β adalah parameter yang diestimasi dengan *maximum likelihood estimation (MLE)*

• **k-Nearest Neighbors (k-NN)**

k-Nearest Neighbor (k-NN) dievaluasi menggunakan SK-Fold untuk menghitung akurasi prediksi berbasis *distance metric* pada tiap lipatan. Resampling membantu memastikan setiap kelas terwakili dengan proporsional pada tetangga terdekat [23]. Rumus jarak Euclidean yang umum digunakan ditunjukkan persamaan 4.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

• **Naïve Bayes (GaussianNB)**

Pada *Gaussian Naïve Bayes (GNB)*, *SK-Fold* digunakan untuk mengevaluasi performa probabilistik klasifikasi pada tiap lipatan, sedangkan resampling membantu menyeimbangkan jumlah sampel antar kelas. GNB bekerja berdasarkan *bayes theorem* dengan asumsi distribusi normal pada fitur [24], rumus teorema ditunjukkan persamaan 5:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (5)$$

e) **Evaluasi Model**

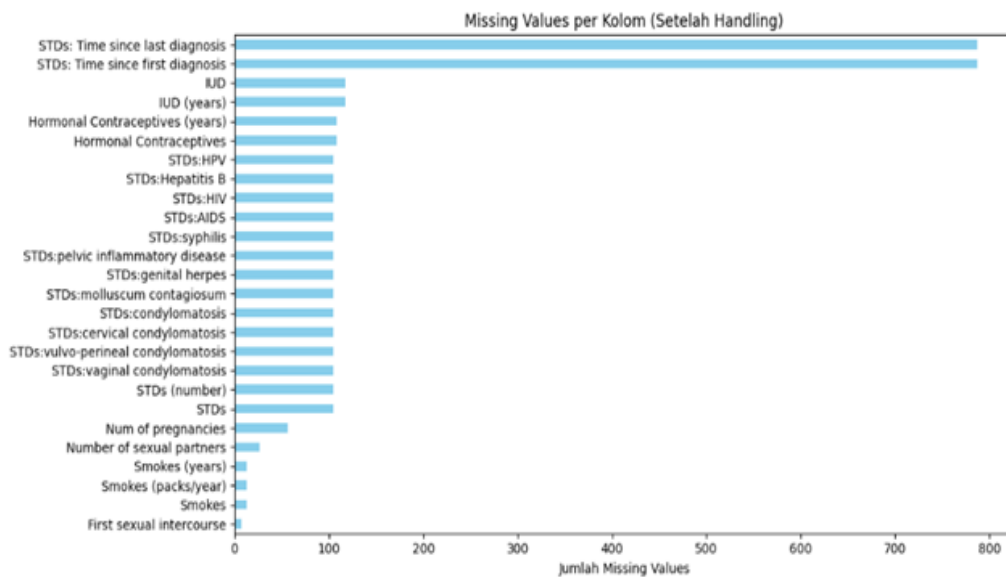
Dengan kerangka *Stratified K-Fold Cross-Validation*, seluruh metrik ini dihitung pada setiap *fold*, kemudian dirata-ratakan untuk memperoleh nilai akhir yang lebih reliabel. *Accuracy* memberikan gambaran umum, *precision* menilai akurasi prediksi positif, *Recall* mengukur kemampuan menangkap kasus kanker serviks, *F1-Score* menyajikan keseimbangan *precision* dan *Recall*, sedangkan AUC-ROC menunjukkan kemampuan model membedakan kelas positif dan negatif secara konsisten. Kombinasi metrik ini memastikan evaluasi model tidak hanya menekankan kinerja global, tetapi juga menangkap performa spesifik dalam mengidentifikasi kasus kanker serviks yang langka namun kritis. Dalam mengevaluasi performa model klasifikasi, pemilihan metrik yang tepat sangatlah krusial, terlebih ketika data yang digunakan tidak seimbang. Metrik seperti AUC-ROC (*Area Under the Receiver Operating Characteristic Curve*), *precision*, *Recall*, dan *F1-Score* sering dipakai untuk mengukur seberapa baik model bekerja. Menurut penelitian [25], nilai AUC ≥ 0.90 tergolong *excellent*, 0.80–0.90 termasuk *good*, 0.70–0.80 masuk kategori *fair*, sementara di bawah 0.70 dianggap *poor*. Panduan ini juga didukung oleh penelitian diagnostik terkini yang menggunakan AUC sebagai tolok ukur kemampuan diskriminasi model [26]. Tak hanya AUC, metrik *precision*, *Recall*, dan F1-skor juga turut diperhitungkan untuk menilai keseimbangan antara prediksi positif dan negatif. Penelitian terbaru di bidang *machine learning* untuk kesehatan menyebutkan bahwa *precision* sekitar 0.91, *Recall* 0.83, dan *F1-Score* 0.87 termasuk dalam kategori performa yang sangat baik dalam penerapan dunia nyata [27]. Mengacu pada standar ini, evaluasi model dalam penelitian ini mengevaluasi efektivitas teknik resampling seperti *Random Undersampling* dan *SMOTE*. Ringkasan interpretasi metrik nilai AUC-ROC [25][26] dan *precision*, *Recall*, dan F1-skor [27] ditunjukkan tabel 1 dibawah ini.

