Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

Analisis Kinerja Random Forest dan XGBoost serta Ensemble Soft Voting untuk Klasifikasi Stroke pada Dataset yang Tidak Seimbang

Bagus Fadhil Husain, Siska Kurnia Gusti*, Muhammad Fikry, Reski Mai Candra

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1]12250115278@students.uin-suska.ac.id, [2,*]siskakurniagusti@uin-suska.ac.id, [3]muhammad.fikry@uin-suska.ac.id, [4]reski.candra@uin-suska.ac.id
Email Penulis Korespondensi: siskakurniagusti@uin-suska.ac.id

Abstrak−Stroke adalah salah satu penyebab utama disabilitas dan kematian di dunia, sehingga diperlukan metode prediksi yang mampu membantu deteksi dini risiko stroke. Namun, karakteristik dataset medis yang umumnya memiliki distribusi kelas tidak seimbang menyebabkan proses klasifikasi menjadi lebih sulit. Penelitian ini bertujuan menganalisis kinerja Random Forest, XGBoost, dan Ensemble Soft Voting pada klasifikasi stroke menggunakan dataset Framingham Heart Study. Tahapan penelitian meliputi deduplikasi data, penanganan missing value menggunakan MICE, pembersihan outlier, standarisasi data, Hybrid Feature Selection berbasis literatur medis dan ANOVA F-Test, penyeimbangan kelas menggunakan SMOTE-ENN, serta optimasi model menggunakan GridSearchCV. Evaluasi dilakukan menggunakan Stratified K-Fold Cross Validation dengan metrik akurasi, presisi, recall, F1-Score, dan ROC-AUC. Hasil penelitian menunjukkan bahwa Random Forest memperoleh F1-Score dan recall tertinggi sebesar 36.52% dan 67.74%, XGBoost menghasilkan akurasi dan presisi tertinggi sebesar 82.98% dan 25.71%, sedangkan Ensemble Soft Voting memperoleh ROC-AUC tertinggi sebesar 81.29%. Meskipun demikian, Ensemble Soft Voting belum mampu memberikan peningkatan performa yang signifikan dibandingkan model tunggal karena tingginya overlapping antar kelas menyebabkan Random Forest dan XGBoost menghasilkan pola prediksi yang relatif serupa, sehingga proses Soft Voting tidak mampu meningkatkan performa klasifikasi secara optimal. Temuan ini menunjukkan bahwa karakteristik dataset memiliki pengaruh yang lebih dominan terhadap performa klasifikasi dibandingkan peningkatan kompleksitas model melalui pendekatan ensemble.

Kata Kunci: Stroke; Machine Learning; Random Forest; XGBoost; Ensemble Soft Voting

Abstract−Stroke is one of the leading causes of disability and death worldwide, making accurate prediction methods essential for early stroke risk detection. However, medical datasets are commonly characterized by class imbalance, which increases the complexity of the classification task. This study aims to analyze the performance of Random Forest, XGBoost, and Ensemble Soft Voting for stroke classification using the Framingham Heart Study dataset. The proposed framework consists of data deduplication, missing value imputation using MICE, outlier removal, data standardization, Hybrid Feature Selection based on medical literature and ANOVA F-Test, class balancing using SMOTE-ENN, and model optimization using GridSearchCV. Model performance was evaluated using Stratified K-Fold Cross Validation based on accuracy, precision, recall, F1-score, and ROC-AUC. The results show that Random Forest achieved the highest F1-score and recall of 36.52% and 67.74%, respectively, while XGBoost obtained the highest accuracy and precision of 82.98% and 25.71%. Ensemble Soft Voting achieved the highest ROC-AUC of 81.29%. Nevertheless, Ensemble Soft Voting did not provide a significant performance improvement over the individual models because the high degree of class overlap caused Random Forest and XGBoost to produce highly similar prediction patterns, limiting the effectiveness of the Soft Voting mechanism. These findings indicate that dataset characteristics have a greater impact on classification performance than increasing model complexity through an ensemble approach.

Keywords: Stroke; Machine Learning; Random Forest; XGBoost; Ensemble Soft Voting

1. PENDAHULUAN

Stroke adalah salah satu penyakit yang termasuk sebagai penyebab angka kematian dan disabilitas tertinggi di seluruh dunia [1], [2]. Berdasarkan data World Stroke Organization 2025, terdapat sekitar 12 juta orang mengalami stroke setiap tahunnya, dengan sekitar 6,5 juta penderita stroke meninggal dunia [3]. Selain itu, lebih dari 100 juta orang di dunia tercatat pernah mengalami stroke. Penyakit stroke adalah sindrom klinis yang ditandai dengan munculnya gangguan fungsi otak secara cepat akibat adanya kerusakan pada sistem pembuluh darah otak [4]. Kondisi ini terjadi karena terganggunya aliran darah menuju otak yang mengakibatkan kerusakan jaringan otak, disabilitas permanen, bahkan kematian [5]. Dampak yang sering dialami penderita stroke antara lain kelemahan otot pada bagian tangan dan kaki yang mengakibatkan turunnya kemampuan untuk melakukan aktivitas sehari-hari.

Terjadinya penyakit stroke dipengaruhi berbagai macam faktor risiko yaitu hipertensi, diabetes, obesitas, kadar kolesterol tinggi, kebiasaan merokok, serta faktor kesehatan lainnya [6]. Hingga saat ini, belum ada pengobatan yang mampu menyembuhkan stroke secara total setelah mengalaminya [7], [8]. Oleh karena itu, deteksi dan prediksi dini terhadap risiko stroke sangat diperlukan agar penanganan dapat dilakukan lebih awal sebelum kondisi pasien memburuk [9]. Seiring dengan pesatnya perkembangan teknologi digital, *Machine Learning* mulai banyak dimanfaatkan dalam bidang kesehatan untuk membantu analisis data medis dan mendeteksi pola penyakit berdasarkan riwayat kesehatan pasien serta membantu proses diagnosis medis [10], [11].

Beberapa penelitian menunjukkan bahwa berbagai algoritma *machine learning* bisa digunakan untuk mengidentifikasi risiko stroke. Salah satu yang sering digunakan adalah algoritma Random Forest, yaitu metode *ensemble learning* yang membangun beberapa pohon keputusan sekaligus dan menggabungkan hasil dari prediksi tiap pohonnya untuk memperoleh hasil prediksi yang lebih stabil dan efektif [12]. Penelitian yang dilakukan oleh [13] menunjukkan bahwa Random Forest menghasilkan akurasi sebesar 92.5% dalam klasifikasi stroke, lebih tinggi dibandingkan Naive Bayes dan K-Nearest Neighbor yang hanya mencapai akurasi sebesar 71.9% dan 73.6%. Penelitian lain oleh [14]

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 244
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

menggunakan kombinasi Random Forest dan metode ADASYN dengan *GridSearchCV* optimization, menghasilkan performa yang baik dengan akurasi 96.5% hingga 96.8% dan recall dari 93.7% hingga 94.3%. Penelitian [15] mengenai Random Forest juga menunjukkan performa yang baik dengan akurasi 93.6%, presisi 91.4%, recall 96.1%, dan juga F1-score sebesar 93.7%. Hasil tersebut mengindikasikan bahwa Random Forest memiliki performa yang baik dalam melakukan klasifikasi data medis.

Selain Random Forest, algoritma XGBoost juga menunjukkan performa yang sangat baik dalam berbagai penelitian klasifikasi penyakit. XGBoost merupakan pengembangan dari *gradient boosting* yang dirancang untuk meningkatkan kinerja model serta mengurangi risiko *overfitting* [16]. Pada penelitian [17], algoritma XGBoost berhasil memperoleh akurasi sebesar 95%, lebih tinggi dibandingkan Random Forest dengan akurasi 94% dan Logistic Regression sebesar 82%. Penelitian lain oleh [18], menunjukkan bahwa XGBoost yang dikombinasikan dengan seleksi fitur memperoleh akurasi 91.4%, lebih tinggi dibandingkan SVM yang hanya 89%. Selain itu, penelitian [19] pada dataset Wisconsin Breast Cancer Diagnostic menunjukkan bahwa XGBoost mampu menghasilkan nilai ROC-AUC sebesar 0.99. Hasil tersebut membuktikan bahwa XGBoost memiliki kemampuan yang baik dalam menangani data kesehatan.

Walaupun kombinasi Random Forest dan XGBoost telah banyak digunakan pada berbagai permasalahan klasifikasi, efektivitas pendekatan tersebut pada dataset stroke yang memiliki distribusi kelas tidak seimbang serta tingkat overlapping antar kelas yang tinggi masih belum banyak dibahas. Karakteristik data seperti ini berpotensi menyebabkan model mengalami kesulitan dalam mengenali kelas minoritas meskipun telah diterapkan metode ensemble. Oleh karena itu, penelitian ini berfokus pada analisis kemampuan Random Forest, XGBoost, dan Ensemble Soft Voting dalam menghadapi karakteristik data tersebut melalui penerapan teknik resampling SMOTE-ENN.

Ensemble learning merupakan salah satu teknik dari *Machine Learning* yang menggabungkan beberapa model untuk memperoleh hasil prediksi dengan kinerja yang lebih akurat juga stabil daripada model tunggal [16]. Teknik ini memanfaatkan kekuatan berbagai model untuk mendapatkan hasil yang lebih stabil dan juga akurat daripada model tunggal. Dalam berbagai penelitian sebelumnya, metode ensemble terbukti mampu meningkatkan akurasi pada berbagai kasus klasifikasi di bidang kesehatan. Selain meningkatkan akurasi, metode ensemble juga mampu mengurangi risiko kesalahan prediksi yang dapat terjadi ketika hanya mengandalkan satu model klasifikasi. Salah satu pendekatan yang dapat diterapkan adalah kombinasi algoritma XGBoost dan Random Forest untuk meningkatkan efektivitas dalam klasifikasi penyakit stroke [20].

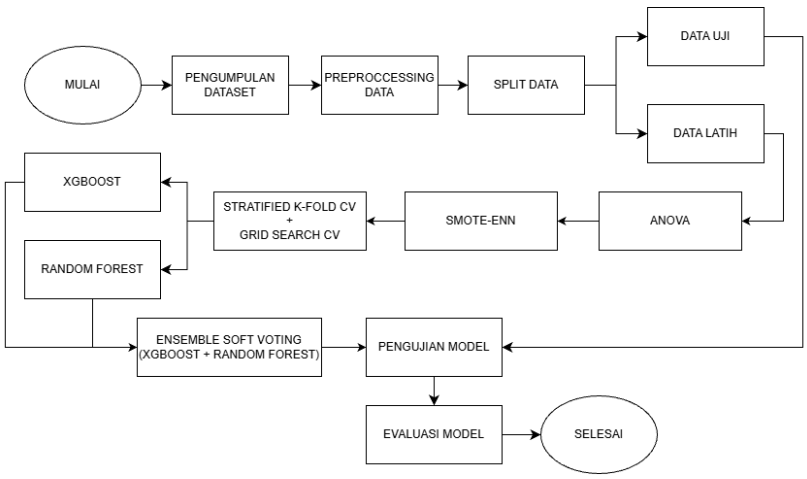
Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengevaluasi efektivitas metode *ensemble learning* berbasis kombinasi algoritma Random Forest dan XGBoost dalam klasifikasi penyakit stroke. Penelitian ini juga bertujuan untuk menganalisis kemampuan model ensemble dalam meningkatkan performa klasifikasi dibandingkan penggunaan model tunggal melalui pengukuran nilai akurasi, presisi, recall, F1-score, dan ROC-AUC.

