

Komparasi Random Forest dan XGBoost Berbasis SMOTE dalam Penanganan Ketidakseimbangan Data untuk Prediksi Risiko Stroke

Comparison of Random Forest and SMOTE-Based XGBoost in Handling Data Imbalance for Stroke Risk Prediction

Chairunnisa Desti Arzety¹, Bayu Wijaya Putra^{2,*}, M Rudi Sanjaya², Dwi Rosa Indah¹

¹ Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia

² Fakultas Ilmu Komputer, Program Studi Manajemen Informatika, Universitas Sriwijaya, Palembang, Indonesia

Email: ¹chairunnisadesti12@gmail.com, ^{2,*}bayuwisata@gmail.com, ²m.rudi.sjy@ilkom.unsri.ac.id, ¹indah812@gmail.com

Email Penulis Korespondensi: bayuwisata@gmail.com

Received: 15/07/2026, Revised: 03/08/2026, Accepted: 01/09/2026, Published: 22/09/2026

Abstrak—Ketidakseimbangan distribusi kelas menjadi salah satu kendala utama dalam membangun model *machine learning* untuk memprediksi stroke, sebab kondisi ini dapat membuat model lebih dominan mempelajari karakteristik kelas mayoritas ketimbang kelas minoritas. Penelitian ini ditujukan untuk menguji perbandingan sekaligus mengevaluasi kinerja antara algoritma *Random Forest* serta *XGBoost* dalam memprediksi risiko stroke, dengan menerapkan SMOTE sebagai teknik penyeimbangan data. Penelitian menggunakan *Kaggle Stroke Prediction Dataset* yang terdiri atas 5.110 data pasien. Tahapan penelitian meliputi *data cleaning*, representasi fitur kategorikal menggunakan TF-IDF, normalisasi fitur numerik melalui *Standard Scaling*, pembagian data dengan teknik *Stratified Train-Test Split*, penerapan SMOTE pada data latih, serta pembangunan model dengan algoritma *Random Forest* dan *XGBoost*. Kinerja model dievaluasi menggunakan sejumlah metrik, yakni *Accuracy*, *Precision*, *Recall*, *F1-Score*, *ROC-AUC*, *Confusion Matrix*, dan *Feature Importance*. Hasil pengujian memperlihatkan bahwa *XGBoost* menghasilkan *Recall* 0,1600, *F1-Score* sebesar 0,1702, serta *ROC-AUC* pada angka 0,7877, lebih tinggi dibandingkan *Random Forest* yang memperoleh *Recall* 0,0800, *F1-Score* 0,1111, dan *ROC-AUC* 0,7068. Meskipun menunjukkan peningkatan dibanding *Random Forest*, nilai *Recall* dan *F1-Score* yang masih tergolong rendah mengindikasikan kemampuan mendeteksi kasus stroke belum optimal sehingga masih memerlukan pengembangan lebih lanjut. Selain itu, penerapan SMOTE digunakan untuk menyeimbangkan distribusi kelas pada data latih sehingga proses pembelajaran model menjadi lebih memperhatikan kelas minoritas. Penelitian ini memberikan kontribusi berupa penerapan *pipeline preprocessing* yang konsisten, meliputi *data cleaning*, *Standard Scaling*, representasi fitur kategorikal menggunakan TF-IDF, *Stratified Train-Test Split*, dan SMOTE pada data latih, sehingga perbandingan performa *Random Forest* dan *XGBoost* dapat dilakukan secara lebih objektif melalui penggunaan metrik evaluasi yang komprehensif pada dataset prediksi stroke yang mengalami *class imbalance*.

Kata Kunci: Stroke; Machine Learning; Random Forest; XGBoost; SMOTE; Class Imbalance

Abstract—Class imbalance is one of the main obstacles in building machine learning models for stroke prediction, as this condition tends to make models learn the characteristics of the majority class more dominantly than those of the minority class. This study aims to examine the comparison and evaluate the performance between the Random Forest and XGBoost algorithms in predicting stroke risk, by applying SMOTE as a data balancing technique. The study uses the Kaggle Stroke Prediction Dataset, consisting of 5,110 patient records. The research stages include data cleaning, categorical feature representation using TF-IDF, numerical feature normalization through Standard Scaling, data splitting using the Stratified Train-Test Split technique, application of SMOTE on the training data, and model development using the Random Forest and XGBoost algorithms. Model performance was evaluated using several metrics, namely Accuracy, Precision, Recall, F1-Score, ROC-AUC, Confusion Matrix, and Feature Importance. The test results show that XGBoost achieved a Recall of 0.1600, an F1-Score of 0.1702, and an ROC-AUC of 0.7877—all higher than those of Random Forest, which achieved a Recall of 0.0800, an F1-Score of 0.1111, and an ROC-AUC of 0.7068. Although these results represent an improvement over Random Forest, the still relatively low Recall and F1-Score values indicate that the model's ability to detect stroke cases is not yet optimal and therefore requires further development. Furthermore, SMOTE was applied to balance the class distribution in the training data so that the model's learning process would give greater attention to the minority class. This study contributes by implementing a consistent preprocessing pipeline, including data cleaning, Standard Scaling, TF-IDF-based categorical feature representation, Stratified Train-Test Split, and SMOTE on the training data, enabling a more objective comparison of Random Forest and XGBoost through the use of comprehensive evaluation metrics on an imbalanced stroke prediction dataset.

Keywords: Stroke; Machine Learning; Random Forest; XGBoost; SMOTE; Class Imbalance

1. PENDAHULUAN

Stroke masih menjadi salah satu penyakit yang tidak menular dan memiliki kontribusi besar terhadap tingginya rate kematian serta banyaknya kasus disabilitas jangka panjang di banyak negara. Efeknya tidak hanya terbatas pada meningkatnya angka kematian, tetapi juga mencakup penurunan kualitas hidup karena disabilitas yang bersifat permanen, biaya pengobatan dan rehabilitasi yang tinggi, serta berkurangnya produktivitas pasien setelah mereka mengalami stroke [1]. Faktor-faktor yang saling berkaitan mempengaruhi kemungkinan terjadinya stroke, termasuk usia, hipertensi, gangguan jantung, kadar glukosa darah, *Body Mass Index*, kebiasaan merokok, serta ciri-ciri demografis seseorang [2], [3]. Kompleksitas hubungan antara berbagai faktor risiko ini membuat sulit untuk mengidentifikasi individu yang mungkin akan mengalami stroke jika hanya menggunakan metode statistik konvensional, yang biasanya menganggap adanya hubungan linier antar variabel. Keterbatasan ini bisa menghalangi

deteksi dini, sehingga bisa menghambat pasien dengan risiko tinggi mendapatkan perawatan medis yang sesuai pada waktu yang tepat. Dengan demikian, diperlukan pendekatan prediktif yang mampu menangkap hubungan kompleks ini agar dapat mengidentifikasi risiko stroke dengan lebih akurat dan membantu dalam proses pengambilan keputusan klinis [1], [4].

Perkembangan *Machine Learning* telah memberikan alternatif yang menjanjikan dalam membangun model prediksi stroke berbasis data medis. Dibandingkan metode statistik konvensional, algoritma *Machine Learning* mampu mempelajari hubungan nonlinier antar variabel dan mengolah data berdimensi tinggi sehingga lebih adaptif terhadap karakteristik data klinis yang kompleks [5], [6]. Berbagai penelitian menunjukkan bahwa pemanfaatan data demografis, riwayat penyakit, indikator klinis, dan faktor gaya hidup dapat meningkatkan kemampuan model dalam mengidentifikasi pasien berisiko stroke [1], [2], [4]. Selain meningkatkan performa prediksi, penelitian terkini juga menekankan pentingnya interpretabilitas model melalui penerapan *Explainable Artificial Intelligence* (XAI) dan analisis *feature importance* agar hasil klasifikasi lebih transparan dan dapat mendukung pengambilan keputusan klinis [7], [8], [9].

Meskipun riset mengenai pemanfaatan *Machine Learning* untuk memprediksi risiko stroke terus berkembang, proses pembangunan modelnya masih dibayangi persoalan mendasar, yakni komposisi kelas data yang timpang. Kondisi ini lazim dijumpai pada dataset prediksi stroke, dengan jumlah pasien yang terkena stroke jauh lebih sedikit daripada pasien yang tidak mengalaminya. Akibat ketimpangan ini, algoritma cenderung lebih dominan mempelajari ciri-ciri dari kelas mayoritas, sehingga kemampuannya mengenali pola pada kelas minoritas menjadi terbatas [10], [11]. Kondisi serupa juga ditemukan pada *Kaggle Stroke Prediction Dataset* yang dimanfaatkan dalam penelitian ini. Dataset tersebut memuat 5.110 data pasien, yang terdiri atas 4.861 data non-stroke dan 249 data stroke, dengan rasio ketidakseimbangan mencapai sekitar 19,52:1. Akibatnya, nilai *accuracy* yang diperoleh model dapat terlihat tinggi, tetapi belum mampu menggambarkan performa sebenarnya dalam mendeteksi pasien yang memiliki risiko stroke. Dalam konteks layanan kesehatan, keadaan ini berpotensi meningkatkan jumlah *false negative*, yaitu kondisi ketika pasien yang sebenarnya mengalami stroke diklasifikasikan sebagai tidak berisiko, sehingga diagnosis dan penanganan medis yang diperlukan dapat tertunda [1]. Salah satu teknik umum untuk mengatasinya adalah SMOTE, metode oversampling yang menghasilkan data sintesis pada kelas minoritas guna menyeimbangkan data latih tanpa mengurangi informasi kelas mayoritas [12], [13]. Sejumlah studi klasifikasi berbasis *Machine Learning* turut mengonfirmasi bahwa penggunaan SMOTE efektif dalam meningkatkan kemampuan model untuk mengenali data dari kelas minoritas [14].