Tabel 1. Ringkasan tabel interpretasi metrik evaluasi

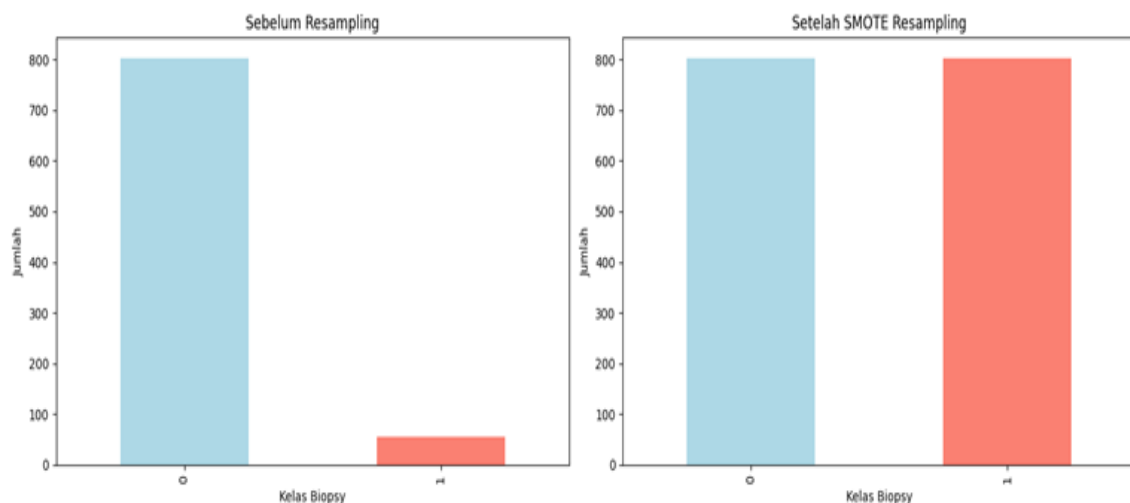
Metrik	Nilai ≥ 0.90	0.80 – 0.89	0.70 – 0.79	< 0.70
<i>AUC-ROC</i>	<i>Excellent</i> (sangat baik)	<i>Good</i> (baik)	<i>Fair</i> (cukup)	<i>Poor</i> (buruk)
<i>Precision</i>	Sangat baik (minim FP)	Baik	Cukup	Rendah
<i>Recall (Sensitivity)</i>	Sangat baik (tinggi TP)	Baik	Cukup	Rendah
<i>F1-score</i>	Sangat baik (seimbang P-R)	Baik	Cukup	Rendah

3. Hasil dan Analisis

Langkah awal sebelum dilakukan pengujian, tahapan *preprocessing* dilakukan untuk memastikan data siap digunakan dalam proses pelatihan model. *Dataset* kanker serviks dari UCI Repository berisi 858 observasi dengan 32 atribut. Banyak fitur pada *dataset* memiliki nilai kosong seperti ditunjukkan pada gambar 2. Penanganannya dilakukan dengan imputasi median agar distribusi data tetap terjaga. Kemudian semua variabel kategorikal diubah menjadi numerik menggunakan label *encoding* agar dapat diproses oleh algoritma berbasis numerik. Fitur dengan skala berbeda dinormalisasi menggunakan *StandardScaler* sehingga tiap fitur memiliki distribusi dengan mean = 0 dan standar deviasi = 1. Masalah utama pada *dataset* ini adalah *imbalance* (jumlah pasien dengan kanker serviks jauh lebih sedikit dibanding yang sehat). Jika dibiarkan, model cenderung bias memprediksi mayoritas (sehat). Solusi *SMOTE (Synthetic Minority Over-sampling Technique)* digunakan untuk menambah sampel sintesis pada kelas minoritas sehingga distribusi kelas menjadi lebih seimbang ditunjukkan gambar 3.