Penelitian ini memberikan kontribusi melalui evaluasi komprehensif terhadap performa Random Forest, XGBoost, dan Ensemble Soft Voting pada dataset stroke dengan karakteristik kelas tidak seimbang dan overlapping antar kelas yang tinggi. Selain membandingkan performa model, penelitian ini juga menganalisis pengaruh karakteristik data terhadap kemampuan model dalam mendeteksi kasus stroke sehingga dapat memberikan gambaran mengenai keterbatasan pendekatan ensemble pada kondisi data tersebut.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian akan melakukan beberapa tahapan seperti yang ada pada **Gambar 1**.



Gambar 1. Tahapan Penelitian

Penelitian ini dilakukan untuk melakukan klasifikasi penyakit stroke menggunakan metode ensemble learning berbasis kombinasi algoritma XGBoost dan Random Forest. Tahapan dimulai dari pengumpulan dataset, preprocessing

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 245
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

data, pembagian data, seleksi fitur hibrida, penanganan ketidakseimbangan kelas (*resampling*), proses pelatihan model, hingga evaluasi performa model. Tahapan penelitian bertujuan agar memperoleh model klasifikasi dengan performa yang optimal dalam mendeteksi penyakit stroke.

2.2 Pengumpulan Data

Dataset yang digunakan pada penelitian ini adalah Framingham Heart Study Dataset yang diperoleh dari Kaggle (https://www.kaggle.com/datasets/shreyjain601/framingham-heart-study?resource=download). Dataset ini berasal dari Framingham Heart Study, yaitu studi kohort jangka panjang yang dikembangkan oleh National Heart, Lung, and Blood Institute (NHLBI) sejak tahun 1948 untuk mengidentifikasi faktor-faktor risiko penyakit kardiovaskular. Dataset ini dipilih karena memiliki sumber data yang jelas, terdokumentasi dengan baik, serta memuat berbagai faktor risiko yang berkaitan dengan kejadian stroke, seperti usia, tekanan darah, kadar kolesterol, diabetes, indeks massa tubuh, dan kebiasaan merokok. Selain itu, dataset ini telah digunakan dalam penelitian machine learning di bidang kesehatan, termasuk penelitian [21] yang memanfaatkan Framingham Heart Study Dataset dalam analisis dislipidemia dan kardiovaskular. Meskipun demikian, pada penelitian ini dataset dimanfaatkan secara khusus untuk membangun model klasifikasi penyakit stroke dengan memanfaatkan atribut-atribut faktor risiko yang tersedia. Oleh karena itu, Framingham Heart Study Dataset dinilai relevan untuk pengembangan model klasifikasi stroke pada penelitian ini. Dataset ini terdiri atas 11.627 baris data dengan 39 atribut. Dataset ini memiliki karakteristik distribusi kelas yang sangat tidak seimbang, yaitu sebanyak 1.061 data Stroke dan 10.566 data No Stroke. Dataset ini terdiri dari beberapa atribut meliputi *RANDID*, *SEX*, *TOTCHOL*, *AGE*, *SYSBP*, dan atribut lainnya. Seperti yang ditunjukkan pada **Tabel 1**.

Tabel 1. Dataset Awal

RANDID	SEX	TOTCHOL	AGE	…	TIMECVD	TIMEDTH	TIMEHYP
2448	1	195	39	...	6438	8766	8766
2448	1	209	52	...	6438	8766	8766
6238	2	250	46	...	8766	8766	8766
6238	2	260	52	...	8766	8766	8766
...	...	...	...	...	...	...	...
9999312	2	-	50	...	8766	8766	4201

2.3 Preprocessing

Sebelum proses pemodelan, data akan terlebih dahulu melalui tahap preprocessing untuk menghasilkan dataset yang lebih terstruktur dan siap digunakan. Tahap ini berperan dalam meningkatkan kualitas data, mengurangi atribut yang tidak diperlukan, serta menyesuaikan format data agar sesuai dengan kebutuhan algoritma *Machine Learning*. Proses preprocessing pada penelitian ini terdiri atas seleksi atribut berdasarkan literatur medis, deduplikasi data, penanganan missing value, dan penghapusan data *outlier*.

2.3.1 Seleksi Atribut Berbasis Literatur Medis

Sebelum memasuki tahapan pembersihan data secara komputasional, dataset awal yang memiliki 39 atribut direduksi terlebih dahulu secara manual. Proses seleksi mengacu pada WHO Stroke Fact Sheet (2025) yang mengidentifikasi hipertensi, merokok, diabetes, kadar kolesterol tinggi, obesitas, peningkatan kadar glukosa darah, usia, dan jenis kelamin sebagai faktor risiko utama stroke [22]. Selain itu, penelitian [23] menegaskan bahwa tekanan darah sistolik tinggi merupakan salah satu kontributor terbesar terhadap beban stroke global, disertai faktor risiko lain seperti diabetes, merokok, dan gangguan profil lipid. Berdasarkan pertimbangan tersebut, dipilih 12 atribut dan 1 label target yang memiliki relevansi klinis terhadap penyebab penyakit stroke, seperti yang dapat dilihat pada **Tabel 2**.

Tabel 2. Atribut Hasil Seleksi

Atribut	Keterangan
AGE	Usia
SEX	Jenis kelamin
TOTCHOL	Total kadar kolesterol
SYSBP	Tekanan sistolik
DIABP	Tekanan diastolik
CURSMOKE	Status perokok aktif
CIGPDAY	Jumlah konsumsi rokok per hari
BMI	Indeks Massa Tubuh
DIABETES	Status diabetes
BPMEDS	Penggunaan obat antihipertensi atau obat tekanan darah
GLUCOSE	Kadar glukosa darah
PREVHYP	Riwayat hipertensi
STROKE	Label target

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 246
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

2.3.2 Deduplikasi Data

Karena Framingham Dataset bersifat longitudinal, maka dilakukan proses deduplikasi. Proses ini mengambil rekam medis pada periode terakhir berdasarkan *RANDID* untuk mendapatkan representasi kondisi kesehatan terbaru. Setelah proses deduplikasi, jumlah data menyusut dari 11627 baris menjadi 4434 data pasien unik. Pada tahap ini, distribusi kelas yang perlihatkan dataset ini sangat tidak seimbang, yaitu sebanyak 4019 data kelas No Stroke dan 415 data kelas Stroke.

2.3.3 Penanganan Missing Value

Karakteristik data medis sering kali memiliki nilai yang hilang atau missing values pada beberapa fiturnya. Sebelum proses imputasi dilakukan, dataset sudah terlebih dahulu dideduplikasi berdasarkan RANDID dengan mempertahankan observasi terakhir dari setiap pasien sehingga setiap individu hanya direpresentasikan oleh satu baris data. Dengan demikian, proses imputasi dilakukan pada data yang bersifat cross-sectional dan tidak melibatkan pengamatan berulang dari pasien yang sama. Penelitian ini tidak membuang baris pasien yang memiliki data kosong, melainkan mengisinya menggunakan metode MICE (*Multivariate Imputation by Chained Equations*). MICE bekerja dengan cara memprediksi dan mengisi nilai yang hilang secara iteratif dengan memodelkan setiap fitur sebagai fungsi dari fitur-fitur lainnya, sehingga integritas dan jumlah pasien tetap terjaga [24].

2.3.4 Pembersihan Outlier

Untuk mengurangi pengaruh nilai ekstrem yang berpotensi mengganggu proses pembelajaran model dan meningkatkan kualitas data, dilakukan pembersihan data *outlier*. Pembersihan menggunakan dua pendekatan statistik yang disesuaikan dengan karakteristik distribusi atributnya [25]. Pertama, Metode Z-Score diterapkan pada atribut yang berdistribusi normal, di mana data yang menyimpang lebih dari tiga kali lipat standar deviasi dari nilai rata-rata diklasifikasikan sebagai anomali dan dihapus. Kemudian yang kedua, Metode IQR (Interquartile Range) diterapkan pada atribut yang distribusinya *skewed.* Karena metode IQR didasarkan pada rentang antarkuartil sehingga sifatnya lebih kebal (*robust*) terhadap nilai ekstrem, data yang berada terlalu jauh di bawah batas kuartil bawah atau melampaui batas kuartil atas diidentifikasi sebagai *outlier* dan dihapus guna menjamin stabilitas prediksi model.

2.4 Pembagian Data

Setelah preprocessing, dataset akan dipisahkan menjadi data latih dan juga data uji dengan metode *train-test split*. Pada kasus ini digunakan skenario pembagian data 70:30, 80:20, dan juga 90:10 guna mengevaluasi pengaruh jumlah data latih dan juga data uji terhadap performa model klasifikasi penyakit stroke. Selain itu, proses pembagian data akan menerapkan parameter *stratify* agar proporsi kelas target tetap terjaga pada data latih dan juga data uji sehingga komposisi kelas pada kedua subset tetap representatif terhadap dataset asli.

2.5 Standarisasi Data dan Seleksi Fitur ANOVA F-Test

Setelah proses pembagian data selesai dilakukan, atribut numerik distandarisasi menggunakan metode StandardScaler pada data latih. Atribut yang distandarisasi meliputi AGE, TOTCHOL, SYSBP, DIABP, CIGPDAY, BMI, dan GLUCOSE. Standarisasi bertujuan untuk menyamakan skala antar atribut sehingga setiap atribut memiliki rata-rata mendekati nol dan juga standar deviasi mendekati satu. Tahap ini penting karena atribut yang digunakan memiliki rentang nilai yang berbeda-beda. Dengan adanya standarisasi, setiap atribut dapat memberikan pengaruh yang lebih stabil dalam proses pembelajaran model. Untuk menghindari terjadinya data leakage, parameter standardisasi dihitung menggunakan data latih dan selanjutnya diterapkan pada data uji.