Perkembangan penelitian prediksi stroke menggunakan *Machine Learning* menunjukkan pergeseran dari penerapan algoritma tunggal menuju pemanfaatan pendekatan yang lebih komprehensif. Berbagai penelitian mulai mengombinasikan tahapan *preprocessing*, *feature engineering*, *feature selection*, serta *ensemble learning* untuk memperoleh model dengan performa yang lebih baik. Penelitian [4] mengusulkan metode *hybrid machine learning* untuk prediksi stroke dengan mengintegrasikan proses *feature selection* ke dalam algoritma klasifikasi, sehingga efisiensi model meningkat tanpa mengorbankan kemampuan prediksinya. Penelitian [5] membangun model *stacked machine learning* yang didukung oleh tahapan *preprocessing* dan *feature selection*, di mana hasil pengujiannya memperlihatkan performa yang lebih unggul dibandingkan penggunaan satu algoritma saja. Sementara itu, penelitian [6] memanfaatkan kombinasi teknik *ensemble machine learning* dan *data mining* untuk meningkatkan ketepatan prediksi stroke pada data klinis yang memiliki karakteristik kompleks. Walaupun memberikan hasil yang menjanjikan, ketiga penelitian tersebut masih lebih menitikberatkan pada pengembangan arsitektur model beserta penerapan *feature selection*. Akibatnya, penelitian yang secara spesifik mengkaji perbandingan kinerja antar-algoritma ensemble tunggal, khususnya *Random Forest* dan *XGBoost*, dalam menghadapi kondisi ketidakseimbangan kelas pada data masih tergolong terbatas jumlahnya.

Permasalahan *class imbalance* kerap menjadi perhatian utama dalam beragam kajian yang mengangkat topik prediksi stroke berbasis *Machine Learning*. Penelitian [10] mengembangkan model prediksi stroke dengan memanfaatkan dataset yang distribusi kelasnya tidak seimbang, dan hasilnya memperlihatkan bahwa penanganan ketidakseimbangan kelas meningkatkan kemampuan model mengidentifikasi data kelas minoritas. Temuan tersebut sejalan dengan hasil penelitian [15] yang mengembangkan *explainable machine learning pipeline* melalui penggabungan metode penyeimbangan data dan pendekatan interpretabilitas model, sehingga proses prediksi menjadi lebih mudah dipahami sekaligus tetap memberikan kinerja yang baik. Hasil serupa juga dilaporkan pada penelitian [11] yang menyatakan bahwa penggunaan strategi untuk menangani *imbalanced data* memberikan kontribusi terhadap peningkatan performa model klasifikasi dalam memprediksi stroke. Beragam temuan ini mengindikasikan bahwa penanganan distribusi kelas yang tidak seimbang berperan penting dalam meningkatkan kemampuan model mendeteksi kasus stroke yang berkedudukan sebagai kelas minoritas. Meski begitu, teknik penyeimbangan data yang diterapkan pada masing-masing penelitian masih beragam, sehingga performa yang dicapai belum menunjukkan konsistensi yang seragam di antara berbagai studi.

Berbagai studi telah melakukan analisis perbandingan algoritma guna menemukan metode yang paling unggul dalam memprediksi stroke. Penelitian [16] menguji kinerja *Support Vector Machine*, *Random Forest*, dan *XGBoost*, dan mendapati bahwa algoritma berbasis ensemble sanggup menghasilkan capaian yang kompetitif dibandingkan pendekatan lain. Hasil yang sejalan juga diperoleh pada penelitian [17] yang membandingkan performa *XGBoost*, *Random Forest*, dan *Logistic Regression* pada klasifikasi stroke. Sementara itu, penelitian [18] yang dilaksanakan pada domain di luar bidang kesehatan memperlihatkan bahwa *Random Forest* mampu menghasilkan kinerja

klasifikasi yang stabil dalam analisis sentimen. Temuan-temuan tersebut mengindikasikan bahwa *Random Forest* dan *XGBoost* merupakan algoritma dengan daya saing yang tinggi. Akan tetapi, hasil evaluasi keduanya masih sangat dipengaruhi oleh karakteristik dataset yang digunakan, tahapan *preprocessing*, metode yang diterapkan untuk menangani *class imbalance*, serta pengaturan parameter pada masing-masing model.

Berdasarkan berbagai penelitian terdahulu, penerapan *Machine Learning* telah menunjukkan potensi yang besar dalam meningkatkan kemampuan prediksi risiko stroke melalui penerapan berbagai pendekatan klasifikasi. Penelitian [4] mengembangkan metode *hybrid machine learning* dengan mengintegrasikan *feature selection* ke dalam proses klasifikasi, sedangkan [5] mengusulkan pendekatan *stacked machine learning* yang dipadukan dengan tahapan *preprocessing* dan *feature selection*. Sementara itu, [10] berfokus pada penanganan *class imbalance*, sedangkan [17] melakukan perbandingan beberapa algoritma klasifikasi, termasuk *Random Forest* dan *XGBoost*, pada kasus prediksi stroke. Meskipun memberikan hasil yang menjanjikan, keempat penelitian tersebut menggunakan pendekatan, konfigurasi model, dan tahapan *preprocessing* yang berbeda sehingga belum memberikan perbandingan yang objektif terhadap performa *Random Forest* dan *XGBoost* ketika diterapkan pada dataset yang sama dengan *pipeline preprocessing* yang konsisten.

Selain perbedaan pendekatan klasifikasi, strategi penanganan *class imbalance* pada penelitian terdahulu juga masih menunjukkan variasi. Penelitian [15] menggabungkan metode penyeimbangan data dengan pendekatan *Explainable Artificial Intelligence*, penelitian [13] memanfaatkan SMOTE bersama algoritma *XGBoost*, sedangkan [10] menitikberatkan pada pengaruh penanganan *imbalanced data* terhadap peningkatan deteksi kelas minoritas. Di sisi lain, penelitian [14] menunjukkan bahwa integrasi SMOTE dalam *pipeline preprocessing* dapat meningkatkan representasi kelas minoritas pada permasalahan klasifikasi. Namun demikian, penelitian-penelitian tersebut belum secara khusus mengevaluasi pengaruh penerapan SMOTE terhadap perbandingan performa *Random Forest* dan *XGBoost* menggunakan *pipeline preprocessing* yang sama, sehingga kontribusi SMOTE terhadap kedua algoritma tersebut masih belum dapat dibandingkan secara objektif.

Perbedaan juga terlihat pada aspek evaluasi model. Penelitian [8], [10], [15] menekankan pentingnya penggunaan metrik evaluasi yang lebih komprehensif pada dataset dengan karakteristik *class imbalance*, karena nilai *accuracy* yang tinggi belum tentu mencerminkan kemampuan model dalam mendeteksi kelas minoritas. Meskipun demikian, pendekatan evaluasi yang digunakan pada berbagai penelitian masih beragam sehingga hasil komparasi antaralgoritma menjadi sulit dibandingkan secara langsung. Oleh karena itu, masih diperlukan penelitian yang menerapkan *pipeline preprocessing* yang konsisten serta menggunakan metrik evaluasi yang lebih menyeluruh, meliputi *Accuracy*, *Precision*, *Recall*, *F1-Score*, *ROC-AUC*, *Confusion Matrix*, dan *Feature Importance*, agar perbandingan performa *Random Forest* dan *XGBoost* dapat dilakukan secara lebih objektif.

Selain strategi *preprocessing* dan evaluasi model, representasi fitur pada data tabular juga menjadi aspek yang terus berkembang dalam penelitian *Machine Learning*. Berbagai penelitian menunjukkan bahwa transformasi atribut ke dalam bentuk numerik berperan penting dalam meningkatkan kualitas pembelajaran model [19], [20], [21]. Mengacu pada perkembangan tersebut, penelitian ini menerapkan tahapan *preprocessing* yang sama pada kedua algoritma, yang meliputi *data cleaning*, *Stratified Train-Test Split*, transformasi fitur melalui *Standard Scaling* pada atribut numerik dan TF-IDF pada representasi fitur kategorikal, serta SMOTE pada data latih. Selanjutnya, kinerja *Random Forest* dan *XGBoost* dianalisis menggunakan metrik evaluasi yang lebih menyeluruh, yaitu *Accuracy*, *Precision*, *Recall*, *F1-Score*, *ROC-AUC*, *Confusion Matrix*, serta *Feature Importance*. Dengan menggunakan alur pengolahan data yang konsisten, perbandingan performa kedua algoritma dapat dilakukan dengan lebih objektif, karena hasil yang diperoleh lebih mencerminkan karakteristik masing-masing algoritma, bukan dipengaruhi oleh perbedaan tahapan *preprocessing*, terutama dalam mendeteksi kelas minoritas pada dataset prediksi stroke yang mengalami *class imbalance*.

