Jurnal Teknik Informatika dan Teknologi Informasi
Volume. 5 Nomor. 3 Desember 2025
E-ISSN .: 2827-9387, P-ISSN .: 2827-9379, Hal 261-272
DOI: https://doi.org/10.55606/jutiti.v5i3.6254
Tersedia: https://journalshub.org/index.php/jutiti

Penerapan dan Perbandingan Algoritma SVM, *Naive Bayes*, dan *Gradient Boosting* dalam Prediksi Stroke

Joseph Melchior Nababan[1*], Iqro Mukti Arto [2], Putra Satria[3], Sumanto[4], Imam Budiawan [5], Roida Pakpahan[6]
[1-6]Prodi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia
E-mail: jojerde254@gmail.com[1],iqrmktart23@gmail.com[2]*
,putrasatriaaa31@gmail.com[3], sumanto@bsi.ac.id[4] ,imam.imb@bsi.ac.id [5], roida.rkh@bsi.ac.id [6]
**Penulis Korespondensi: jojerde254@gmail.com*

Abstract. *Stroke is a major cardiovascular disease that significantly contributes to global mortality and disability rates. Early detection through stroke risk prediction is essential in reducing its impact. This study focuses on evaluating and comparing the performance of three machine learning algorithms—Support Vector Machine (SVM), Naive Bayes (NB), and Gradient Boosting (GB)—in predicting stroke occurrence. The research utilizes the Healthcare Stroke Dataset, which contains 5,109 records and 11 predictor variables. Modeling was performed using Orange Data Mining software, with 70% of the data allocated for training and 30% for testing. The results show that the SVM algorithm achieved the highest performance, obtaining an AUC score of 0.919 and an accuracy of 96.0%, followed by Gradient Boosting with an AUC of 0.885 and accuracy of 95.2%, and Naive Bayes with an AUC of 0.803 and accuracy of 88.2%. Therefore, SVM is identified as the most effective algorithm for predicting stroke risk within this dataset.*

Keywords: *Gradient Boosting; Machine Learning; Naive Bayes; Stroke; Support Vector Machine.*

Abstrak. Penyakit stroke termasuk dalam kategori kardiovaskular dengan angka kematian serta kecacatan yang masih tinggi secara globa. Deteksi dini melalui prediksi risiko stroke berperan penting dalam upaya pencegahan dan pengurangan dampak penyakit ini. Penelitian ini bertujuan untuk mengevaluasi serta membandingkan kinerja tiga algoritma machine learning, yaitu *Support Vector Machine (SVM)*, *Naive Bayes*, dan *Gradient Boosting*, dalam memprediksi kemungkinan terjadinya stroke. Dataset yang digunakan adalah Healthcare Stroke Dataset yang terdiri dari 5.109 data dengan 11 variabel prediktor. Proses pemodelan dilakukan menggunakan perangkat lunak Orange Data Mining, dengan pembagian data sebesar 70% untuk pelatihan dan 30% untuk pengujian. Hasil penelitian menunjukkan bahwa algoritma SVM memberikan performa terbaik dengan nilai AUC sebesar 0,919 dan akurasi 96,0%, diikuti oleh *Gradient Boosting* dengan AUC 0,885 dan akurasi 95,2%, serta *Naive Bayes* dengan AUC 0,803 dan akurasi 88,2%. Berdasarkan hasil tersebut, *SVM* dinyatakan sebagai metode yang paling efektif dalam memprediksi risiko stroke pada dataset yang digunakan.

Kata Kunci: Mesin Vektor Pendukung; Naive Bayes; Pembelajaran Mesin; Peningkatan Gradien; Stroke.