Setelah proses standarisasi selesai dilakukan, penelitian ini menerapkan pendekatan *Hybrid Feature Selection* yang mengombinasikan seleksi atribut berbasis literatur medis dengan metode statistik ANOVA F-Test. Tahap pertama dilakukan berdasarkan literatur medis, selanjutnya atribut yang telah dipilih dievaluasi menggunakan metode ANOVA F-Test melalui SelectKBest untuk mencari fitur yang memiliki hubungan paling signifikan terhadap variabel target. Pendekatan ini bertujuan untuk mempertahankan relevansi klinis atribut sekaligus mengurangi fitur yang kurang relevan sehingga proses pembelajaran model dapat berlangsung secara lebih efektif [26]. Dalam penelitian ini, ANOVA F-Test digunakan sebagai tahap seleksi fitur statistik untuk menyaring atribut yang telah dipilih berdasarkan literatur medis. Oleh karena itu, pemilihan fitur pada penelitian ini mempertimbangkan baik aspek statistik maupun relevansi klinis. Pemilihan metode ini didasarkan pada kemampuannya dalam mengidentifikasi fitur yang memiliki kontribusi statistik rendah sehingga dapat mengurangi dimensi data sebelum proses pembelajaran model dilakukan. Meskipun ANOVA F-Test memiliki keterbatasan karena hanya mengevaluasi hubungan linear antara fitur dan target, Random Forest dan XGBoost tetap mampu mempelajari hubungan non-linear serta interaksi kompleks antar fitur pada tahap pemodelan. Oleh karena itu, ANOVA F-Test pada penelitian ini berfungsi sebagai proses reduksi dimensi awal, sedangkan kemampuan pemodelan hubungan non-linear sepenuhnya ditangani oleh algoritma klasifikasi yang digunakan.

2.6 SMOTE-ENN

Dataset hasil pembagian data masih menunjukkan ketidakseimbangan kelas antara kelompok Stroke dan No Stroke. Oleh karena itu, dilakukan penyeimbangan data dengan menggunakan metode SMOTE-ENN. Metode ini merupakan metode

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 247
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

hybrid yang mengombinasikan teknik oversampling dan undersampling. SMOTE digunakan untuk menghasilkan sampel sintetis pada kelas minoritas sehingga distribusi pada kelas menjadi lebih seimbang. Selanjutnya, ENN digunakan untuk menghapus data yang dianggap sebagai noise atau memiliki kedekatan yang tinggi dengan kelas lain. Kombinasi kedua metode tersebut memungkinkan distribusi data menjadi lebih seimbang sekaligus meningkatkan kualitas data hasil resampling [27]. Proses SMOTE-ENN hanya diterapkan pada data latih agar terhindar dari data *leakage* serta memastikan bahwa proses evaluasi model dilakukan secara objektif dan tidak bias.

2.7 Analisis Distribusi dan Visualisasi Dataset

Analisis distribusi data dilakukan agar memahami karakteristik dari dataset sebelum proses klasifikasi dilakukan. Tahap ini bertujuan untuk mengidentifikasi pola persebaran data, hubungan antar kelas, serta kemungkinan terjadinya *overlapping* antara kelas stroke dan no stroke yang dapat memengaruhi performa model klasifikasi. Visualisasi data dilakukan menggunakan metode *Principal Component Analysis* (PCA) dan juga *t-distributed Stochastic Neighbor Embedding* (t-SNE) dengan mereduksi dimensi data ke dalam ruang dua dimensi. PCA digunakan untuk memvisualisasikan distribusi data secara linear berdasarkan variansi terbesar pada dataset, sedangkan t-SNE digunakan untuk memetakan hubungan non-linear dan mempertahankan kedekatan lokal antar data sehingga pola klaster dan *overlapping* antar kelas dapat diamati dengan lebih jelas.

2.8 *K-Fold Cross Validation*

K-Fold Cross Validation adalah teknik validasi yang sering diterapkan untuk mengukur performa model *Machine Learning* terhadap data yang belum pernah dilihat sebelumnya [28]. Dalam metode ini, data dikelompokkan ke dalam sejumlah subset dengan jumlah data yang seimbang pada setiap subset, setelah itu proses pelatihan dan pengujian model dilakukan secara berulang sebanyak nilai *k* yang digunakan. Secara bergantian, masing-masing subset akan digunakan sebagai data validasi, sedangkan subset lainnya tetap digunakan dalam proses pembelajaran model. Setelah seluruh iterasi selesai dilakukan, rata-rata hasil evaluasi dihitung sebagai estimasi performa model. Dalam penelitian ini digunakan *Stratified K-Fold Cross Validation* dengan variasi nilai *k* mulai dari 3 hingga 20. Penggunaan rentang nilai tersebut didasarkan pada penelitian [28] yang menyimpulkan tidak terdapat nilai *k* yang optimal secara universal untuk seluruh dataset dan algoritma, sehingga pemilihan nilai *k* perlu disesuaikan berdasarkan karakteristik data serta model yang digunakan. Selain itu, penelitian tersebut merekomendasikan agar nilai k diperlakukan sebagai parameter sensitivitas yang perlu diuji pada beberapa skenario untuk mengevaluasi pengaruhnya terhadap bias, variansi, dan performa model. Oleh karena itu, variasi nilai k dari 3 hingga 20 diterapkan untuk mengevaluasi kestabilan performa model terhadap perubahan jumlah fold serta menganalisis pengaruhnya terhadap hasil klasifikasi pada dataset stroke yang digunakan. Konfigurasi terbaik kemudian dipilih berdasarkan rata-rata nilai F1-Score sebagai metrik utama pada data yang tidak seimbang.

2.9 *GridSearchCV*

Selain itu, penelitian ini menerapkan metode *GridSearchCV* untuk memperoleh kombinasi *hyperparameter* terbaik dari algoritma Random Forest dan XGBoost. *GridSearchCV* merupakan metode untuk memperoleh kombinasi parameter terbaik secara sistematis melalui proses evaluasi pada setiap kombinasi parameter yang telah ditentukan. Metode ini memanfaatkan teknik *cross-validation* untuk menguji berbagai kombinasi hyperparameter sehingga mampu menghasilkan konfigurasi model yang lebih optimal dan stabil [29], [30]. Parameter yang diuji pada algoritma Random Forest yaitu *n_estimators*, *max_depth*, *min_samples_leaf*, *min_samples_split*, *max_features*, dan *bootstrap*. Sedangkan pada algoritma XGBoost, parameter yang diuji meliputi *n_estimators*, *max_depth*, *learning_rate*, *colsample_bytree* dan *subsample*. Proses *GridSearchCV* akan mengevaluasi setiap kombinasi menggunakan *cross-validation* untuk mendapatkan performa yang paling optimal [31].

2.10 Random Forest

Random Forest adalah metode klasifikasi berbasis ensemble yang menggunakan teknik bagging dengan membentuk banyak pohon keputusan sekaligus dari sampel data yang berbeda. Hasil akhir dari prediksinya akan didapat dari penggabungan hasil prediksi dari setiap pohon sehingga keputusan yang dihasilkan lebih stabil dan *robust* terhadap variasi data [32], [33]. Pada kasus klasifikasi, hasil prediksi ditentukan berdasarkan kelas yang memperoleh suara terbanyak (*majority voting*). Random Forest menggunakan Gini Index untuk menentukan kualitas pemisahan data pada tiap node pohon keputusan. Nilai Gini Index dapat dihitung menggunakan Persamaan (1).

$$Gini(S) = 1 - \sum_{i=1}^{c} p_i^2 \qquad (1)$$

Keterangan pada rumus tersebut menunjukkan bahwa (S) merepresentasikan *dataset* atau kumpulan data yang digunakan dalam proses analisis. Variabel (c) menyatakan jumlah kelas atau kategori yang terdapat dalam dataset, sedangkan (p_i) merupakan probabilitas dari setiap kelas ke-(i), yang menunjukkan peluang kemunculan masing-masing kelas dalam dataset.

Dalam penelitian ini, optimasi hyperparameter Random Forest dilakukan dengan metode *GridSearchCV* untuk menghasilkan kombinasi terbaik. Nilai-nilai yang digunakan ditunjukkan pada **Tabel 3**.

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 248
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

Tabel 3. Random Forest Hyperparameter

Hyperparameter	Nilai
n_ estimators	100, 200, 300, 400, 500
max_depth	5, 10, 20
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 4
max_features	sqrt, log2
bootstrap	True, False

2.11 XGBoost

Extreme Gradient Boosting atau biasa disebut XGBoost adalah algoritma yang dikembangkan dari metode *Gradient Boosting* untuk meningkatkan akurasi prediksi dan efisiensi komputasi [18]. XGBoost membangun model secara bertahap dengan menambahkan keputusan baru yang sudah memperbaiki kesalahan prediksi dari pohon sebelumnya [34]. XGBoost mengoptimalkan hyperparameter seperti *max_depth*, *learning rate*, dan jumlah pohon (n_estimators) untuk meminimalkan kesalahan prediksi. Dengan akurasi tinggi dan kemampuannya untuk menangani data kompleks, *XGBoost* menjadi algoritma yang efektif dalam klasifikasi penyakit stroke. Sama seperti sebelumnya, XGBoost mencari kombinasi parameter terbaik menggunakan *GridSearchCV*. Hyperparameter yang digunakan diperlihatkan di **Tabel 4**.

Tabel 4. XGBoost Hyperparameter

Hyperparameter	Nilai
n_ estimators	100, 200, 300
max_depth	3, 5, 7, 10
learning_rate	0.01, 0.05, 0.1
colsample_bytree	0.8, 1.0
subsample	0.8, 1.0

2.12 *Ensemble Soft Voting*

Di penelitian ini, algoritma Random Forest dan XGBoost digunakan untuk membangun model klasifikasi penyakit stroke. Hasil prediksi dari kedua model kemudian digabung dengan metode *Ensemble Soft Voting* yang bertujuan untuk meningkatkan kinerja klasifikasi dengan menggabungkan keunggulan masing-masing algoritma. Teknik ensemble learning ini sering digunakan untuk menghasilkan performa yang lebih stabil, akurat, dan menghasilkan performa yang lebih baik dibandingkan penggunaan model tunggal [35]. Metode *Soft Voting* bekerja dengan menggabungkan probabilitas prediksi yang dihasilkan oleh tiap model. Kemudian probabilitas akan dirata-ratakan dan kelas dengan nilai probabilitas yang paling tinggi akan dipilih sebagai hasil prediksi akhir [36].

2.13 Model Evaluasi

Kinerja tiap model akan dianalisis melalui pendekatan *Confusion Matrix* yang berfungsi sebagai dasar perhitungan berbagai metrik. *Confusion Matrix* mengelompokkan hasil klasifikasi ke dalam empat kategori, yaitu TP, TN, FP, dan FN. Berdasarkan nilai tersebut, performa model dihitung menggunakan beberapa metrik, yaitu akurasi, presisi, recall, F1-score, dan ROC-AUC. Metrik-metrik tersebut digunakan untuk mengevaluasi dan membandingkan performa algoritma Random Forest, XGBoost, dan *Ensemble Soft Voting* dalam mengklasifikasikan data Stroke dan No Stroke. Metrik F1-score kemudian digunakan sebagai metrik utama dalam menentukan model terbaik karena dataset yang digunakan menunjukkan ketidakseimbangan kelas yang cukup tinggi. Selain itu, Recall turut dipertimbangkan untuk menilai kemampuan model dalam mengidentifikasi kasus Stroke sebagai kelas minoritas. Sementara itu, ROC-AUC berperan untuk mengevaluasi performa dari tiap model dalam membedakan kelas positif dan juga kelas negatif pada berbagai nilai threshold. Model dengan nilai evaluasi yang lebih tinggi menunjukkan performa klasifikasi yang lebih baik dalam memprediksi risiko stroke.