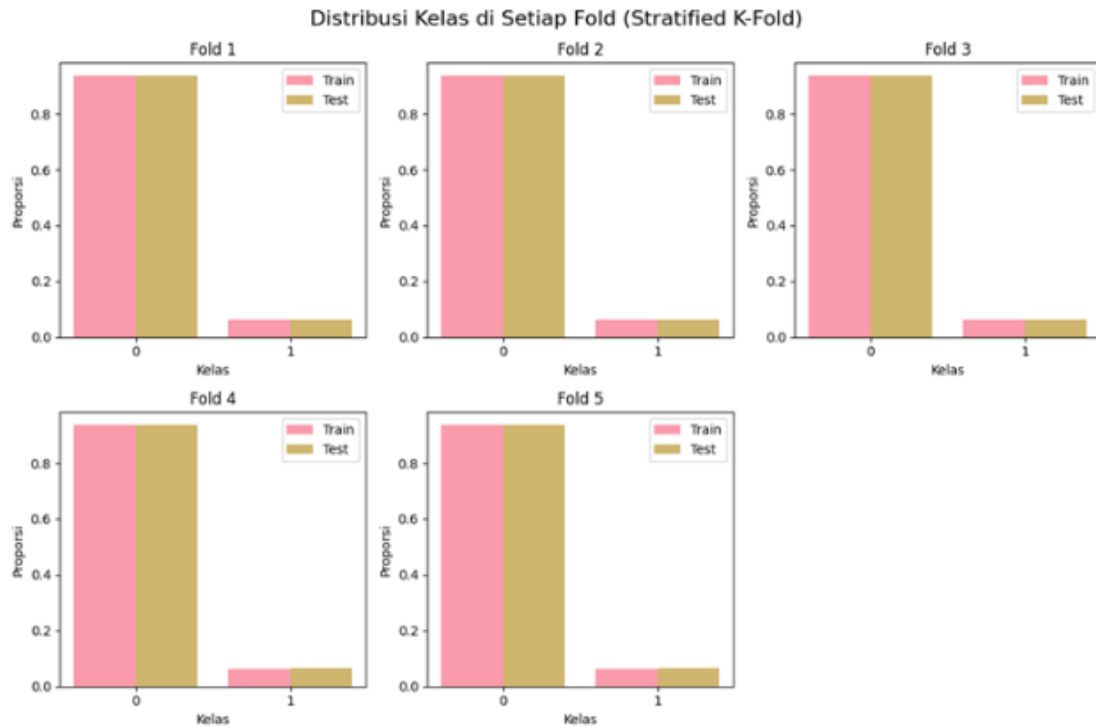


Gambar 2. Missing value pada setiap atribut



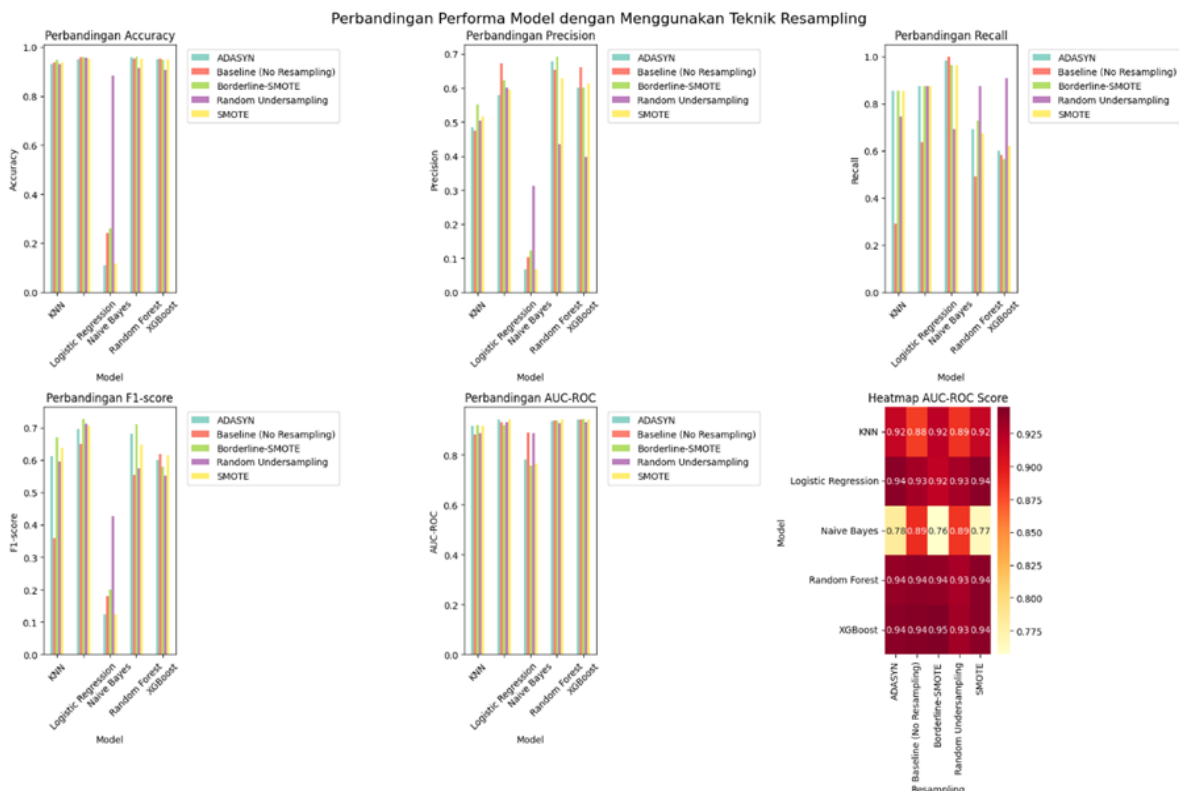
Gambar 3. Distribusi kelas biopsi

Stratified K-Fold (k = 5) digunakan agar distribusi kelas tetap proporsional di setiap *fold*. Dengan cara ini, tiap *fold* akan tetap merepresentasikan jumlah pasien kanker dengan sehat sesuai proporsi *dataset* asli ditunjukkan pada gambar 4. Hal ini mencegah data *leakage* dan menghasilkan evaluasi yang lebih *robust*.



Gambar 4. Distribusi kelas di setiap fold (Stratified K-Fold)

Setelah *preprocessing*, penelitian ini menggunakan lima model pembelajaran mesin yang berbeda seperti pada Gambar 5, *Logistic Regression (LR)*, *Random Forest (RF)*, *XGBoost (XGB)*, *K-Nearest Neighbor (KNN)*, dan *Naïve Bayes (NB)*. Pemilihan model didasarkan pada keberagaman pendekatan LR mewakili model linear, RF dan XGB mewakili pendekatan *ensemble* berbasis pohon, KNN sebagai model *instance-based learning*, dan NB sebagai model probabilistik. Kelima model ini dievaluasi pada kondisi *imbalance* maupun setelah dilakukan teknik *resampling*.



Gambar 5. Perbandingan performa model dengan menggunakan teknik *resampling*

Berdasarkan evaluasi dari model klasifikasi dan teknik *resampling* yang digunakan, didapat bahwa performa algoritma sangat dipengaruhi oleh pendekatan penanganan ketidakseimbangan kelas. RF menunjukkan performa *baseline* yang kuat dengan akurasi dan AUC-ROC yang tinggi (masing-masing 0.95 dan 0.93), namun memiliki *Recall* yang rendah (0.49), mengindikasikan bias terhadap kelas mayoritas. Penerapan *Random Undersampling* berhasil meningkatkan *Recall* secara drastis menjadi 0.87, meskipun diiringi penurunan *precision*. Teknik *oversampling* seperti *SMOTE* dan *ADASYN* mulai menciptakan keseimbangan yang lebih baik, sementara *Borderline-SMOTE* menghasilkan performa terbaik dengan akurasi 0.96, *F1-Score* 0.71, dan AUC 0.94, menjadikannya kombinasi yang paling stabil untuk model ini.