Berdasarkan uraian tersebut, penelitian ini bertujuan membandingkan performa antara *Random Forest* dan *XGBoost* untuk memprediksi risiko stroke pada dataset yang tidak seimbang, dengan SMOTE sebagai teknik penyeimbang data. Kontribusi utama penelitian ini terletak pada penerapan *pipeline preprocessing* yang konsisten, meliputi *data cleaning*, *Standard Scaling*, representasi fitur kategorikal menggunakan TF-IDF, *Stratified Train-Test Split*, dan SMOTE pada data latih, sehingga perbandingan performa kedua algoritma dapat dilakukan secara lebih objektif menggunakan metrik evaluasi yang komprehensif, yaitu *Accuracy*, *Precision*, *Recall*, *F1-Score*, *ROC-AUC*, *Confusion Matrix*, dan *Feature Importance*. Temuan penelitian ini diharapkan dapat menjadi referensi dalam pengembangan model prediksi stroke pada dataset dengan karakteristik *class imbalance* sekaligus menjadi dasar bagi penelitian selanjutnya dalam mengembangkan metode klasifikasi yang lebih efektif. Penelitian ini turut mendukung pencapaian *Sustainable Development Goal (SDG) 3: Good Health and Well-being* melalui pemanfaatan kecerdasan buatan untuk mendukung pengembangan deteksi dini stroke.

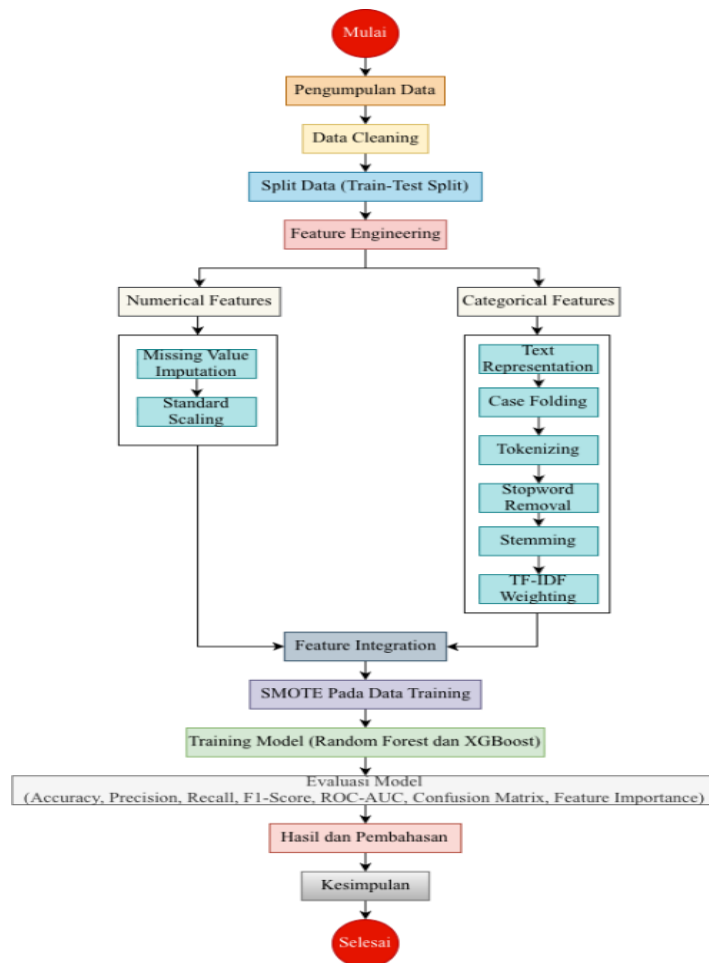
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menerapkan pendekatan kuantitatif melalui metode eksperimen guna menganalisis sekaligus membandingkan performa algoritma *Random Forest* serta *XGBoost* untuk memprediksi risiko stroke pada dataset medis yang distribusi kelasnya tidak seimbang. Pelaksanaan penelitian ditempuh melalui beberapa tahapan yang

berurutan, meliputi pengumpulan data, *preprocessing*, pembangunan model, hingga proses evaluasi hasil klasifikasi. Setiap tahapan dirancang agar model yang dihasilkan mampu menangani karakteristik *imbalanced data* dengan memanfaatkan teknik *preprocessing*, metode penyeimbangan data, serta perbandingan dua algoritma *Machine Learning*. Gambaran lengkap mengenai alur penelitian ditampilkan pada Gambar 1.

Keseluruhan tahapan penelitian ini, mulai dari pengolahan data hingga evaluasi hasil, dilaksanakan dengan bahasa pemrograman Python melalui Google Colaboratory. Pustaka *pandas* dimanfaatkan untuk menangani manipulasi data, sementara *scikit-learn* dioptimalkan untuk proses *preprocessing* dan evaluasi model. Masalah ketidakseimbangan kelas (*class imbalance*) dimitigasi melalui pustaka *imbalanced-learn*, sedangkan *Natural Language Toolkit* (NLTK) diterapkan dalam penanganan data kategorikal. Selain itu, implementasi algoritma *XGBoost* dilakukan dengan mengintegrasikan pustaka *xgboost*.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Data dalam penelitian ini bersumber dari Stroke Prediction Dataset yang dipublikasikan oleh Fedesoriano dan dapat diakses melalui platform Kaggle. Dataset tersebut berisi 5.110 data pasien dengan 12 variabel, yang terdiri dari 11 variabel sebagai prediktor dan 1 variabel target, yaitu *stroke*. Variabel prediktor mencakup tiga kelompok informasi, yaitu karakteristik demografis (*gender*, *age*, *ever_married*, *work_type*, dan *Residence_type*), riwayat penyakit (*hypertension* dan *heart_disease*), serta indikator klinis dan gaya hidup (*avg_glucose_level*, *bmi*, dan *smoking_status*). Adapun variabel target disajikan dalam bentuk biner, dengan nilai 1 menandakan pasien yang mengalami stroke dan nilai 0 menandakan pasien yang tidak mengalami stroke.

Dataset tersebut dipilih karena memuat berbagai faktor risiko yang relevan terhadap kejadian stroke serta telah banyak dimanfaatkan sebagai *benchmark dataset* dalam penelitian prediksi stroke berbasis *Machine Learning*. Penggunaan dataset yang telah banyak digunakan pada penelitian sebelumnya memungkinkan proses evaluasi dan perbandingan hasil penelitian dilakukan secara lebih objektif serta mendukung reproduktibilitas penelitian [5], [10].

2.3 Preprocessing Data

Untuk menjamin kualitas dan kesiapan data sebelum memasuki tahap klasifikasi, dilakukan serangkaian tahapan *preprocessing*. Tahapan ini meliputi *data cleaning*, *Split data*, *feature engineering*, serta *feature integration*.

2.3.1 Data Cleaning

Sebelum proses pemodelan dilakukan, tahap *data cleaning* diterapkan untuk meningkatkan kualitas dataset dengan menangani berbagai permasalahan yang berpotensi memengaruhi kinerja model klasifikasi. Pada penelitian ini, proses tersebut mencakup penghapusan atribut *id*, pemeriksaan kesesuaian tipe data pada setiap variabel, serta penanganan *missing value* pada variabel *bmi* menggunakan teknik *median imputation*. Penggunaan nilai median sebagai metode imputasi dipilih karena lebih baik terhadap keberadaan nilai ekstrem (*outlier*) dibandingkan nilai rata-rata, sehingga dinilai lebih sesuai untuk karakteristik data medis [4], [6].

2.3.2 Split Data dan Penyeimbangan Kelas

Pembagian data pada penelitian ini dipisah menjadi dua kelompok, yaitu 80% digunakan sebagai data latih (*training set*) dan 20% digunakan sebagai data uji (*testing set*) dengan menerapkan metode *Stratified Train-Test Split*. Penerapan metode stratifikasi bertujuan untuk mempertahankan agar proporsi kelas *stroke* dan *non-stroke* tetap terjaga pada kedua subset data, sehingga distribusi masing-masing kelas tetap seimbang secara proporsional selama proses pelatihan maupun pada tahap evaluasi model [5]. Setelah proses pembagian data, tahapan *feature engineering* diterapkan dengan mekanisme *fit* pada data latih dan *transform* pada data uji sebelum kedua jenis fitur diintegrasikan menjadi satu matriks fitur.