3. HASIL DAN PEMBAHASAN

Dataset di penelitian ini terdiri atas 11.627 baris data dan 39 atribut yang berasal dari Framingham Heart Study. Dataset tersebut memiliki distribusi kelas yang sangat tidak seimbang, yaitu 1061 data *Stroke* dan 10566 data *No Stroke*. Setelah melalui tahapan pembersihan awal deduplikasi, diperoleh 4.434 rekam medis pasien unik. Dataset tersebut memiliki karakteristik ketimpangan kelas yang sangat tinggi (*highly imbalanced*), dengan 4.019 data *No Stroke* dan hanya 415 data *Stroke*. Tahapan analisis diawali dengan tahap preprocessing, pembagian data serta seleksi fitur menggunakan metode ANOVA F-Test. Kemudian, proses validasi model menggunakan *Stratified K-Fold Cross Validation* dengan variasi nilai k mulai dari 3 hingga 20. Pada tahap pemodelan, algoritma Random Forest dan XGBoost dioptimasi menggunakan *GridSearchCV* dan digabungkan menggunakan metode *Ensemble Soft Voting*. Tahap akhir dilakukan dengan mengevaluasi kinerja model dengan *Confusion Matrix* dan juga kurva ROC-AUC. Tahap evaluasi dilakukan untuk

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 249
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

mengukur kemampuan algoritma Random Forest, XGBoost, dan *Ensemble Soft Voting* dalam mengklasifikasikan data *Stroke* dan *No Stroke* serta menentukan model dengan performa terbaik berdasarkan nilai akurasi, presisi, recall, F1-score, dan ROC-AUC.

3.1 Hasil Preprocessing

Hasil preprocessing memperlihatkan adanya perubahan pada struktur dataset setelah dilakukan seleksi atribut, deduplikasi data, penanganan missing value, dan pembersihan outlier. Dataset awal yang terdiri dari 39 atribut mengalami pengurangan atribut setelah proses seleksi fitur. Perubahan struktur atribut setelah seleksi atribut berdasarkan literatur medis ditunjukkan pada **Tabel 5**.

Tabel 5. Atribut Setelah Seleksi Fitur

Tahapan	Jumlah Atribut
Dataset Awal	39 atribut
Seleksi Atribut Berdasarkan Literatur Medis	12 atribut

Selanjutnya, karena dataset bersifat longitudinal maka dilakukan proses deduplikasi berdasarkan atribut RANDID dengan mempertahankan rekam medis terakhir setiap pasien. Hasil deduplikasi ditunjukkan pada Tabel 6.

Tabel 6. Jumlah Data Setelah Deduplikasi

Tahapan	Jumlah Data
Dataset Awal	11.627 baris data
Hasil deduplikasi	4.434 baris data

Kemudian dilakukan penanganan missing value menggunakan metode MICE. Sebelum proses imputasi, ditemukan nilai hilang pada atribut TOTCHOL, CIGPDAY, BMI, BPMEDS, dan GLUCOSE. Setelah proses imputasi dilakukan, seluruh missing value berhasil ditangani sehingga tidak terdapat lagi nilai kosong pada dataset Tahap berikutnya adalah pembersihan outlier menggunakan kombinasi metode Z-Score dan IQR. Hasil proses ini mengurangi jumlah data yang bisa dilihat pada **Tabel 7**.

Tabel 7. Jumlah Data Setelah Penanganan Outlier

Tahapan	Jumlah Data
Dataset Awal	11.627 baris data
Hasil deduplikasi	4.434 baris data
Hasil Penanganan Outlier	3.817 baris data

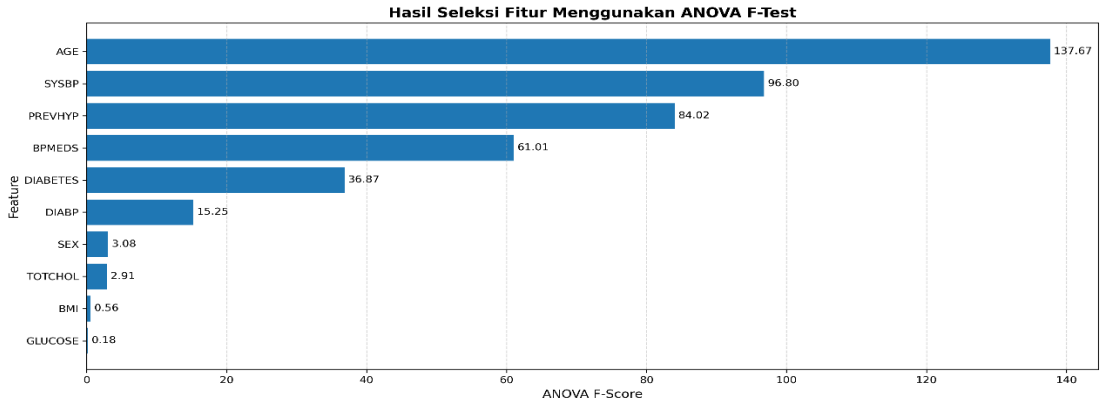
Berdasarkan hasil preprocessing yang telah dilakukan, kualitas dataset mengalami peningkatan melalui proses seleksi atribut, deduplikasi data, penanganan missing value, dan pembersihan outlier. Jumlah data berkurang dari 11.627 observasi menjadi 3.817 observasi yang siap digunakan pada tahap analisis selanjutnya. Selain itu, seluruh nilai hilang berhasil ditangani menggunakan metode MICE sehingga tidak terdapat missing value pada dataset. Hasil preprocessing ini akan memperoleh dataset yang bersih, konsisten, dan representatif untuk digunakan diproses seleksi fitur, penyeimbangan data, serta pembentukan model klasifikasi stroke.

3.2 Hasil Seleksi Fitur ANOVA F-Test

Setelah proses preprocessing selesai dilakukan, seleksi fitur dengan ANOVA F-Test diterapkan untuk mengevaluasi tingkat hubungan setiap atribut terhadap variabel target (STROKE). Penelitian ini menggunakan metode SelectKBest dengan nilai $k = 10$, sehingga hanya 10 atribut dengan nilai F-score tertinggi yang akan digunakan pada tahap klasifikasi. Hasil seleksi fitur menggunakan ANOVA F-Test ditunjukkan **Gambar 2**.

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 250
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds



Gambar 2. Visualisasi Hasil Seleksi Fitur ANOVA F-Test

Berdasarkan nilai F-score yang diperoleh, atribut AGE, SYSBP, PREVHYP, dan BPMEDS menjadi fitur yang paling berpengaruh dalam membedakan kelas Stroke dan No Stroke. Secara umum, hasil ANOVA menunjukkan bahwa faktor usia dan hipertensi memiliki kontribusi yang dominan terhadap kejadian stroke pada dataset ini. Sepuluh atribut yang terpilih setelah dilakukannya seleksi fitur ANOVA ditunjukkan pada **Tabel 8**.

Tabel 8. Sepuluh Fitur Terpilih Hasil ANOVA F-Test

Atribut	Keterangan
AGE	Usia
SEX	Jenis kelamin
TOTCHOL	Total kadar kolesterol
SYSBP	Tekanan sistolik
DIABP	Tekanan diastolik
BMI	Indeks Massa Tubuh
DIABETES	Status diabetes
BPMEDS	Penggunaan obat antihipertensi atau obat tekanan darah
GLUCOSE	Kadar glukosa darah
PREVHYP	Riwayat hipertensi

3.3 Hasil Penyeimbangan Data Menggunakan SMOTE-ENN

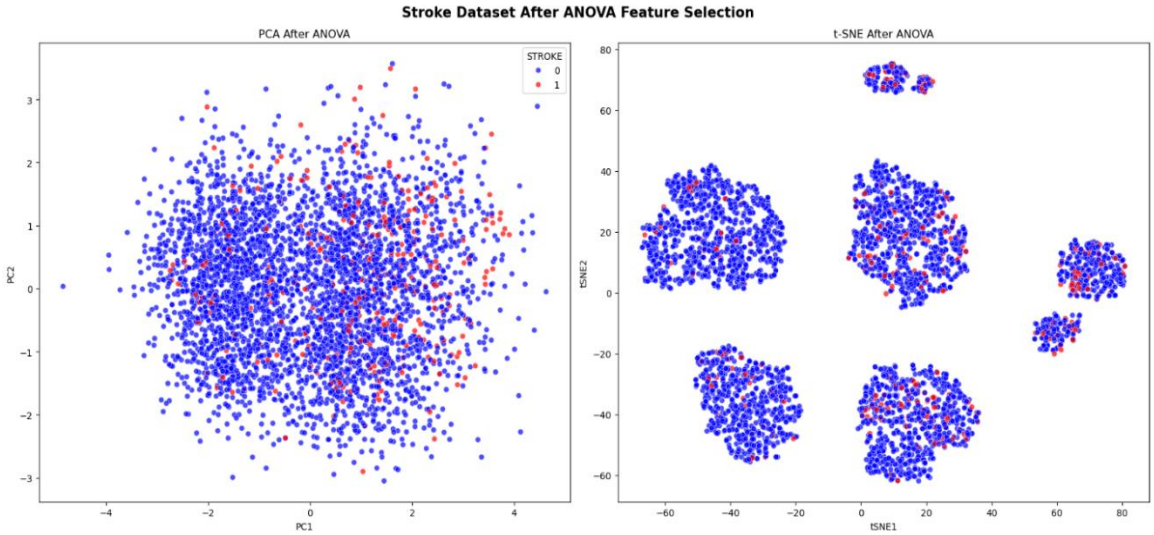
Setelah proses seleksi fitur menggunakan ANOVA F-Test selesai dilakukan, tahap berikutnya adalah penyeimbangan data menggunakan metode SMOTE-ENN. Metode ini diterapkan hanya pada data latih untuk mencegah terjadinya *data leakage* dan menjaga objektivitas proses evaluasi model. Sebelum proses penyeimbangan dilakukan, distribusi kelas pada data latih menunjukkan ketimpangan yang cukup tinggi dengan 3.154 data No Stroke dan 281 data Stroke. Setelah penerapan SMOTE-ENN, distribusi kelas berubah menjadi 2.067 data No Stroke dan 2.887 data Stroke. Hasil tersebut menunjukkan bahwa SMOTE-ENN berhasil meningkatkan representasi kelas minoritas sekaligus mengurangi data yang dianggap sebagai noise melalui mekanisme Edited Nearest Neighbour (ENN).