1. LATAR BELAKANG

Stroke termasuk penyakit tidak menular yang memberikan kontribusi besar terhadap tingginya angka kematian dan kecacatan di seluruh dunia. Menurut laporan *World Health Organization (WHO)*, terdapat sekitar 15 juta kasus baru setiap tahunnya, dan sebagian besar penderitanya mengalami gangguan pada fungsi motorik maupun kognitif. Di Indonesia sendiri, stroke juga menempati posisi teratas sebagai penyebab kematian dan kecacatan jangka panjang pada kelompok usia dewasa dan produktif. Penyakit ini tidak hanya berdampak pada individu yang terkena, tetapi juga menimbulkan beban sosial dan ekonomi bagi keluarga serta sistem kesehatan secara keseluruhan (M & Subbulakshmi, 2024; Melnykova et al., 2025)

Berbagai faktor risiko dapat meningkatkan peluang seseorang mengalami stroke, antara lain hipertensi, kadar gula darah yang tidak stabil, obesitas, kebiasaan merokok,

Naskah Masuk: 16 Juni 2025; Revisi: 05 Agustus 2025; Diterima: 23 September 2025; Tersedia: 27 November 2025.

dan pola hidup yang kurang sehat. Oleh karena itu, deteksi dini terhadap risiko stroke menjadi langkah krusial untuk menurunkan angka kejadian serta tingkat keparahan penyakit ini (Mizwar, Rahim, Sunyoto, & Arief, 2022; Sari et al., 2025)

Meskipun demikian, metode diagnosis konvensional masih memiliki sejumlah keterbatasan, seperti waktu pemeriksaan yang cukup lama, biaya yang mahal, serta ketergantungan pada tenaga medis yang berpengalaman. Karena itu, diperlukan pendekatan alternatif berbasis teknologi data yang dapat mempercepat proses identifikasi risiko stroke secara lebih efisien (Guhdar, Ismail Melhum, & Luqman Ibrahim, 2023; Hendro Martono & Sulistianingsih, 2024)

Kemajuan teknologi *machine learning* membuka peluang besar dalam bidang kesehatan, khususnya untuk memprediksi penyakit secara otomatis melalui analisis data pasien. Dengan memanfaatkan data historis, algoritma *machine learning* mampu mengidentifikasi pola-pola yang berkaitan dengan risiko stroke, sehingga dapat mendukung pengambilan keputusan medis secara lebih cepat dan akurat (Asadi, Rahimi, Daeechini, & Paghe, 2024; Muhamad Indra, Siti Ernawati, & Ilham Maulana, 2024)

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk mengevaluasi dan membandingkan kinerja tiga algoritma pembelajaran mesin, yaitu *Support Vector Machine (SVM)*, *Naïve Bayes (NB)*, dan *Gradient Boosting (GB)* dalam melakukan prediksi risiko stroke dengan menggunakan *Healthcare Stroke Dataset*. Analisis dilakukan dengan bantuan perangkat lunak *Orange Data Mining* untuk menilai performa tiap model berdasarkan nilai *Area Under Curve (AUC)* dan akurasi. Temuan dari penelitian ini diharapkan dapat menjadi dasar bagi pengembangan sistem prediksi risiko stroke yang lebih efisien dan tepat guna di masa mendatang (M & Subbulakshmi, 2024; Mizwar et al., 2022).

2. KAJIAN TEORITIS

Stroke adalah gangguan pada aliran darah menuju otak yang mengakibatkan kerusakan jaringan karena kekurangan suplai oksigen. Penyakit ini merupakan salah satu penyebab utama kematian dan kecacatan jangka panjang di seluruh dunia. Menurut laporan **World Health Organization (WHO)**, lebih dari 15 juta orang mengalami stroke setiap tahunnya, dan sebagian besar di antaranya mengalami gangguan serius pada fungsi motorik maupun kognitif. Sejumlah faktor seperti tekanan darah tinggi, kadar gula darah yang tidak stabil, obesitas, kebiasaan merokok, serta gaya hidup tidak sehat menjadi pemicu utama meningkatnya risiko terjadinya stroke (Pasiolo et al., 2025). Oleh karena itu, kemampuan untuk melakukan deteksi

dini terhadap potensi stroke menjadi hal yang sangat penting guna mencegah komplikasi serius (Rahmawati, Hafid, & Sunandar, 2023)