Meskipun pembagian data telah dilakukan secara terstratifikasi, distribusi kelas pada data latih masih menunjukkan kondisi *class imbalance*. Atas dasar tersebut, metode SMOTE diterapkan hanya pada data latih setelah proses *Feature Engineering* dan *Feature Integration* selesai. Data uji tidak mengalami proses penyeimbangan kelas sehingga tetap mempertahankan distribusi aslinya untuk menghasilkan evaluasi model yang lebih objektif serta menghindari terjadinya *data leakage*. Metode ini membangkitkan sampel sintetis berdasarkan kedekatan antar data pada kelas minoritas dengan memanfaatkan konsep *k-nearest neighbors*, sehingga jumlah sampel kelas minoritas dapat ditingkatkan tanpa melakukan penggandaan data secara langsung [12]. Prosedur pembentukan sampel sintetis menggunakan algoritma SMOTE dijelaskan pada Persamaan (1).

$$x_{baru} = x_i + \lambda(x_{nn} - x_i) \quad (1)$$

Dimana x_i adalah data kelas minoritas, x_{nn} merupakan tetangga terdekat (*nearest neighbor*), dan λ merupakan nilai acak yang berada di rentang 0 sampai 1. Sejumlah penelitian mengenai prediksi stroke menunjukkan bahwa penggunaan SMOTE memperkuat kemampuan model dalam mendeteksi kelas minoritas, terutama melalui peningkatan nilai *Recall* dan *F1-Score* [10], [13], [15].

2.3.3 Feature Engineering

Tahap *feature engineering* bertujuan menghasilkan representasi fitur yang sesuai dengan karakteristik setiap atribut agar dapat dimanfaatkan secara optimal dalam proses pembelajaran model. Pada penelitian ini, proses *feature engineering* dilakukan setelah pembagian data menggunakan *Stratified Train-Test Split*, sehingga seluruh transformasi fitur diterapkan dengan mekanisme *fit* pada data latih (*training set*) dan selanjutnya digunakan untuk mentransformasikan data uji (*testing set*) guna menjaga konsistensi pembelajaran serta menghindari *data leakage*.

Pada fitur numerik, diterapkan Standard Scaling untuk menyeragamkan rentang nilai antar variabel sehingga perbedaan skala tidak mempengaruhi proses pembelajaran model klasifikasi. Transformasi dilakukan menggunakan nilai rata-rata (*mean*) dan simpangan baku (*standard deviation*) dari data latih, kemudian diterapkan pada data uji. Transformasi Standard Scaling dihitung menggunakan Persamaan (2).

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

Dimana x merepresentasikan nilai data, μ melambangkan rata-rata data, dan σ menunjukkan standar deviasi.

Sementara itu, atribut kategorikal yang terdiri atas *gender*, *ever_married*, *work_type*, *Residence_type*, dan *smoking_status* terlebih dahulu direpresentasikan dalam bentuk teks (*text-based representation*) dengan menggabungkan nilai kategori setiap sampel menjadi satu dokumen tekstual. Representasi ini kemudian ditransformasikan menjadi vektor numerik menggunakan metode TF-IDF sehingga dapat diintegrasikan dengan fitur numerik dalam satu *pipeline preprocessing*. Pendekatan representasi fitur yang sesuai pada data tabular diketahui berperan dalam meningkatkan kemampuan pembelajaran model *machine learning* [19], [20], [21].

Sebelum proses pembobotan TF-IDF, representasi teks terlebih dahulu melalui tahapan *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Pada tahap *stopword removal*, token yang merepresentasikan kategori penting tetap dipertahankan agar informasi atribut tidak hilang selama proses pra-proses. Selanjutnya, hasil *preprocessing* ditransformasikan menggunakan TF-IDF sehingga setiap token memperoleh bobot berdasarkan tingkat kemunculannya pada keseluruhan representasi data. Sedangkan implementasinya mengacu pada konsep representasi fitur berbasis Python yang dijelaskan oleh [22]. Perhitungan bobot TF-IDF mengikuti Persamaan (3)-(5).

$$TF(t, d) = \frac{f(t, d)}{\sum f(t, d)} \quad (3)$$

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (4)$$

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \quad (5)$$

Dimana variabel $(f(t, d))$ mendefinisikan frekuensi kemunculan token (t) dalam dokumen (d) , N menunjukkan total keseluruhan dokumen, adapun $df(t)$ merepresentasikan banyaknya dokumen yang memuat token (t) . Melalui pembobotan tersebut, setiap representasi kategorikal diubah menjadi vektor numerik yang kemudian digunakan bersama fitur numerik pada tahap integrasi fitur sebelum proses pembentukan model klasifikasi.

2.3.4 Feature Integration

Pada tahap *feature integration*, fitur numerik yang telah ditransformasikan menggunakan *Standard Scaling* digabungkan dengan fitur kategorikal yang telah direpresentasikan dalam bentuk vektor berbobot menggunakan TF-IDF. Proses tersebut menghasilkan satu matriks fitur numerik terpadu yang merepresentasikan seluruh atribut pada setiap sampel sebagai data masukan (*input*) dalam proses klasifikasi.

Matriks fitur hasil integrasi selanjutnya digunakan pada tahap penyeimbangan kelas menggunakan SMOTE pada data latih, kemudian dimanfaatkan dalam proses pelatihan model *Random Forest* dan *XGBoost*. Dengan mengintegrasikan kedua jenis fitur ke dalam satu *pipeline preprocessing*, seluruh atribut dapat diproses secara konsisten selama pembentukan model.

2.4 Training Model

Penelitian ini melakukan evaluasi komparatif antara dua algoritma klasifikasi berbasis *ensemble learning*, yakni *Random Forest* serta *XGBoost*. Kedua model dikembangkan menggunakan dataset yang telah melalui tahapan *preprocessing* serta penyeimbangan kelas dengan prosedur yang sama. Dengan menerapkan alur pengolahan data yang sama, hasil perbandingan diharapkan dapat lebih mencerminkan karakteristik bawaan dari masing-masing algoritma, sehingga perbedaan performa yang diperoleh tidak dipengaruhi oleh variasi pada proses pengolahan data.

Konfigurasi model pada penelitian ini ditetapkan dengan parameter yang berbeda untuk masing-masing algoritma. Algoritma *Random Forest* menggunakan $n_estimators = 300$ dan $max_depth = None$, sedangkan *XGBoost* dikonfigurasi dengan $n_estimators = 300$, $max_depth = 4$, serta $learning_rate = 0,05$. Selain itu, kedua model menerapkan $random_state = 42$ untuk memastikan hasil eksperimen tetap konsisten sekaligus mendukung reproduksibilitas penelitian.

2.5 Evaluasi Model

Evaluasi terhadap kinerja model klasifikasi, menggunakan sejumlah metrik agar kemampuan prediksi risiko stroke dapat dinilai secara lebih menyeluruh, terutama pada dataset dengan karakteristik *class imbalance*. Penggunaan berbagai metrik dalam evaluasi model telah lazim diterapkan pada penelitian klasifikasi penyakit [23]. Metrik yang diterapkan dalam penelitian ini meliputi *Accuracy*, *Precision*, *Recall*, *F1-Score*, *ROC-AUC*, *Confusion Matrix*, dan *Feature Importance*. Perhitungan untuk metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* mengacu pada Persamaan (6)-(9).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

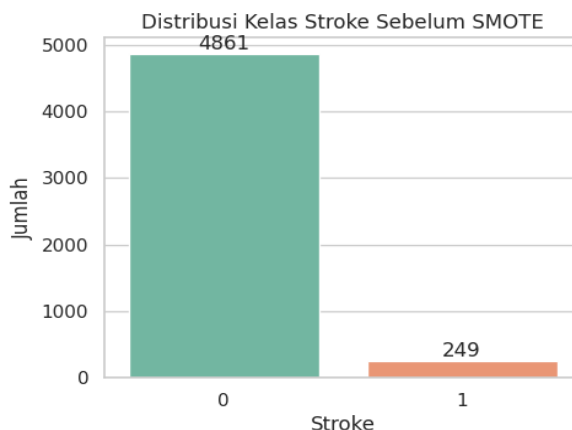
Setiap metrik evaluasi tersebut berfungsi untuk mengukur aspek performa model yang berlainan. *Accuracy* mencerminkan tingkat ketepatan prediksi secara menyeluruh, sementara *Precision* menggambarkan tingkat ketelitian model dalam mengenali data yang tergolong kelas positif. Khusus pada kasus prediksi stroke, *Recall* menjadi ukuran yang krusial karena menggambarkan sejauh mana model mampu mendeteksi pasien yang benar-benar memiliki risiko stroke. Di sisi lain, *F1-Score* dipakai untuk melihat keseimbangan antara *Precision* dan *Recall*, sedangkan *ROC-AUC* berfungsi mengukur kemampuan model membedakan kelas positif dan negatif pada berbagai nilai *threshold*. *Confusion Matrix* digunakan untuk menelaah sebaran hasil klasifikasi, sementara *Feature Importance* mengidentifikasi variabel-variabel paling berpengaruh saat proses prediksi [1], [8], [9].