3.4 Visualisasi Dataset

Berdasarkan visualisasi pada **Gambar 3**, hasil *Principal Component Analysis* (PCA) menunjukkan bahwa distribusi data antara kelas Stroke dan No Stroke masih mengalami tingkat overlapping yang cukup tinggi sehingga batas pemisah kelas secara linear belum terlihat dengan jelas. Nilai explained variance ratio sebesar 26.40% pada komponen utama pertama dan 12.40% pada komponen utama kedua menghasilkan total explained variance sebesar 38.80%. Hasil tersebut menunjukkan bahwa dua komponen PCA hanya mampu merepresentasikan sebagian informasi dari keseluruhan dataset. Hal tersebut menunjukkan dataset memiliki struktur data yang kompleks dan tersebar pada banyak dimensi. Sementara itu, visualisasi menggunakan *t-distributed Stochastic Neighbor Embedding* (t-SNE) memperlihatkan pembentukan beberapa klaster data yang lebih jelas dibandingkan PCA karena mampu mempertahankan hubungan non-linear antar observasi. Meskipun demikian, distribusi kelas Stroke dan No Stroke masih menunjukkan tingkat *overlapping* yang cukup tinggi pada sebagian besar klaster. Kondisi ini mengindikasikan bahwa karakteristik kedua kelas memiliki kemiripan yang cukup tinggi sehingga proses klasifikasi menjadi lebih menantang. Tingginya *overlapping* yang terlihat pada visualisasi juga menjadi indikasi awal bahwa proses klasifikasi kemungkinan akan menghadapi tantangan dalam membedakan kedua kelas secara optimal. Hasil Visualisasi PCA dan t-SNE ditunjukkan pada **Gambar 3**.

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 251
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds



Gambar 3. Visualisasi PCA dan t-SNE

3.5 Penerapan Metode Random Forest

Setelah melalui tahap preprocessing, seleksi fitur, dan juga penyeimbangan data, tahap berikutnya dataset digunakan pada proses pemodelan Random Forest. Algoritma ini dipilih karena kemampuannya yang baik dalam menangani data dengan jumlah atribut yang banyak serta mampu mengurangi risiko *overfitting* melalui pembentukan banyak pohon keputusan (*decision tree*). Pada penelitian ini, proses validasi dilakukan menggunakan *Stratified K-Fold Cross Validation* dengan variasi nilai *k=3* hingga 20 agar distribusi kelas tetap seimbang pada proses validasi. Selain itu, optimasi *hyperparameter* dilakukan menggunakan *GridSearchCV* untuk menghasilkan kombinasi terbaik yang mampu menghasilkan performa klasifikasi secara optimal. Parameter yang diuji meliputi *n_estimators*, *max_depth*, *min_samples_leaf*, *min_samples_split*, *max_features*, dan *bootstrap*. Penentuan konfigurasi terbaik dipilih berdasarkan nilai F1-Score sebagai metrik utama karena dataset yang digunakan punya karakteristik kelas yang tidak seimbang. Pemilihan hasil terbaik berdasarkan F1-Score dinilai lebih representatif dalam kasus klasifikasi data tidak seimbang, dibandingkan dengan akurasi yang cenderung bias terhadap kelas mayoritas. Apabila terdapat lebih dari satu konfigurasi yang menghasilkan nilai F1-Score yang sama, maka pemilihan dilakukan dengan mempertimbangkan nilai recall, presisi, ROC-AUC, dan akurasi secara berurutan. Berdasarkan seluruh pengujian yang dilakukan, nilai *k=5* mendapatkan hasil terbaik pada algoritma Random Forest. Meskipun demikian, beberapa nilai k lainnya juga menghasilkan performa yang relatif mendekati sehingga pemilihan k=5 tidak dimaksudkan sebagai nilai optimal universal, melainkan konfigurasi terbaik pada dataset dan skenario penelitian ini. Oleh karena itu, hasil evaluasi yang disajikan pada subbab ini merupakan hasil terbaik yang diperoleh pada nilai *k=5*. Hasil optimasi hyperparameter terbaik ditunjukkan pada **Tabel 9**.

Tabel 9. Parameter Terbaik Random Forest

Parameter	Nilai
n_estimators	200
max_depth	None
min_samples_leaf	1
min_samples_split	2
max_features	sqrt
bootstrap	False

Selanjutnya dilakukan pengujian model menggunakan tiga variasi pembagian data. Hasil pengujian model Random Forest ditunjukkan pada **Tabel 10**.

Tabel 10. Hasil Terbaik Evaluasi Model Random Forest

Rasio	K-Fold	Akurasi	Presisi	Recall	F1-Score	ROC-AUC
70:30	k=5	78.01%	17.62%	45.74%	25.44%	69.41%
80:20	k=5	79.06%	19.75%	51.61%	28.57%	75.33%
90:10	k=5	80.89%	25.00%	67.74%	36.52%	80.32%

Berdasarkan **Tabel 10**, performa Random Forest menunjukkan peningkatan seiring bertambahnya data latih yang digunakan. Pada rasio 70:30, model menghasilkan F1-Score sebesar 25.44% dan ROC-AUC sebesar 69.41%. Ketika rasio pembagian data ditingkatkan menjadi 80:20, nilai F1-Score meningkat menjadi 28.57% dengan ROC-AUC sebesar 75.33%. Performa terbaik diperoleh pada rasio 90:10 dengan nilai akurasi sebesar 80.89%, presisi sebesar 25.00%, recall sebesar 67.74%, F1-Score sebesar 36.52%, dan ROC-AUC sebesar 80.32%. Hasil tersebut menunjukkan bahwa Random

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 252
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

Forest mampu mengidentifikasi sebagian besar kasus Stroke dengan cukup baik berdasarkan nilai recall sebesar 67.74%. Selain itu, nilai ROC-AUC sebesar 80.32% menandakan bahwa model memiliki kinerja yang baik dalam membedakan kelas Stroke dan No Stroke. Meskipun demikian, nilai presisi dan F1-Score yang masih relatif rendah menandakan bahwa model mengalami kesulitan dalam mengklasifikasikan kasus stroke secara konsisten. Hal ini diduga karena karakteristik dataset yang sangat tidak seimbang serta adanya *overlapping* antara kelas stroke dan no stroke yang terlihat pada hasil visualisasi PCA dan t-SNE. Meskipun metode SMOTE-ENN telah membantu mengurangi ketimpangan kelas pada data latih, kompleksitas distribusi data masih menjadi tantangan dalam proses klasifikasi.

3.6 Penerapan Metode XGBoost

Pada tahap berikutnya dilakukan pemodelan menggunakan algoritma XGBoost. Algoritma ini dirancang supaya meningkatkan performa prediksi serta mengurangi kesalahan klasifikasi melalui pembelajaran secara bertahap (*boosting*). Setiap pohon yang dibentuk secara bertahap akan memperbaiki kesalahan prediksi dari pohon sebelumnya sehingga model mampu mempelajari pola data yang kompleks dan non-linear. Sama seperti Random Forest, proses pelatihan XGBoost dilakukan menggunakan *Stratified K-Fold Cross Validation* dan optimasi hyperparameter menggunakan *GridSearchCV*. Parameter yang diuji meliputi *n_estimators*, *max_depth*, *learning_rate*, *colsample_bytree* dan *subsample*. Sama seperti Random Forest, algoritma XGBoost diuji menggunakan variasi nilai *K-Fold Cross Validation* mulai dari k=3 hingga k=20. Penentuan konfigurasi terbaik dilakukan menggunakan cara yang sama seperti pada Random Forest, yaitu berdasarkan nilai F1-Score sebagai metrik utama. Berdasarkan hasil pengujian, nilai k=4 menghasilkan performa terbaik dibandingkan variasi nilai k lainnya. Perbedaan performa antar nilai k relatif tidak terlalu besar, yang menunjukkan bahwa model memiliki tingkat kestabilan yang cukup baik terhadap perubahan jumlah fold pada proses validasi. Oleh karena itu, hasil evaluasi yang disajikan pada subbab ini merupakan hasil terbaik yang diperoleh pada nilai k=4. Hasil hyperparameter terbaik ditunjukkan pada **Tabel 11.**

Tabel 11. Parameter Terbaik XGBoost

Hyperparameter	Nilai
n_estimators	300
max_depth	7
learning_rate	0.1
colsample_bytree	1.0
subsample	0.8

Selanjutnya dilakukan pengujian model menggunakan tiga variasi pembagian data yaitu 70:30, 80:20, dan 90:10. Hasil pengujian model XGBoost ditunjukkan pada **Tabel 12**.

Tabel 12. Hasil Terbaik Evaluasi Model XGBoost

Rasio	K-Fold	Akurasi	Presisi	Recall	F1-Score	ROC-AUC
70:30	k=4	76.96%	17.56%	48.94%	25.84%	71.66%
80:20	k=4	81.02%	21.77%	51.61%	30.62%	76.90%
90:10	k=4	82.98%	25.71%	58.06%	35.64%	81.22%

Berdasarkan hasil evaluasi pada **Tabel 12**, performa XGBoost menunjukkan peningkatan seiring bertambahnya data latih yang digunakan. Pada rasio 70:30, model menghasilkan F1-Score sebesar 25.84% dengan ROC-AUC sebesar 71.66%. Ketika rasio pembagian data ditingkatkan menjadi 80:20, nilai F1-Score meningkat menjadi 30.62% dengan ROC-AUC sebesar 76.90%. Performa terbaik diperoleh pada rasio 90:10 dengan akurasi sebesar 82.98%, presisi sebesar 25.71%, recall sebesar 58.06%, F1-Score sebesar 35.64%, dan ROC-AUC sebesar 81.22%. Hasil tersebut menandakan bahwa XGBoost memiliki performa yang cukup baik dalam membedakan kelas Stroke dan No Stroke, yang tercermin dari nilai ROC-AUC sebesar 81.22%. Namun demikian, tingginya nilai akurasi sebesar 82.98% perlu diinterpretasikan secara hati-hati karena dataset yang digunakan memiliki distribusi kelas yang tidak seimbang. Pada kondisi tersebut, model masih dapat memperoleh akurasi tinggi akibat dominasi kelas mayoritas (*No Stroke*), sementara kemampuan mendeteksi kelas Stroke masih terbatas sebagaimana ditunjukkan oleh nilai presisi sebesar 25.71% dan F1-Score sebesar 35.64%. Oleh karena itu, evaluasi performa pada penelitian ini lebih menitikberatkan pada nilai F1-Score dan recall dibandingkan akurasi. Rendahnya nilai presisi dan F1-Score menunjukkan bahwa model masih mengalami kesulitan dalam mengidentifikasi seluruh pola pada kelas Stroke secara konsisten. Kondisi ini diduga karena ketidakseimbangan distribusi kelas serta tingginya tingkat *overlapping* antara kelas Stroke dan No Stroke yang terlihat pada hasil visualisasi PCA dan t-SNE sebelumnya.