Pada LR, *baseline* sudah menunjukkan stabilitas yang cukup dengan *Recall* 0.63 dan AUC 0.93. Semua teknik *resampling*, khususnya *Random Undersampling*, *SMOTE*, dan *ADASYN*, berhasil mempertahankan *Recall* tinggi sekitar 0.87. *Borderline-SMOTE* memberikan hasil *F1-Score* tertinggi (0.73), menawarkan generalisasi yang baik dengan *Recall* tinggi tanpa mengorbankan *precision* secara berlebihan. Sementara itu, XGBoost menunjukkan konsistensi yang tinggi sejak *baseline* dengan *F1-Score* 0.61. Teknik *resampling* tidak memberikan peningkatan signifikan, dan justru menyebabkan penurunan *precision* pada *Random Undersampling*. Model ini paling cocok digunakan jika fokus utama adalah metrik AUC-ROC.

KNN merupakan model yang paling terbantu oleh teknik *resampling*. Performa *baseline*-nya rendah dengan *Recall* hanya 0.29, namun *oversampling* dengan *SMOTE* dan *ADASYN* berhasil mendongkrak *Recall* hingga di atas 0.85. *Borderline-SMOTE* menghasilkan performa terbaik untuk KNN dengan *F1-Score* 0.67 dan AUC 0.92. Sebaliknya, Naive Bayes menunjukkan ketidakcocokan dengan *dataset* ini. Model ini cenderung memprediksi hampir semua *instance* sebagai kelas minoritas (*Recall* 1.0), yang mengakibatkan akurasi sangat rendah (0.24) dan *precision* yang buruk. Berbagai teknik *resampling* tidak mampu memperbaiki performanya secara signifikan, sehingga tidak direkomendasikan untuk penggunaan lebih lanjut.

Tabel 2. Performa model dengan Accuracy, Precision, Recall, F1-Score, ROC-AUC, dan PR-AUC

<i>Resampling</i>	Model	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	AUC-ROC
<i>Baseline (No Resampling)</i>	<i>Random Forest</i>	0.952230	0.651746	0.490909	0.554444	0.938892
<i>Baseline (No Resampling)</i>	<i>Logistic Regression</i>	0.958058	0.672367	0.636364	0.648177	0.931161
<i>Baseline (No Resampling)</i>	XGBoost	0.953407	0.660202	0.581818	0.617662	0.942683
<i>Baseline (No Resampling)</i>	KNN	0.937060	0.473571	0.290909	0.358041	0.882270
<i>Baseline (No Resampling)</i>	Naive Bayes	0.240908	0.102151	1.000.000	0.178444	0.889841
<i>Random Undersampling</i>	<i>Random Forest</i>	0.913804	0.434536	0.872727	0.573784	0.929116
<i>Random Undersampling</i>	<i>Logistic Regression</i>	0.954556	0.601594	0.872727	0.710869	0.930292
<i>Random Undersampling</i>	XGBoost	0.905590	0.397629	0.909091	0.552296	0.931462
<i>Random Undersampling</i>	KNN	0.931307	0.503593	0.745455	0.594716	0.886995
<i>Random Undersampling</i>	Naive Bayes	0.883476	0.311271	0.690909	0.427832	0.886300
<i>SMOTE</i>	<i>Random Forest</i>	0.952237	0.628531	0.672727	0.647835	0.941050
<i>SMOTE</i>	<i>Logistic Regression</i>	0.953393	0.595746	0.872727	0.706271	0.940343
<i>SMOTE</i>	XGBoost	0.951081	0.612424	0.618182	0.614267	0.940083
<i>SMOTE</i>	KNN	0.937094	0.516253	0.854545	0.637680	0.915867
<i>SMOTE</i>	Naive Bayes	0.117659	0.065674	0.963636	0.122958	0.765739
<i>ADASYN</i>	<i>Logistic Regression</i>	0.951061	0.579152	0.872727	0.694744	0.940444
<i>ADASYN</i>	XGBoost	0.948735	0.600606	0.600000	0.599736	0.941443
<i>ADASYN</i>	KNN	0.930097	0.483333	0.854545	0.611696	0.917231
<i>ADASYN</i>	Naive Bayes	0.109520	0.066099	0.981818	0.123856	0.782752
<i>Borderline-SMOTE</i>	<i>Random Forest</i>	0.961553	0.692424	0.727273	0.709091	0.937660
<i>Borderline-SMOTE</i>	<i>Logistic Regression</i>	0.958051	0.623571	0.872727	0.726906	0.919474
<i>Borderline-SMOTE</i>	XGBoost	0.947579	0.601515	0.563636	0.578866	0.945164
<i>Borderline-SMOTE</i>	KNN	0.946403	0.552129	0.854545	0.669657	0.921017
<i>Borderline-SMOTE</i>	Naive Bayes	0.260683	0.122300	0.963636	0.200958	0.757027
<i>Borderline-SMOTE</i>	Naive Bayes	0.260683	0.122300	0.963636	0.200958	0.757027