3. HASIL DAN PEMBAHASAN

3.1 Analisis Ketidakseimbangan Data

Penelitian ini memanfaatkan dataset yang terdiri atas 5.110 data pasien dengan distribusi kelas yang tidak seimbang. Dari keseluruhan data tersebut, terdapat 4.861 sampel yang termasuk ke dalam kelas *Tidak Stroke* dan 249 sampel pada kelas *Stroke*, sehingga rasio ketidakseimbangan kelas mencapai sekitar 19,52:1. Perbedaan jumlah sampel yang cukup besar menunjukkan bahwa representasi kelas minoritas jauh tertinggal dibandingkan kelas mayoritas. Kondisi

tersebut dapat membuat model klasifikasi lebih banyak mempelajari karakteristik kelas mayoritas, sehingga kemampuannya dalam mengenali pola maupun mengidentifikasi data pada kelas minoritas menjadi kurang optimal.



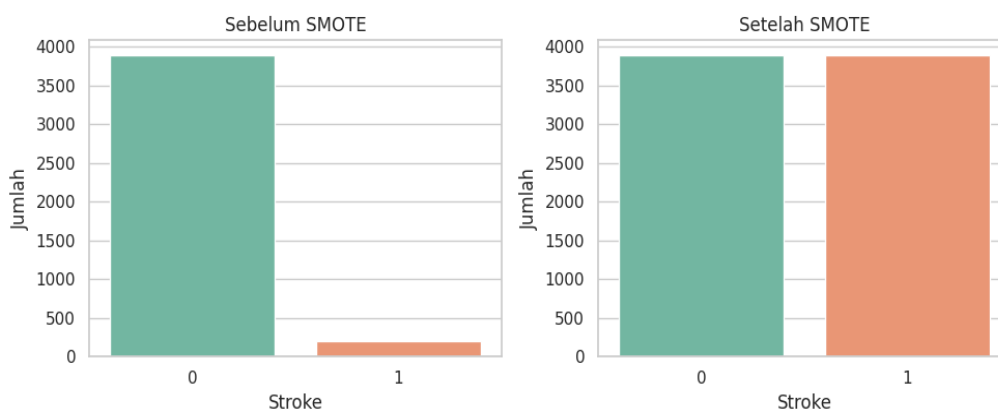
Gambar 2. Distribusi Kelas Stroke Sebelum SMOTE

Class imbalance merupakan tantangan fundamental dalam pengembangan model prediksi penyakit berbasis *machine learning*, karena kondisi ini dapat menghambat kemampuan model dalam membedakan distribusi kelas secara akurat. Dalam situasi tersebut, model sering kali menunjukkan nilai *accuracy* yang tinggi namun gagal dalam mengidentifikasi sampel dari kelas minoritas [10], [11]. Dalam konteks prediksi stroke, kondisi tersebut dapat meningkatkan kemungkinan terjadinya *false negative*, yaitu ketika pasien yang sebenarnya memiliki risiko stroke diklasifikasikan sebagai tidak berisiko. Kekeliruan prediksi seperti ini dapat menimbulkan dampak klinis yang sangat serius karena berpotensi menyebabkan keterlambatan dalam proses diagnosis maupun pemberian terapi atau penanganan medis yang sesuai [1].

Berdasarkan karakteristik *dataset* tersebut, penelitian ini mengimplementasikan SMOTE sebagai strategi penyeimbangan data sebelum memasuki tahap klasifikasi. Penerapan SMOTE bertujuan untuk mengoptimalkan representasi kelas minoritas dalam *training set*, sehingga proses pembelajaran model menjadi lebih seimbang tanpa harus mereduksi jumlah sampel pada kelas mayoritas. Keputusan untuk menggunakan metode ini juga didasarkan pada berbagai studi terdahulu yang menunjukkan bahwa SMOTE terbukti efektif dalam meningkatkan kapabilitas model untuk mengidentifikasi kelas minoritas dalam kasus prediksi *stroke* [13], [15].

3.2 Hasil Penerapan SMOTE

Sebelum proses pembentukan model dilakukan, *Synthetic Minority Over-sampling Technique* (SMOTE) dilakukan hanya pada data latih guna mengatasi ketidakseimbangan distribusi kelas. Sebelum tahap *oversampling*, data latih terdiri atas 3.889 sampel *Tidak Stroke* dan 199 sampel *Stroke*. Setelah SMOTE diterapkan, jumlah data pada kelas minoritas bertambah secara signifikan hingga mencapai jumlah yang sama dengan kelas mayoritas, sehingga masing-masing kelas memiliki 3.889 sampel.



Gambar 3. Distribusi Kelas Sebelum dan Setelah SMOTE

Temuan tersebut menunjukkan bahwa penerapan SMOTE mampu memperbanyak representasi kelas minoritas melalui pembentukan sampel sintetis, tanpa mengurangi data yang berasal dari kelas mayoritas. Proses penyeimbangan hanya dilakukan pada data latih agar distribusi asli pada data uji tetap dipertahankan. Dengan cara tersebut, hasil evaluasi dapat menggambarkan kondisi yang lebih realistis sekaligus mencegah terjadinya *data leakage* yang dapat menyebabkan estimasi performa model menjadi bias [12].

Hasil penerapan SMOTE pada penelitian ini selaras dengan temuan [10] dan [13], yang memperlihatkan bahwa teknik tersebut efektif dalam memperkuat representasi kelas minoritas. Dengan meningkatnya jumlah sampel pada kelas tersebut, model memiliki peluang yang lebih baik untuk mempelajari pola-pola yang sebelumnya kurang terwakili. Oleh karena itu, penyeimbangan data diharapkan mampu menciptakan proses pembelajaran yang lebih proporsional sebelum dilakukan evaluasi terhadap kinerja algoritma *Random Forest* dan *XGBoost*.

3.3 Perbandingan Performa Random Forest dan XGBoost

Penilaian kinerja model dilakukan dengan mempertimbangkan beberapa metrik evaluasi, yaitu *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *ROC-AUC*. Rincian lengkap hasil evaluasi dari masing-masing model yang dikembangkan dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan Performa *Random Forest* dan *XGBoost*

Model	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>ROC-AUC</i>
<i>Random Forest</i>	0,9374	0,1818	0,0800	0,1111	0,7068
<i>XGBoost</i>	0,9237	0,1818	0,1600	0,1702	0,7877

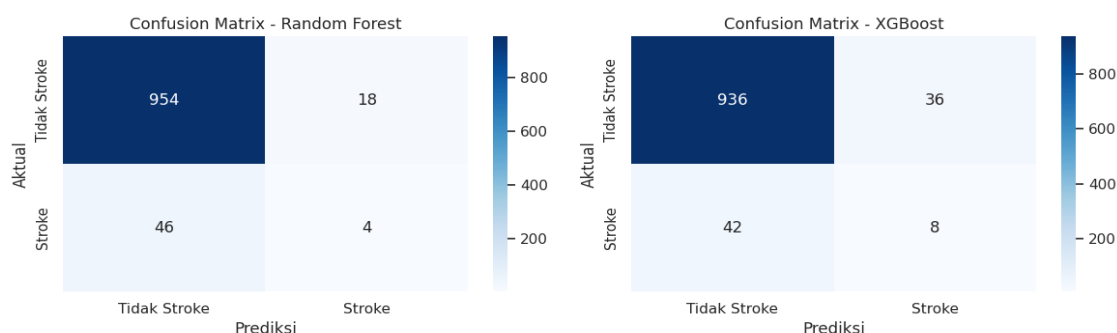
Merujuk pada hasil evaluasi, algoritma *Random Forest* menghasilkan *Accuracy* yang sedikit unggul dibandingkan *XGBoost*. Meskipun demikian, pada dataset yang memiliki karakteristik *class imbalance*, metrik *Accuracy* tidak dapat dijadikan patokan tunggal dalam menilai performa model karena cenderung dipengaruhi oleh dominasi kelas mayoritas [10], [15]. Atas dasar tersebut, penilaian performa model juga dilengkapi dengan metrik *Recall*, *F1-Score*, dan *ROC-AUC*, yang dinilai lebih mampu menggambarkan kemampuan model dalam mengenali data dari kelas minoritas.

Hasil pengujian mengungkapkan bahwa *XGBoost* menghasilkan nilai *Recall*, *F1-Score*, dan *ROC-AUC* yang lebih tinggi dibandingkan *Random Forest*. Peningkatan tersebut mengindikasikan bahwa *XGBoost* memiliki kemampuan yang relatif lebih baik dalam mengenali kelas minoritas pada dataset yang digunakan. Namun demikian, nilai *Recall* sebesar 0,1600 dan *F1-Score* sebesar 0,1702 masih menunjukkan bahwa kemampuan model dalam mendeteksi seluruh kasus stroke belum optimal. Nilai *F1-Score* yang lebih tinggi pada *XGBoost* mencerminkan adanya peningkatan keseimbangan antara *Precision* dan *Recall* dibandingkan *Random Forest*, sedangkan peningkatan *ROC-AUC* menunjukkan kemampuan diskriminasi kelas yang lebih baik. Meskipun *Random Forest* memperoleh nilai *Accuracy* yang sedikit lebih tinggi, selisih tersebut relatif kecil dan kurang representatif sebagai indikator utama pada dataset dengan karakteristik *class imbalance*. Oleh sebab itu, metrik *Recall*, *F1-Score*, dan *ROC-AUC* dinilai lebih tepat dijadikan sebagai dasar evaluasi model pada dataset medis dengan karakteristik *class imbalance* [8], [10], [15].