3.7 Penerapan Metode *Ensemble Soft Voting*

Setelah diperoleh model Random Forest dan XGBoost yang telah dioptimasi, dilakukan proses penggabungan model menggunakan metode *Ensemble Soft Voting*. Metode ini bekerja dengan menggabungkan probabilitas prediksi yang dihasilkan oleh masing-masing model. Probabilitas dari Random Forest dan XGBoost kemudian dirata-ratakan untuk menghasilkan keputusan klasifikasi akhir. Pendekatan *ensemble* diterapkan untuk memanfaatkan keunggulan tiap

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 253
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

algoritma sehingga dapat meningkatkan stabilitas dan akurasi prediksi disbanding model tunggal. Pada penelitian ini, proses evaluasi Ensemble Soft Voting dilakukan menggunakan skenario pembagian data dan validasi yang sama dengan model sebelumnya agar hasil yang diperoleh dapat dibandingkan secara objektif. Penentuan konfigurasi terbaik juga dilakukan berdasarkan nilai F1-Score sebagai metrik utama. Metode *Ensemble Soft Voting* juga dievaluasi menggunakan variasi nilai *K-Fold Cross Validation* mulai dari k=3 hingga k=20. Berdasarkan seluruh pengujian yang dilakukan, nilai k=4 menghasilkan performa terbaik dibandingkan nilai k lainnya sehingga digunakan sebagai dasar analisis pada subbab ini. Hasil evaluasi metode *Ensemble Soft Voting* diperlihatkan pada **Tabel 13**.

Tabel 13. Hasil Evaluasi *Ensemble Soft Voting*

Rasio	K-Fold	Akurasi	Presisi	Recall	F1-Score	ROC-AUC
70:30	k=4	77.40%	17.39%	46.81%	25.36%	71.28%
80:20	k=4	80.24%	20.53%	50.00%	29.11%	76.81%
90:10	k=4	82.20%	25.33%	61.29%	35.85%	81.29%

Berdasarkan hasil evaluasi pada Tabel 13, performa Ensemble Soft Voting menunjukkan peningkatan seiring bertambahnya proporsi data latih yang digunakan. Pada rasio 70:30, model menghasilkan F1-Score sebesar 25.36% dengan ROC-AUC sebesar 71.28%. Ketika rasio pembagian data ditingkatkan menjadi 80:20, nilai F1-Score meningkat menjadi 29.11% dan ROC-AUC menjadi 76.81%. Performa terbaik diperoleh pada rasio 90:10 dengan akurasi 82.20%, presisi 25.33%, recall 61.29%, F1-Score 35.85%, dan ROC-AUC sebesar 81,29%. Hasil tersebut menunjukkan bahwa metode Ensemble Soft Voting mampu memberikan performa yang baik dalam membedakan kelas Stroke dan No Stroke, yang ditandai dari nilai ROC-AUC sebesar 81.29%. Selain itu, nilai recall 61.29% mengindikasikan bahwa sebagian besar kasus Stroke berhasil dikenali oleh model. Meskipun demikian, nilai presisi dan F1-Score yang rendah menandakan bahwa model masih menghadapi tantangan dalam mengidentifikasi seluruh pola pada kelas Stroke secara konsisten. Kondisi ini diduga dipengaruhi oleh tidak seimbangnya distribusi kelas serta tingginya tingkat *overlapping* antara kelas Stroke dan No Stroke yang telah terlihat pada hasil visualisasi PCA dan t-SNE sebelumnya. Jika dibandingkan dengan model tunggal, *Ensemble Soft Voting* menghasilkan performa yang cukup baik dengan nilai ROC-AUC paling tinggi di antara semua model yang diuji. Namun, peningkatan performa dibandingkan Random Forest dan XGBoost relatif kecil, terutama pada nilai F1-Score yang tetap berada di bawah Random Forest. Hal ini menunjukkan bahwa kedua model dasar menghasilkan pola prediksi yang relatif serupa sehingga proses penggabungan probabilitas melalui Soft Voting belum mampu menghasilkan variasi prediksi yang cukup untuk meningkatkan performa klasifikasi secara signifikan. Akibatnya, kesalahan klasifikasi yang terjadi pada kedua model masih cenderung terjadi pada observasi yang sama sehingga peningkatan performa secara keseluruhan menjadi terbatas. Indikasi tersebut terlihat dari nilai F1-Score ketiga model yang relatif berdekatan, yaitu 36.52% pada Random Forest, 35.64% pada XGBoost, dan 35.85% pada Ensemble Soft Voting. Kedekatan performa tersebut menunjukkan bahwa kedua model dasar cenderung mempelajari pola keputusan yang serupa sehingga manfaat diversitas model yang umumnya menjadi keunggulan pendekatan ensemble tidak terbentuk secara optimal.

3.8 Perbandingan dan Evaluasi Performa Model

Setelah dilakukan proses pelatihan model dan evaluasi model menggunakan Random Forest, XGBoost, dan *Ensemble Soft Voting*, dilakukan perbandingan performa untuk mengidentifikasi model yang memiliki kinerja terbaik dalam klasifikasi penyakit stroke. Perbandingan dilakukan berdasarkan hasil dari metrik akurasi, presisi, recall, F1-score, dan juga ROC-AUC terbaik dari tiap model. Perbandingan performa terbaik dari tiap model ditunjukkan pada **Tabel 14**.

Tabel 14. Perbandingan Performa Model Terbaik

Model	Akurasi	Presisi	Recall	F1-Score	ROC-AUC
Random Forest	80.89%	25.00%	67.74%	36.52%	80.32%
XGBoost	82.98%	25.71%	58.06%	35.64%	81.22%
Ensemble Soft Voting	82.20%	25.33%	61.29%	35.85%	81.29%

Berdasarkan **Tabel 14**, setiap model menunjukkan keunggulan pada metrik yang berbeda. Algoritma XGBoost memperoleh nilai akurasi tertinggi sebesar 82,98%. Akan tetapi, tingginya akurasi tidak secara langsung menunjukkan kemampuan klasifikasi yang lebih baik karena dataset memiliki distribusi kelas yang tidak seimbang. Oleh karena itu, penelitian ini menggunakan F1-Score sebagai metrik utama dan recall sebagai pertimbangan tambahan, karena kedua metrik tersebut lebih representatif dalam mengevaluasi kemampuan model mengenali kelas Stroke sebagai kelas minoritas. Sementara itu, Random Forest menghasilkan nilai recall dan juga F1-Score tertinggi, yaitu sebesar 67.74% dan 36.52%. Adapun nilai ROC-AUC tertinggi diperoleh oleh Ensemble Soft Voting sebesar 81.29%. Pada penelitian ini, model terbaik dipilih berdasarkan nilai F1-Score karena dataset yang digunakan pada studi ini memiliki distribusi kelas tidak seimbang. Selain itu, nilai recall turut menjadi pertimbangan penting karena penelitian berfokus pada identifikasi kasus Stroke sebagai kelas minoritas. Berdasarkan kedua metrik tersebut, Random Forest menunjukkan performa terbaik dengan F1-Score sebesar 36.52% dan recall sebesar 67.74%. Meskipun Ensemble Soft Voting menghasilkan nilai ROC-AUC tertinggi dan XGBoost memperoleh akurasi yang lebih besar, selisih performa antar model relatif kecil. Selain itu, pengujian menggunakan variasi nilai K-Fold dari 3 hingga 20 menunjukkan bahwa performa ketiga model tidak

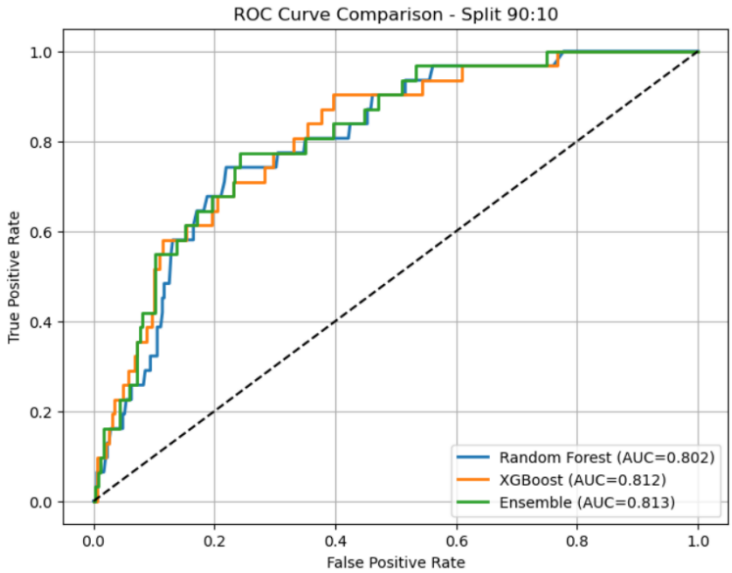
© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 254
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

mengalami fluktuasi yang ekstrem pada berbagai nilai k yang diuji. Temuan ini mengindikasikan bahwa performa model relatif stabil terhadap perubahan jumlah fold sehingga pemilihan nilai k terbaik pada penelitian ini dilakukan untuk mengidentifikasi konfigurasi yang paling sesuai pada dataset yang digunakan, bukan untuk menentukan nilai optimal yang bersifat universal. Hasil ini menunjukkan bahwa ketiga model memiliki performa yang cukup sebanding dalam mempelajari pola data stroke. Hal tersebut menunjukkan bahwa proses penggabungan probabilitas melalui Soft Voting belum mampu memberikan variasi prediksi yang cukup besar karena kedua model dasar masih menghasilkan pola kesalahan klasifikasi yang relatif serupa. Namun demikian, Random Forest dinilai lebih sesuai untuk penelitian ini karena mampu memberikan keseimbangan yang lebih baik antara presisi dan juga recall, yang diindikasikan dari nilai F1-Score tertinggi yang berhasil dicapai.

Berdasarkan seluruh hasil pengujian yang dilakukan, Random Forest terpilih sebagai model terbaik pada penelitian ini karena menghasilkan nilai F1-Score dan recall tertinggi. Meskipun demikian, perbedaan performa antar model tidak terlalu besar, Random Forest, XGBoost, dan Ensemble Soft Voting menunjukkan kemampuan yang relatif sebanding dalam mempelajari pola pada dataset ini. Namun, rendahnya nilai presisi dan F1-Score pada seluruh model menunjukkan bahwa tantangan utama penelitian bukan berasal dari pemilihan algoritma, melainkan dari karakteristik dataset yang memiliki distribusi kelas tidak seimbang dan tingkat *overlapping* yang tinggi. Tantangan utama yang masih dihadapi seluruh model diduga berasal dari ketidakseimbangan distribusi kelas dan tingginya tingkat *overlapping* antar kelas yang menyebabkan proses pemisahan data menjadi lebih kompleks.

Berdasarkan **Gambar 4**, kurva ROC yang dihasilkan dari pengujian pada rasio pembagian data 90:10 dan k=4, yang merupakan skenario pengujian dengan nilai ROC-AUC terbaik, mengindikasikan bahwa ketiga model memiliki performa yang baik dalam membedakan kelas Stroke dan No Stroke. Hal ini bisa dilihat dari nilai ROC-AUC yang seluruhnya berada di atas 0.80. *Ensemble Soft Voting* memperoleh nilai ROC-AUC tertinggi sebesar 0.813, diikuti oleh XGBoost sebesar 0.812 dan Random Forest sebesar 0.802.