Secara keseluruhan, kombinasi Borderline-SMOTE dengan RF atau LR memberikan hasil yang paling optimal dan seimbang untuk menangani masalah klasifikasi pada data yang tidak seimbang ini. Hasil lengkap evaluasi performa model dengan Accuracy, Precision, Recall, F1-Score, ROC-AUC, dan PR-AUC ditunjukkan pada tabel 2.

4. Simpulan

Hasil penelitian ini menemukan bahwa kombinasi teknik resampling, khususnya *Borderline-SMOTE*, dengan algoritma ensemble seperti *Random Forest* (RF) menghasilkan kinerja paling optimal untuk klasifikasi data tidak seimbang pada *dataset* kanker serviks. Secara kuantitatif, kombinasi *Borderline-SMOTE* + RF mencapai akurasi 0.96, F1-Score 0.71, dan AUC-ROC 0.94, dengan nilai *Recall* yang tinggi (0.87), menunjukkan kemampuan yang seimbang dan stabil dalam mendeteksi kasus kanker. Temuan ini berbeda dengan beberapa penelitian sebelumnya, seperti [18] yang lebih menekankan *Logistic Regression* (LR) tanpa eksplorasi mendalam teknik resampling. Penelitian ini mengisi celah (gap) tersebut dengan membuktikan bahwa resampling adaptif secara signifikan mampu meningkatkan *Recall*—sebuah metrik rusial dalam konteks medis. Sebagai contoh, pada model K-Nearest Neighbors (KNN), penerapan *Borderline-SMOTE* berhasil meningkatkan *Recall* dari 0.29 (*baseline*) menjadi 0.87, sementara pada LR, teknik yang sama mempertahankan *Recall* tinggi sekitar 0.87 dengan *F1-Score* tertinggi 0.73. Dengan demikian, penelitian ini membuktikan bahwa integrasi teknik resampling seperti *SMOTE*, *ADASYN*, dan khususnya *Borderline-SMOTE* ke dalam *pipeline* klasifikasi berbasis Stratified K-Fold dapat meningkatkan stabilitas performa dan memperkecil risiko bias kelas. Temuan ini dapat menjadi landasan bagi pengembangan sistem deteksi dini kanker serviks yang lebih andal serta penelitian lanjutan tentang optimasi metode resampling adaptif.

Referensi

- [1] World Health Organization (WHO), “Cervical Cancer,” *World Health Organization (WHO)*, Mar. 2024. <https://www.who.int/news-room/fact-sheets/detail/Cervical-Cancer> (accessed Aug. 01, 2025).
- [2] N. Al Mudawi and A. Alazeb, “A Model for Predicting Cervical Cancer Using Machine learning Algorithms,” *Sensors*, vol. 22, no. 11, p. 4132, May 2022, doi: 10.3390/s22114132.
- [3] M. L. R. University of California, Irvine, “Cervical Cancer (Risk Factors) Data Set,” *UCI Repository*, 2017.
- [4] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, “An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics,” *Inf. Sci. (Ny)*, vol. 250, pp. 113–141, Nov. 2013, doi: 10.1016/j.ins.2013.07.007.
- [5] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from Imbalanced Data Sets*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-98074-4.
- [6] Y. Shi, Y. Wan, K. Wu, and X. Chen, “Non-negativity and locality constrained Laplacian sparse coding for image classification,” *Expert Syst. Appl.*, vol. 72, pp. 121–129, Apr. 2017, doi: 10.1016/j.eswa.2016.12.012.
- [7] M. L. R. University of California, Irvine, “Cervical Cancer (Risk Factors) Data Set,” *UCI Repository*, 2017. <https://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk>
- [8] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, “An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics,” *Inf. Sci. (Ny)*, vol. 250, pp. 113–141, Nov. 2013, doi:

- 10.1016/j.ins.2013.07.007.
- [9] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from Imbalanced Data Sets*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-98074-4.
- [10] Y. Shi, Y. Wan, K. Wu, and X. Chen, "Non-negativity and locality constrained Laplacian sparse coding for image classification," *Expert Syst. Appl.*, vol. 72, pp. 121–129, Apr. 2017, doi: 10.1016/j.eswa.2016.12.012.
- [11] N. Almgren and H. Alshamlan, "A survey on hybrid feature selection methods in microarray gene expression data for *Cancer* classification," *IEEE Access*, vol. 7, pp. 78533–78548, 2019, doi: 10.1109/ACCESS.2019.2922987.
- [12] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, Oct. 2018, doi: 10.1016/j.neunet.2018.07.011.
- [13] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (*SMOTE*) for handling class imbalance," *Inf. Sci. (Ny)*, vol. 505, pp. 32–64, Dec. 2019, doi: 10.1016/j.ins.2019.07.070.
- [14] G. Kovács, "An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced *datasets*," *Appl. Soft Comput.*, vol. 83, p. 105662, Oct. 2019, doi: 10.1016/j.asoc.2019.105662.
- [15] H. Mo, H. Sun, J. Liu, and S. Wei, "Developing window behavior models for residential buildings using XGBoost algorithm," *Energy Build.*, vol. 205, pp. 1–23, 2019, doi: 10.1016/j.enbuild.2019.109564.
- [16] M. Z. Islam, V. Estivill-Castro, M. A. Rahman, and T. Bossomaier, "Combining K-MEANS and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering," *Expert Syst. Appl.*, vol. 91, pp. 402–417, 2018, doi: 10.1016/j.eswa.2017.09.005.
- [17] A. Geron, *Aurélien Géron - Hands on Machine learning with Scikit Learn Keras and TensorFlow. 2nd Edition-O'Reilly Media (2019)*, vol. 1, no. 0. O'Reilly Media, 2019. [Online]. Available: [http://repo.iain-tulungagung.ac.id/5510/5/BAB 2.pdf](http://repo.iain-tulungagung.ac.id/5510/5/BAB%202.pdf)
- [18] K. Fernandes, J. S. Cardoso, and J. Fernandes, "Transfer Learning with Partial Observability Applied to *Cervical Cancer* Screening," 2017, pp. 243–250. doi: 10.1007/978-3-319-58838-4_27.
- [19] A. Javed, B. S. Lee, and D. M. Rizzo, "A benchmark study on time series clustering," *arXiv*, vol. 1, no. June, p. 100001, 2020, doi: 10.1016/j.mlwa.2020.100001.
- [20] L. Breiman, *Springer Series in Statistics The Elements of*. CRC Press, 2017.
- [21] T. Chen, "XGBoost : A Scalable Tree Boosting System".
- [22] C. Bishop, *Information Science and Statistics*. Springer International Publishing, 2006.
- [23] K. Murphy, *Machine learning: A Probabilistic Perspective*, 1st ed. The MIT Press, 2012.
- [24] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2012.
- [25] J. N. Mandrekar, "Receiver Operating Characteristic Curve in Diagnostic Test Assessment," *J. Thorac. Oncol.*, vol. 5, no. 9, pp. 1315–1316, Sep. 2010, doi: 10.1097/JTO.0b013e3181ec173d.
- [26] Ş. K. Çorbacioğlu and G. Aksel, "Receiver operating characteristic curve analysis in diagnostic accuracy studies," *Turkish J. Emerg. Med.*, vol. 23, no. 4, pp. 195–198, Oct. 2023,

doi: 10.4103/tjem.tjem_182_23.

- [27] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Sci. Rep.*, vol. 12, no. 1, p. 5979, Apr. 2022, doi: 10.1038/s41598-022-09954-8.