Nilai *Recall*, *F1-Score*, dan *ROC-AUC* yang lebih tinggi pada *XGBoost* dibandingkan *Random Forest* diduga berasal dari mekanisme *gradient boosting* yang membangun model secara bertahap dengan memberi bobot lebih besar terhadap sampel yang sebelumnya mengalami kesalahan klasifikasi. Setelah distribusi kelas diperbaiki melalui penerapan SMOTE, mekanisme ini memungkinkan *XGBoost* membentuk *decision boundary* yang lebih peka dalam membedakan karakteristik kelas minoritas dibandingkan pendekatan *bagging* yang digunakan oleh *Random Forest* [24], [25]. Temuan tersebut memperkuat pentingnya pemilihan metrik evaluasi yang sesuai untuk dataset dengan karakteristik *class imbalance*, karena penggunaan *Accuracy* sebagai satu-satunya indikator belum mampu menggambarkan kemampuan model dalam mendeteksi pasien berisiko stroke secara memadai.

3.4 Analisis Confusion Matrix

Confusion Matrix digunakan sebagai bagian dari proses evaluasi untuk menganalisis secara lebih rinci kemampuan setiap model dalam mengelompokkan pasien ke dalam kelas *Stroke* maupun *Tidak Stroke*. Melalui analisis ini, setiap jenis kesalahan klasifikasi yang dihasilkan oleh masing-masing algoritma dapat diidentifikasi, sehingga karakteristik kinerja model dapat dipahami dan dievaluasi secara lebih menyeluruh.



Gambar 4. *Confusion Matrix* Model *Random Forest* dan *XGBoost*

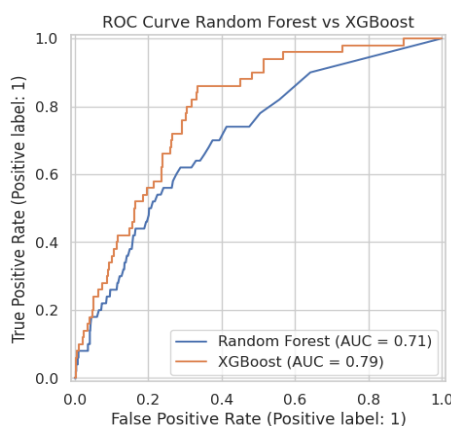
Berdasarkan Gambar 4, *Random Forest* mampu mengklasifikasikan 954 pasien *Tidak Stroke* sebagai *true negative* dan 4 pasien *Stroke* sebagai *true positive*. Namun, model ini masih menghasilkan 46 kasus *false negative*,

yakni kondisi di mana pasien dengan risiko *stroke* diprediksi sebagai tidak berisiko. Sebaliknya, *XGBoost* menunjukkan performa lebih baik dengan meningkatkan jumlah *true positive* menjadi 8 kasus dan menekan jumlah *false negative* menjadi 42 kasus. Meskipun terjadi peningkatan jumlah *false positive* dari 18 menjadi 36 kasus, hasil ini mengindikasikan bahwa *XGBoost* memiliki kemampuan yang relatif lebih baik dalam mengidentifikasi kelas minoritas dibandingkan *Random Forest*, meskipun jumlah *false negative* yang masih tinggi menunjukkan bahwa kemampuan deteksi kasus *stroke* belum optimal.

Dalam konteks prediksi *stroke*, peningkatan jumlah *true positive* yang disertai dengan penurunan *false negative* merupakan indikator performa yang lebih krusial dibandingkan sekadar capaian *accuracy* yang tinggi. Hal ini dikarenakan kesalahan klasifikasi berupa *false negative* yang menyebabkan pasien sebenarnya berisiko tidak teridentifikasi, sehingga menghambat proses diagnosis dan penanganan medis yang tepat waktu, serta meningkatkan risiko komplikasi hingga fatalitas [10]. Meskipun *XGBoost* menghasilkan jumlah *false positive* yang lebih tinggi, peningkatan jumlah *true positive* menunjukkan adanya peningkatan kemampuan deteksi dibandingkan *Random Forest*. Namun demikian, jumlah *false negative* yang masih cukup besar menunjukkan bahwa model masih memerlukan pengembangan lebih lanjut sebelum dapat diterapkan pada skenario yang memerlukan tingkat sensitivitas tinggi.

3.5 Analisis ROC Curve

Kinerja model juga dianalisis menggunakan kurva ROC untuk menilai kemampuan model dalam membedakan kelas *Stroke* dan *Tidak Stroke* pada berbagai nilai *threshold*. Penilaian tersebut didasarkan pada hubungan antara *True Positive Rate* dan *False Positive Rate*, sehingga kurva ROC menjadi salah satu ukuran yang tepat untuk mengevaluasi kemampuan diskriminasi model, terutama ketika diterapkan pada dataset yang memiliki distribusi kelas tidak seimbang.



Gambar 5. ROC Curve Perbandingan Model

Berdasarkan Gambar 5, kurva ROC yang dihasilkan oleh *XGBoost* berada di atas kurva *Random Forest* pada sebagian besar rentang *threshold*, yang diskriminasi yang relatif lebih baik pada dataset yang digunakan dalam penelitian ini. Hasil tersebut sejalan dengan nilai *ROC-AUC*, di mana *XGBoost* memperoleh skor sebesar 0,7877, lebih tinggi dibanding *Random Forest* yang mencapai 0,7068, atau mengalami peningkatan sekitar 11,44%. Temuan ini mengindikasikan bahwa *XGBoost* memiliki kemampuan yang lebih baik dalam membedakan kelas *Stroke* dan *Tidak Stroke* pada berbagai nilai *threshold*. Namun demikian, peningkatan kemampuan diskriminasi ini belum sepenuhnya diikuti oleh kemampuan deteksi kelas minoritas yang tinggi, sebagaimana ditunjukkan oleh nilai *Recall* yang masih relatif rendah.

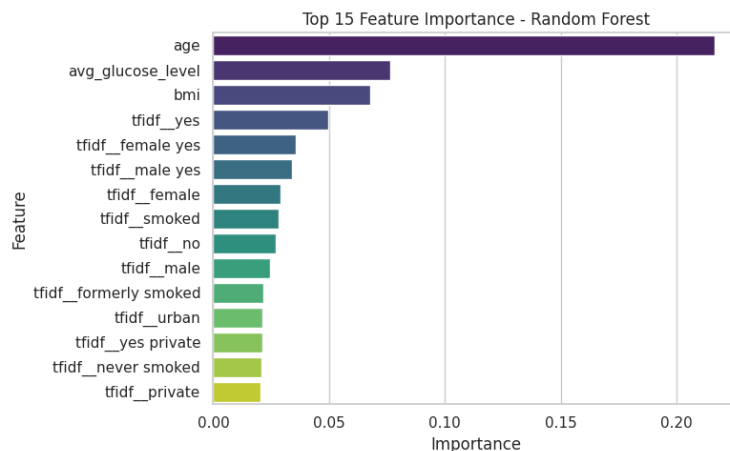
Secara umum, model dengan nilai *ROC-AUC* di atas 0,70 merepresentasikan kapabilitas diskriminasi model yang baik, sedangkan nilai yang mendekati 0,80 mencerminkan efektivitas klasifikasi yang semakin baik. Oleh karena itu, capaian skor *ROC-AUC* yang diraih *XGBoost* mengindikasikan bahwa model memiliki kemampuan diskriminasi yang lebih baik dibandingkan *Random Forest* pada dataset yang digunakan dalam penelitian ini. Namun demikian, nilai tersebut belum cukup untuk menggambarkan kemampuan model dalam mendeteksi seluruh kasus *stroke* sehingga tetap perlu diinterpretasikan bersama metrik lain, seperti *Recall* dan *F1-Score*.

Berdasarkan perbedaan nilai *ROC-AUC* antara kedua model tersebut, mekanisme *gradient boosting* dalam *XGBoost* dapat menghasilkan batas keputusan yang lebih baik daripada *Random Forest* pada dataset yang digunakan dalam penelitian ini. Untuk memberikan gambaran yang lebih lengkap mengenai kinerja model, nilai *ROC-AUC* harus diinterpretasikan bersamaan dengan metrik *Recall* dan *F1-Score*, karena kemampuan diskriminasi yang lebih baik ini tidak sepenuhnya berarti peningkatan kemampuan dalam mendeteksi semua kasus *stroke*.