Gambar 4. Kurva ROC Perbandingan Tiap Model

Perbedaan nilai ROC-AUC antar model relatif kecil, mengindikasikan bahwa ketiganya memiliki kemampuan klasifikasi yang hampir setara. Meskipun demikian, penggabungan model melalui metode Ensemble Soft Voting mampu menghasilkan kemampuan pemisahan kelas yang sedikit lebih baik daripada model tunggal. Hal ini menunjukkan bahwa kombinasi probabilitas prediksi dari XGBoost dan Random Forest dapat memberikan peningkatan performa, meskipun peningkatan yang diperoleh tidak terlalu signifikan. Hasil tersebut juga memperkuat temuan pada Tabel 14 yang menunjukkan bahwa setiap model memiliki keunggulan pada metrik yang berbeda. Random Forest menghasilkan nilai F1-Score dan recall tertinggi, XGBoost memperoleh akurasi dan presisi terbaik, sedangkan Ensemble Soft Voting menghasilkan nilai ROC-AUC tertinggi. Dengan demikian, ketiga model menunjukkan performa yang relatif sebanding dalam mengklasifikasikan data stroke.

Meskipun nilai ROC-AUC yang diperoleh tergolong baik, nilai presisi dan F1-Score yang masih relatif rendah mengindikasikan bahwa proses klasifikasi belum sepenuhnya optimal. Kondisi ini diduga dipengaruhi oleh ketidakseimbangan distribusi kelas serta tingginya tingkat *overlapping* antara kelas Stroke dan No Stroke yang telah ditunjukkan pada hasil visualisasi PCA dan t-SNE sebelumnya. Akibatnya, sebagian data masih sulit dipisahkan secara tegas sehingga memengaruhi performa klasifikasi yang dihasilkan.

3.9 Pembahasan

Berdasarkan hasil pengujian, model Random Forest, XGBoost, dan Ensemble Soft Voting, memperlihatkan kemampuan yang cukup baik dalam mengklasifikasikan data stroke. Hal ini terlihat dari nilai ROC-AUC seluruh model yang cukup

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 255
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Data Science
ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

tinggi di atas 80%, yang mengindikasikan bahwa model mampu membedakan data Stroke dan No Stroke dengan cukup baik. Namun, nilai presisi dan F1-Score yang rendah menandakan bahwa masih terdapat sejumlah kasus yang belum dapat diklasifikasikan secara optimal, khususnya pada data kelas Stroke yang jumlahnya lebih sedikit daripada kelas No Stroke.

Hasil visualisasi menggunakan PCA dan t-SNE menunjukkan adanya tingkat *overlapping* yang cukup tinggi antara kedua kelas. Kondisi ini menunjukkan bahwa karakteristik pasien Stroke dan No Stroke pada dataset memiliki kemiripan yang cukup besar sehingga batas pemisah antar kelas menjadi sulit dibentuk secara jelas. Akibatnya, model masih mengalami kesulitan dalam mengenali seluruh pola yang membedakan kedua kelas tersebut.

Tingginya overlapping yang terlihat pada visualisasi PCA maupun t-SNE menunjukkan bahwa atribut yang digunakan belum cukup mampu membantu model untuk membentuk batas keputusan yang jelas antara kelas Stroke dan No Stroke. Meskipun atribut yang dipilih memiliki relevansi klinis, representasi fitur yang tersedia masih belum cukup mampu untuk memisahkan kedua kelas secara optimal. Kondisi ini menjelaskan mengapa seluruh model masih menghasilkan nilai presisi dan F1-Score yang relatif rendah meskipun nilai ROC-AUC berada di atas 80%.

Penerapan metode SMOTE-ENN terbukti membantu mengurangi masalah ketidakseimbangan kelas pada data latih sehingga kemampuan model dalam mengenali kelas Stroke menjadi lebih baik. Selain itu, proses seleksi fitur menggunakan ANOVA F-Test juga berhasil mengidentifikasi atribut yang memiliki hubungan paling kuat terhadap variabel target. Hasil seleksi menunjukkan bahwa AGE, SYSBP, PREVHYP, BPMEDS, dan DIABETES merupakan atribut yang paling berpengaruh dalam proses klasifikasi. Temuan ini sejalan dengan berbagai penelitian medis yang menyatakan bahwa usia, hipertensi, dan diabetes merupakan faktor risiko penting yang berhubungan dengan kejadian stroke.

Meskipun metode SMOTE-ENN berhasil meningkatkan representasi kelas minoritas pada data latih, peningkatan tersebut belum mampu mengatasi tingginya overlapping antar kelas. Hal ini menunjukkan bahwa permasalahan utama pada dataset bukan hanya ketidakseimbangan distribusi kelas, tetapi juga kemiripan karakteristik antara pasien Stroke dan No Stroke. Akibatnya, sampel sintetis yang dihasilkan SMOTE masih berada pada wilayah fitur yang saling tumpang tindih sehingga proses pembersihan menggunakan ENN belum mampu membentuk batas keputusan yang lebih jelas. Kondisi tersebut menyebabkan peningkatan performa model setelah proses resampling menjadi relatif terbatas.

Berdasarkan hasil perbandingan model, Random Forest menghasilkan nilai recall dan F1-Score tertinggi, sedangkan XGBoost memperoleh nilai akurasi dan presisi tertinggi. Di sisi lain, Ensemble Soft Voting menghasilkan nilai ROC-AUC tertinggi dibandingkan model lainnya. Namun, perbedaan performa antar model relatif kecil. Temuan ini menunjukkan bahwa pemilihan algoritma bukan merupakan faktor utama yang menentukan keberhasilan klasifikasi pada dataset Framingham. Karakteristik data berupa ketidakseimbangan kelas dan tingginya overlapping antar kelas justru memberikan pengaruh yang lebih besar terhadap performa seluruh model yang diuji..

Hasil penelitian juga menunjukkan bahwa peningkatan kompleksitas model tidak secara otomatis meningkatkan performa klasifikasi stroke. Meskipun XGBoost dan Ensemble Soft Voting memiliki mekanisme pembelajaran yang lebih kompleks dibandingkan Random Forest, peningkatan performa yang diperoleh relatif kecil. Temuan ini mengindikasikan bahwa keterbatasan utama penelitian bukan terletak pada algoritma yang digunakan, melainkan pada karakteristik dataset yang memiliki tingkat overlapping tinggi serta representasi fitur yang belum cukup mampu untuk membedakan kelas. Oleh karena itu, peningkatan kualitas feature engineering diperkirakan akan memberikan dampak yang lebih besar dibandingkan hanya meningkatkan kompleksitas model klasifikasi.

3.10 Keterbatasan Penelitian

Penelitian ini memiliki keterbatasan yang memengaruhi performa model. Meskipun metode SMOTE-ENN telah diterapkan, metode tersebut belum mampu mengatasi tingginya *overlapping* antara kelas Stroke dan No Stroke. Kondisi ini menunjukkan bahwa permasalahan utama pada dataset tidak hanya disebabkan oleh distribusi kelas yang tidak seimbang, tetapi juga oleh kemiripan karakteristik antar kelas sehingga proses resampling belum mampu membentuk batas keputusan yang lebih jelas. Selain itu, penelitian ini hanya memanfaatkan atribut yang tersedia pada dataset Framingham Heart Study tanpa melakukan proses feature engineering yang lebih mendalam, sehingga representasi fitur yang digunakan masih belum mampu membedakan kedua kelas secara optimal. Penelitian ini juga menggunakan ANOVA F-Test sebagai tahap seleksi fitur statistik untuk menyaring atribut yang telah dipilih berdasarkan literatur medis.. Meskipun metode tersebut efektif untuk mengurangi fitur yang memiliki kontribusi statistik rendah, ANOVA F-Test hanya mengevaluasi hubungan linear antara fitur dan variabel target sehingga belum mampu menangkap hubungan non-linear maupun interaksi kompleks antar fitur yang berpotensi dimanfaatkan oleh algoritma berbasis pohon seperti Random Forest dan XGBoost. Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan feature engineering yang menghasilkan fitur yang lebih mampu membedakan kelas Stroke dan No Stroke, mengevaluasi metode seleksi fitur berbasis model seperti Boruta atau *Recursive Feature Elimination* (RFE), serta membandingkan metode penanganan ketidakseimbangan kelas seperti Borderline-SMOTE, ADASYN, maupun *cost-sensitive learning*.

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 256
This work is licensed under a Creative Commons Attribution 4.0 International License.

ISSN 2745-7966 (Media Online)
Vol 5, No 3, June 2026 | Page 244–258
DOI : 10.47065/bulletinds.v5i3.10306
https://ejurnal.seminar-id.com/index.php/bulletinds

4. KESIMPULAN

Penelitian ini menganalisis kinerja Random Forest, XGBoost, dan Ensemble Soft Voting untuk klasifikasi stroke pada dataset Framingham Heart Study yang memiliki karakteristik kelas tidak seimbang. Berdasarkan evaluasi menggunakan F1-Score sebagai metrik utama dan recall sebagai pertimbangan tambahan, Random Forest memberikan keseimbangan performa terbaik dengan F1-Score sebesar 36.52% dan recall sebesar 67.74%, sedangkan Ensemble Soft Voting hanya memberikan peningkatan pada nilai ROC-AUC tanpa mampu meningkatkan F1-Score secara signifikan dibandingkan model tunggal. Hasil penelitian menunjukkan bahwa peningkatan kompleksitas model melalui pendekatan Ensemble Soft Voting tidak secara otomatis meningkatkan performa klasifikasi. Analisis visualisasi PCA dan t-SNE memperlihatkan tingkat overlapping yang tinggi antara kelas Stroke dan No Stroke, sehingga karakteristik kedua kelas sulit dipisahkan oleh model klasifikasi. Selain itu, meskipun SMOTE-ENN berhasil meningkatkan representasi kelas minoritas, metode tersebut belum mampu mengatasi kemiripan distribusi antar kelas sehingga peningkatan performa model menjadi terbatas. Temuan ini mengindikasikan bahwa karakteristik dataset, khususnya ketidakseimbangan kelas dan tingginya overlapping antar kelas, memiliki pengaruh yang lebih dominan terhadap performa klasifikasi dibandingkan pemilihan algoritma maupun peningkatan kompleksitas model melalui pendekatan ensemble. Berdasarkan hasil tersebut, penelitian selanjutnya disarankan untuk mengembangkan feature engineering yang lebih baik, mengeksplorasi metode feature selection berbasis model seperti Boruta atau *Recursive Feature Elimination* (RFE), menerapkan metode penanganan ketidakseimbangan kelas seperti Borderline-SMOTE, ADASYN, atau *cost-sensitive learning*, serta mengembangkan pendekatan ensemble yang memanfaatkan model dengan karakteristik prediksi yang lebih beragam agar mampu meningkatkan kemampuan deteksi kasus Stroke.

REFERENCES

[1] L. Pu, L. Wang, R. Zhang, T. Zhao, Y. Jiang, and L. Han, "Projected Global Trends in Ischemic Stroke Incidence, Deaths and Disability-Adjusted Life Years From 2020 to 2030," *Stroke*, vol. 54, no. 5, pp. 1330–1339, May 2023, doi: 10.1161/STROKEAHA.122.040073/SUPPL_FILE/STROKE-2022-040073FINAL_ELECTRONIC.

[2] V. Saini, L. Guada, and D. R. Yavagal, "Global Epidemiology of Stroke and Access to Acute Ischemic Stroke Interventions," *Neurology*, vol. 97, no. 20_Supplement_2, pp. S6–S16, Nov. 2021, doi: 10.1212/WNL.0000000000012781.

[3] "World Stroke Organization." Accessed: Jun. 03, 2026. [Online]. Available: https://www.world-stroke.org/world-stroke-day-campaign/about-stroke/impact-of-stroke?

[4] S. J. Murphy and D. J. Werring, "Stroke: causes and clinical features," *Medicine*, vol. 48, no. 9, pp. 561–566, Sep. 2020, doi: 10.1016/J.MPMED.2020.06.002.

[5] N. Biswas, K. M. M. Uddin, S. T. Rikta, and S. K. Dey, "A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach," *Healthcare Analytics*, vol. 2, p. 100116, Nov. 2022, doi: 10.1016/J.HEALTH.2022.100116.

[6] T. Gustin Rahayu and U. Faletehan, "The Analysis of Stroke Risk Factors and Stroke Types," *Faletehan Health Journal*, vol. 10, no. 01, pp. 48–53, Mar. 2023, doi: 10.33746/FHJ.V10I01.410.

[7] T. H. Lee *et al.*, "A Systematic Search and Review of Registered Pharmacological Therapies Investigated to Improve Outcomes after a Stroke," *Cerebrovasc. Dis. Extra*, vol. 14, no. 1, pp. 158–164, Oct. 2024, doi: 10.1159/000541703.

[8] A. Arnalich-Montiel, A. Burgos-Santamaría, L. Pazó-Sayós, and B. Quintana-Villamandos, "Comprehensive Management of Stroke: From Mechanisms to Therapeutic Approaches," *International Journal of Molecular Sciences 2024, Vol. 25, Page 5252*, vol. 25, no. 10, p. 5252, May 2024, doi: 10.3390/IJMS25105252.

[9] D. T. Murdiansyah, "Prediksi Stroke Menggunakan Extreme Gradient Boosting," *JIKO (Jurnal Informatika dan Komputer)*, vol. 8, no. 2, pp. 419–426, Sep. 2024, doi: 10.26798/JIKO.V8I2.1295.

[10] Ethem. Alpaydin, "Introduction to machine learning," p. 682, 2020, Accessed: Apr. 16, 2026. [Online]. Available: https://books.google.com/books/about/Introduction_to_Machine_Learning_fourth.html?hl=id&id=uZnSDwAAQBAJ

[11] R. Aggrey, B. Ansah Adjei, N. Adwoa Konadu Dsane, K. Osei Afoduo, and L. Naa Oye Holdbrooke, "The Role of Artificial Intelligence and Machine Learning in Healthcare." doi: 10.36948/ijfmr.2024.v06i06.33477.

[12] R. Kharrisa Andini *et al.*, "Application of ADASYN and Bayesian Optimization to Random Forests for Cervical Cancer Classification," *SISFO Jurnal Ilmiah*, vol. 10, no. 1, pp. 30–39, 2026, doi: 10.29103/sisfo.

[13] F. Adha, H. Airi, T. Suprapti, and A. Bahtiar, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," *E-Link: Jurnal Teknik Elektro dan Informatika*, vol. 18, no. 1, pp. 73–79, May 2023, doi: 10.30587/E-LINK.V18I1.5271.

[14] R. M. Saputra *et al.*, "Improving Cervical Cancer Classification Using ADASYN and Random Forest with GridSearchCV Optimization," *Infotekmesin*, vol. 16, no. 1, pp. 59–67, Jan. 2025, doi: 10.35970/INFOTEKMESIN.V16I1.2552.

[15] M. Fadli and R. A. Saputra, "Klasifikasi dan Evaluasi Performa Model Random Forest untuk Prediksi Stroke," *Jurnal Teknik*, vol. 12, no. 2, Oct. 2023, doi: 10.31000/JT.V12I2.9099.

[16] A. Ramadhani, I. Afrianty, E. Budianita, and S. K. Gusti, "Implementation of XGBoost Ensemble and Support Vector Machine For Gender Classification of Skull Bones," *Bulletin of Informatics and Data Science*, vol. 4, no. 1, pp. 34–41, May 2025, doi: 10.61944/BIDS.V4I1.115.

[17] L. R. Sitompul, A. A. Nababan, M. L. Manihuruk, W. A. Ponsen, and S. Supriyandi, "Comparison of Xgboost, Random Forest and Logistic Regression Algorithms in Stroke Disease Classification," *Sinkron : jurnal dan penelitian teknik informatika*, vol. 9, no. 2, pp. 957–968, Jun. 2025, doi: 10.33395/SINKRON.V9I2.14794.

[18] D.- Andriansyah and E. W. Fridayanthie, "Optimization of Support Vector Machine and XGBoost Methods Using Feature Selection to Improve Classification Performance," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 6, no. 2, pp. 484–493, Jan. 2023, doi: 10.31289/JITE.V6I2.8373.

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 257
This work is licensed under a Creative Commons Attribution 4.0 International License.

[19] M. Ravly Andryan *et al.*, "Komparasi Kinerja Algoritma XGBoost dan Support Vector Machine (SVM) untuk Diagnosis Penyakit Kanker Payudara," *JIKO (Jurnal Informatika dan Komputer)*, vol. 6, no. 1, pp. 1–5, Feb. 2022, doi: 10.26798/JIKO.V6I1.500.
[20] G. Almuzadid and E. R. Subhiyakto, "Stroke Risk Classification Using the Ensemble Learning Method of XGBoost and Random Forest," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 828–837, Jun. 2025, doi: 10.30871/JAIC.V9I3.9528.
[21] S. G. Kwak, "LASSO Regression Analysis: Applications in Dyslipidemia and Cardiovascular Disease Research," *J. Lipid Atheroscler.*, vol. 14, no. 3, p. 289, Sep. 2025, doi: 10.12997/JLA.2025.14.3.289.
[22] "Stroke." Accessed: Jun. 13, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/stroke?
[23] Y. X. Liu *et al.*, "Global burden of disease analysis of stroke due to high systolic blood pressure in different regions, 1990–2021," *BMC Public Health 2025 25:1*, vol. 25, no. 1, pp. 3743-, Nov. 2025, doi: 10.1186/S12889-025-25026-8.
[24] M. M. Kurniawan and B. N. Sari, "Analisis Komparatif Strategi Imputasi Nilai Hilang Pada Dataset Hepatitis UCI Menggunakan XGBoost," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 14, no. 2, pp. 2830–7062, Apr. 2026, doi: 10.23960/JITET.V14I2.9163.
[25] V. Tummalapalli, "Outlier Detection & Treatment for Machine Learning Models", doi: 10.5281/ZENODO.16500050.
[26] H. Nasiri and S. A. Alavi, "A novel framework based on deep learning and ANOVA feature selection method for diagnosis of COVID-19 cases from chest X-ray Images," *Comput. Intell. Neurosci.*, vol. 2022, Sep. 2021, doi: 10.1155/2022/4694567.
[27] I. Riantika, B. Sartono, and K. Anwar Notodiputro, "Effectiveness of SMOTE-ENN to Reduce Complexity in Classification Model," *Indonesian Journal of Statistics and Its Applications*, vol. 8, no. 1, pp. 70–82, Jun. 2024, doi: 10.29244/IJSA.V8I1P70-82.
[28] T. Abedin, H. Xu, and S. Uddin, "The impact of K selection in K-fold cross-validation on bias and variance in supervised learning models," *Scientific Reports 2026 16:1*, vol. 16, no. 1, pp. 6084-, Jan. 2026, doi: 10.1038/s41598-026-37247-x.
[29] R. Muzayanah, D. Ananda, A. Pertiwi, M. Ali, and A. Muslim, "Comparison of gridsearchcv and bayesian hyperparameter optimization in random forest algorithm for diabetes prediction," *Journal of Soft Computing Exploration*, vol. 5, no. 1, pp. 86–91, Apr. 2024, doi: 10.52465/JOSCEX.V5I1.308.
[30] A. A. Syahputra and R. E. Saputro, "Application of the XGBoost Model with Hyperparameter Tuning for Industry Classification for Job Applicants," *Sinkron : jurnal dan penelitian teknik informatika*, vol. 8, no. 3, pp. 1920–1931, Jul. 2024, doi: 10.33395/SINKRON.V8I3.13840.
[31] Sugiarto *et al.*, "Optimizing The XGBoost Model with Grid Search Hyperparameter Tuning for Maximum Temperature Forecasting," *Journal of Applied Data Sciences*, vol. 6, no. 4, pp. 2517–2529, Sep. 2025, doi: 10.47738/JADS.V6I4.885.
[32] F. Q. Aina and Z. Fatah, "Penerapan Data Mining untuk Kelulusan Al-Barqi di Pesantren Salafiyah Syafi'iyah Menggunakan Metode Random Forest," *Jurnal Penelitian Rumpun Ilmu Teknik*, vol. 4, no. 3, pp. 202–210, Aug. 2025, doi: 10.55606/juprit.v4i3.5625.
[33] M. R. Kusuma and A. A. Kurniawan, "Penerapan Teknik Ensemble Learning untuk Deteksi Dini Penyakit Jantung Menggunakan Metode Voting Classifier dan Stacking Classifier," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 30, no. 2, pp. 181–197, Sep. 2025, doi: 10.35760/tr.2025.v30i2.14826.
[34] R. Zizilia *et al.*, "Lung Cancer Classification Using the Extreme Gradient Boosting (XGBoost) Algorithm and Mutual Information for Feature Selection," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 5, pp. 2198–2214, Sep. 2025, doi: 10.32520/stmsi.v14i5.5345.
[35] M. Azim Mim, N. Majadi, and P. Mazumder, "A soft voting ensemble learning approach for credit card fraud detection," *Heliyon*, vol. 10, no. 3, p. e25466, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25466.
[36] P. Chithuloori and J. M. Kim, "Soft voting ensemble classifier for liquefaction prediction based on SPT data," *Artificial Intelligence Review 2025 58:8*, vol. 58, no. 8, pp. 228-, May 2025, doi: 10.1007/S10462-025-11230-W.

© 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 258
This work is licensed under a Creative Commons Attribution 4.0 International License.