3.6 Analisis Feature Importance

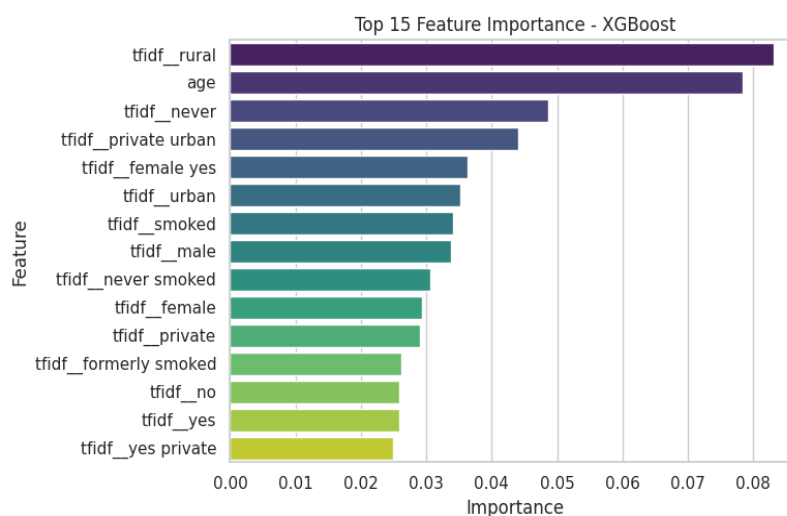
Evaluasi *Feature Importance* dilaksanakan guna mengetahui besarnya kontribusi tiap variabel dalam proses prediksi risiko *stroke* pada masing-masing model yang dibangun. Melalui evaluasi ini, fitur-fitur yang paling berpengaruh

terhadap keputusan klasifikasi dapat diidentifikasi, sehingga karakteristik dan cara kerja model dapat dipahami secara lebih mendalam.



Gambar 6. Grafik *Feature Importance* Model *Random Forest*

Pada model *Random Forest*, variabel *age* merupakan fitur dengan tingkat kepentingan tertinggi, diikuti oleh *avg_glucose_level* dan *bmi*. Hasil tersebut menunjukkan bahwa faktor usia serta indikator metabolik menjadi variabel yang paling dominan dalam membedakan pasien Stroke dan Tidak Stroke. Temuan ini konsisten dengan penelitian [2], [3], serta [4] yang menyatakan bahwa usia yang lebih lanjut, kadar glukosa darah, serta indeks massa tubuh merupakan faktor risiko utama yang berkaitan dengan kejadian stroke.



Gambar 7. Grafik *Feature Importance* Model *XGBoost*

Berbeda dengan *Random Forest*, model *XGBoost* menunjukkan bahwa beberapa fitur hasil transformasi TF-IDF, khususnya *tfidf_rural*, memiliki nilai *feature importance* yang lebih tinggi dibandingkan sebagian besar fitur lainnya. Temuan tersebut menunjukkan bahwa token *rural* pada representasi TF-IDF memberikan kontribusi yang relatif besar terhadap proses prediksi yang dilakukan oleh model *XGBoost* pada dataset penelitian ini. Atribut *Residence_type* yang direpresentasikan melalui token *rural* juga telah banyak dibahas dalam berbagai penelitian mengenai stroke sebagai salah satu karakteristik yang berkaitan dengan perbedaan kondisi kesehatan antarwilayah [26], [27]. Oleh karena itu, tingginya nilai *feature importance* pada *tfidf_rural* menunjukkan bahwa representasi fitur tersebut memiliki nilai prediktif yang tinggi dalam model yang dibangun.

Secara keseluruhan, kedua model menunjukkan pola *feature importance* yang berbeda. *Random Forest* lebih banyak dipengaruhi oleh fitur numerik seperti *age*, *avg_glucose_level*, dan *bmi*, yang merupakan fitur dengan kontribusi terbesar dalam proses prediksi, sehingga hasil penelitian ini tetap konsisten dengan berbagai penelitian sebelumnya mengenai faktor risiko stroke [2], [3], [4]. Sementara itu, *XGBoost* memberikan kontribusi yang lebih besar pada beberapa fitur hasil representasi TF-IDF, khususnya *tfidf_rural*. Perbedaan tersebut menunjukkan bahwa *Random Forest* dan *XGBoost* memanfaatkan karakteristik fitur yang berbeda selama proses pembelajaran model. Oleh karena itu, interpretasi *feature importance* pada kedua model perlu dilakukan secara hati-hati dan dipahami sebagai kontribusi relatif suatu fitur terhadap keputusan model, bukan sebagai dasar untuk menyimpulkan hubungan sebab-akibat maupun tingkat pengaruh klinis suatu variabel terhadap kejadian stroke.

3.7 Pembahasan

Hasil penelitian memperlihatkan bahwa algoritma *XGBoost* memperoleh *Recall*, *F1-Score*, dan *ROC-AUC* yang lebih tinggi dibandingkan *Random Forest* pada dataset prediksi stroke yang mengalami *class imbalance*, meski nilai *Accuracy Random Forest* sedikit lebih unggul. Temuan ini sejalan dengan penelitian [17] yang juga melaporkan *XGBoost* sebagai model dengan performa terbaik dibandingkan *Random Forest* dan *Logistic Regression* pada klasifikasi stroke setelah penerapan SMOTE. Namun, berbeda dari penelitian tersebut yang lebih menitikberatkan pada capaian *Accuracy*, penelitian ini menegaskan bahwa pada dataset dengan distribusi kelas tidak seimbang, interpretasi performa model perlu lebih berfokus pada *Recall*, *F1-Score*, dan *ROC-AUC*, karena ketiga metrik ini lebih representatif dalam menggambarkan kemampuan model mengenali kelas minoritas.

Temuan penelitian ini turut mendukung hasil penelitian [10], [15], yang menegaskan bahwa penanganan *class imbalance* berperan penting dalam meningkatkan kemampuan model mendeteksi kelas minoritas. Meski penerapan SMOTE pada penelitian ini berhasil menghasilkan distribusi data latih yang lebih seimbang dan mendorong performa *XGBoost* melampaui *Random Forest*, nilai *Recall* sebesar 0,1600 mengindikasikan bahwa kemampuan model mendeteksi seluruh kasus stroke masih terbatas. Hal ini menunjukkan bahwa penanganan *class imbalance* saja belum cukup untuk mencapai performa optimal dalam prediksi stroke, sehingga pengembangan metode dan optimasi model lebih lanjut masih diperlukan.

Sebaliknya, hasil penelitian ini berbeda dari [4], yang mencapai performa prediksi lebih tinggi melalui pendekatan *hybrid machine learning* dengan *feature selection*. Perbedaan ini mengindikasikan bahwa pendekatan yang lebih kompleks berpotensi meningkatkan performa model. Kendati demikian, penelitian ini menawarkan kontribusi tersendiri, yakni perbandingan *Random Forest* dan *XGBoost* melalui *pipeline preprocessing* yang konsisten serta evaluasi yang komprehensif pada dataset *class imbalance*, sehingga karakteristik kedua algoritma dapat dibandingkan secara lebih objektif.

4. KESIMPULAN

Hasil yang diperoleh pada penelitian ini menunjukkan bahwa model *XGBoost* memperoleh nilai *Recall*, *F1-Score*, dan *ROC-AUC* yang lebih tinggi dibandingkan *Random Forest* dalam memprediksi risiko stroke pada dataset dengan karakteristik *class imbalance*, meskipun *Random Forest* menghasilkan nilai *Accuracy* yang sedikit lebih tinggi. Temuan tersebut mengindikasikan bahwa penilaian kinerja model pada data medis yang distribusi kelasnya tidak seimbang sebaiknya tidak hanya didasarkan pada *Accuracy*, tetapi juga perlu mempertimbangkan metrik lain yang lebih representatif terhadap kemampuan model dalam mengenali kelas minoritas. Selain itu, penerapan SMOTE yang dilakukan pada data latih berhasil menghasilkan distribusi kelas yang lebih seimbang, sedangkan analisis *feature importance* menunjukkan bahwa *age*, *avg glucose level*, dan *bmi* merupakan fitur yang memberikan kontribusi terbesar terhadap proses prediksi pada model *Random Forest*. Kontribusi penelitian ini terletak pada penerapan *pipeline preprocessing* yang konsisten dan penggunaan metrik evaluasi yang komprehensif sehingga perbandingan performa *Random Forest* dan *XGBoost* dapat dilakukan secara lebih objektif pada dataset prediksi stroke yang mengalami *class imbalance*. Meskipun begitu, nilai *Recall* sebesar 0,1600 menunjukkan bahwa kemampuan model dalam mendeteksi kasus stroke masih terbatas sehingga belum memadai untuk mendukung kebutuhan deteksi dengan sensitivitas tinggi. Oleh sebab itu, penelitian selanjutnya disarankan untuk melakukan optimasi *hyperparameter*, mengeksplorasi metode penyeimbangan data lain, seperti ADASYN dan *Borderline-SMOTE*, serta menguji model pada dataset yang berskala lebih besar dan lebih bervariasi, agar performa prediksi maupun kemampuan generalisasinya dapat ditingkatkan.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: C. D. A., B. W. P., M. R. S., dan D. R. I.;

Metodologi: C. D. A., B. W. P., dan M. R. S.;

Perangkat Lunak: C. D. A.;

Validasi: C. D. A., B. W. P., M. R. S., dan D. R. I.;

Analisis Formal: C. D. A., B. W. P., dan M. R. S.;

Investigasi: C. D. A., B. W. P., dan M. R. S.;

Sumber Daya: C. D. A., B. W. P., dan M. R. S.;

Kurasi Data: C. D. A.;

Penulisan Draf Awal: C. D. A., B. W. P., dan M. R. S.;

Penulisan, Peninjauan, dan Penyuntingan: C. D. A., B. W. P., M. R. S., dan D. R. I.;

Visualisasi: C. D. A., B. W. P., dan M. R. S.;

Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis pertama.

5.3 Pendanaan

Penelitian ini tidak menerima pendanaan eksternal.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (*Not applicable*).

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (*Not applicable*).

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCES

- [1] T. Vu *et al.*, “Machine Learning Approaches for Stroke Risk Prediction: Findings from the Suita Study,” *JCDD*, vol. 11, no. 7, p. 207, Jul. 2024, doi: 10.3390/jcdd11070207.
- [2] A. Papadopoulou, D. Harding, G. Slabaugh, E. Marouli, and P. Deloukas, “Prediction of atrial fibrillation and stroke using machine learning models in UK Biobank,” *Heliyon*, vol. 10, no. 7, p. e28034, Apr. 2024, doi: 10.1016/j.heliyon.2024.e28034.
- [3] S. Sahriar *et al.*, “Unlocking stroke prediction: Harnessing projection-based statistical feature extraction with ML algorithms,” *Heliyon*, vol. 10, no. 5, p. e27411, Mar. 2024, doi: 10.1016/j.heliyon.2024.e27411.
- [4] P. Bathla and R. Kumar, “A hybrid system to predict brain stroke using a combined feature selection and classifier,” *Intelligent Medicine*, vol. 4, no. 2, pp. 75–82, May 2024, doi: 10.1016/j.imed.2023.06.002.
- [5] P. Chakraborty *et al.*, “Predicting stroke occurrences: a stacked machine learning approach with feature selection and data preprocessing,” *BMC Bioinformatics*, vol. 25, no. 1, p. 329, Oct. 2024, doi: 10.1186/s12859-024-05866-8.
- [6] R. Wijaya, F. Saeed, P. Samimi, A. M. Albarrak, and S. N. Qasem, “An Ensemble Machine Learning and Data Mining Approach to Enhance Stroke Prediction,” *Bioengineering*, vol. 11, no. 7, p. 672, Jul. 2024, doi: 10.3390/bioengineering11070672.
- [7] I. Abousaber, “A Novel Explainable Attention-Based Meta-Learning Framework for Imbalanced Brain Stroke Prediction,” *Sensors*, vol. 25, no. 6, p. 1739, Mar. 2025, doi: 10.3390/s25061739.
- [8] J. Lee, K. M. Park, and S. Park, “Interpretable machine learning for prediction of clinical outcomes in acute ischemic stroke,” *Front. Neurol.*, vol. 14, p. 1234046, Sep. 2023, doi: 10.3389/fneur.2023.1234046.
- [9] P. N. Srinivasu, U. Sirisha, K. Sandeep, S. P. Praveen, L. P. Maguluri, and T. Bikku, “An Interpretable Approach with Explainable AI for Heart Stroke Prediction,” *Diagnostics*, vol. 14, no. 2, p. 128, Jan. 2024, doi: 10.3390/diagnostics14020128.
- [10] N. Melnykova, Y. Patercha, S. Skopivskyi, M. Farion, S. Fedushko, and K. Drohomyska, “Machine learning for stroke prediction using imbalanced data,” *Sci Rep*, vol. 15, no. 1, p. 33773, Sep. 2025, doi: 10.1038/s41598-025-01855-w.
- [11] A. Rehman, T. Alam, M. Mujahid, F. S. Alamri, B. A. Ghofaili, and T. Saba, “RDET stacking classifier: a novel machine learning based approach for stroke prediction using imbalance data,” *PeerJ Computer Science*, vol. 9, p. e1684, Nov. 2023, doi: 10.7717/peerj-cs.1684.
- [12] K. Abhishek, *Machine Learning for Imbalanced Data: Tackle imbalanced datasets using machine learning and deep learning techniques*, 1st ed. Birmingham: Packt Publishing Limited, 2023.
- [13] Y. Li *et al.*, “Hematoma expansion prediction based on SMOTE and XGBoost algorithm,” *BMC Med Inform Decis Mak*, vol. 24, no. 1, p. 172, Jun. 2024, doi: 10.1186/s12911-024-02561-9.
- [14] A. Y. Imran, M. R. Sanjaya, B. W. Putra, and G. E. H. Cahyadi, “Optimization Of Sentiment Analysis On Tokopedia User Reviews Using Gridsearchcv And Smote With Machine Learning Algorithms,” *JURNAL INOVTEK POLBENG - SERI INFORMATIKA*, vol. 10, no. 3, 2025, doi: 10.35314/5ax8km80.
- [15] C. Kokkotis *et al.*, “An Explainable Machine Learning Pipeline for Stroke Prediction on Imbalanced Data,” *Diagnostics*, vol. 12, no. 10, p. 2392, Oct. 2022, doi: 10.3390/diagnostics12102392.
- [16] H. A. Nabila and Endang Wahyu Pamungkas, “Perbandingan Algoritma Machine Learning: Svm, Random Forest, Dan Xgboost Untuk Prediksi Stroke,” *rabit*, vol. 10, no. 2, pp. 1098–1110, Jul. 2025, doi: 10.36341/rabit.v10i2.6444.
- [17] L. R. Sitompul, A. A. Nababan, M. L. Manihuruk, W. A. Ponsen, and S. Supriyadi, “Comparison of Xgboost, Random Forest and Logistic Regression Algorithms in Stroke Disease Classification,” *Sinkron*, vol. 9, no. 2, pp. 957–968, Jun. 2025, doi: 10.33395/sinkron.v9i2.14794.
- [18] M. R. Sanjaya, A. Khoiriah, R. I. Heroza, and B. W. Putra, “Analysis Of User Reviews For The Mytelkomsel App Using Naïve Bayes And Random Forest Methods,” *Jurnal Testing dan Implementasi Sistem Informasi*, vol. 4(1), pp. 32–48, 2026, doi: 10.55583/jtisi.v4i1.2220.



- [19] N. Hollmann *et al.*, “Accurate predictions on small data with a tabular foundation model,” *Nature*, vol. 637, no. 8045, pp. 319–326, Jan. 2025, doi: 10.1038/s41586-024-08328-6.
- [20] M. J. Kim, L. Grinsztajn, and G. Varoquaux, “CARTE: Pretraining and Transfer for Tabular Learning,” May 31, 2024, *arXiv*: arXiv:2402.16785. doi: 10.48550/arXiv.2402.16785.
- [21] M. Mráz, B. Das, A. Gupta, L. Purucker, and F. Hutter, “Towards Benchmarking Foundation Models for Tabular Data With Text,” Jul. 10, 2025, *arXiv*: arXiv:2507.07829. doi: 10.48550/arXiv.2507.07829.
- [22] Munaldi, *Algoritma Random Forest dan XGBoost Menggunakan Python*. Indonesia: Ganesha Kreasi Semesta, 2025. [Online]. Available: <https://repository.ganeshakreasisemesta.com/publications/640026/algoritma-random-forest-dan-xgboost-menggunakan-python>
- [23] B. W. Putra, “Deep Neural Network Classification for ECG Arrhythmia Detection with Optimized Feature Autoencoder,” *J. Adv. Res. Appl. Mech.*, vol. 133, no. 1, pp. 78–86, Feb. 2025, doi: 10.37934/aram.133.1.7886.
- [24] R. Chen *et al.*, “A study on predicting the length of hospital stay for Chinese patients with ischemic stroke based on the XGBoost algorithm,” *BMC Med Inform Decis Mak*, vol. 23, no. 1, p. 49, Mar. 2023, doi: 10.1186/s12911-023-02140-4.
- [25] W. Hong *et al.*, “A Comparison of XGBoost, Random Forest, and Nomograph for the Prediction of Disease Severity in Patients With COVID-19 Pneumonia: Implications of Cytokine and Immune Cell Profile,” *Front. Cell. Infect. Microbiol.*, vol. 12, p. 819267, Apr. 2022, doi: 10.3389/fcimb.2022.819267.
- [26] G. Howard, “Rural-urban differences in stroke risk,” *Preventive Medicine*, vol. 152, p. 106661, Nov. 2021, doi: 10.1016/j.ypmed.2021.106661.
- [27] A. Machin, S. P. Widiawan, A. Lefi, M. S. Ardhi, N. Husni, and M. W. Rosyich, “Stroke Risk Factors Based on The Framingham Stroke Risk Score Among Urban vs Rural Elderly Communities,” *Malaysian Journal of Medicine and Health Sciences*, vol. 20, no. SUPP9, pp. 35–41, 2024, doi: 10.47836/mjmhs20.s9.6.