Perkembangan teknologi informasi, terutama di bidang *machine learning*, telah memberikan pengaruh besar terhadap dunia kedokteran.(Handayani, Science, & 2024, 2024) Teknologi *machine learning* memungkinkan komputer mempelajari data pasien dan menemukan pola tersembunyi yang sulit dikenali secara manual oleh manusia. Pendekatan ini dapat mempercepat proses diagnosis penyakit secara lebih efisien dan akurat tanpa sepenuhnya bergantung pada metode konvensional yang umumnya membutuhkan waktu serta biaya tinggi. Selain itu, teknologi ini juga berperan sebagai alat bantu bagi tenaga medis dalam mengambil keputusan yang lebih objektif dan berbasis data.

Dalam upaya memprediksi risiko terjadinya stroke, berbagai algoritma *machine learning* telah banyak diterapkan, dan masing-masing menghasilkan performa yang berbeda.. (Ayuningtyas, And, & 2023, 2023) menyatakan bahwa *Support Vector Machine (SVM)* bekerja dengan mencari *hyperplane* paling optimal untuk memisahkan antar kelas data. Metode ini dinilai efektif dalam menangani data berdimensi tinggi, meskipun performanya dapat menurun ketika berhadapan dengan data yang tidak seimbang. Sementara itu, menurut (Rahmawati et al., 2023) *Naive Bayes* menggunakan pendekatan probabilistik berdasarkan *Teorema Bayes*, dan dianggap sesuai untuk dataset dengan banyak fitur serta pola distribusi yang relatif sederhana. Adapun metode *Gradient Boosting*, sebagaimana dijelaskan oleh (Handayani et al., 2024) mengombinasikan beberapa model lemah (*weak learners*) guna meningkatkan akurasi prediksi, serta terbukti efektif dalam mengolah data non-linear.

Hasil-hasil penelitian terdahulu menunjukkan bahwa performa ketiga algoritma tersebut dapat ditingkatkan melalui berbagai pendekatan optimasi. Misalnya, penggunaan teknik SMOTE mampu meningkatkan performa SVM pada dataset yang tidak seimbang (Pasiolo et al., 2025), sedangkan penerapan *Particle Swarm Optimization (PSO)* terbukti efektif dalam mengoptimalkan parameter model (Ayuningtyas et al., 2023). Di sisi lain, *Gradient Boosting* yang dikombinasikan dengan *Grid Search* menunjukkan kestabilan performa pada data medis (Handayani et al., 2024)

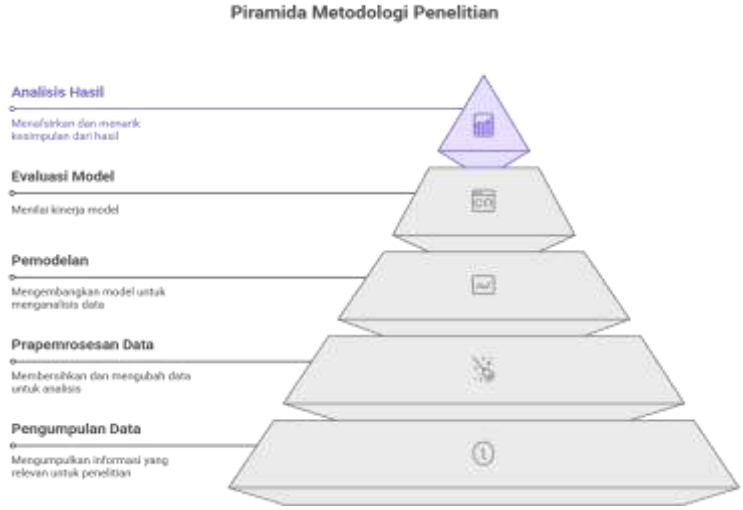
Penilaian kinerja model *machine learning* umumnya dilakukan dengan memanfaatkan beberapa metrik evaluasi, seperti Accuracy, Precision, Recall, F1-Score, dan Area Under Curve (AUC). Metrik accuracy digunakan untuk mengukur tingkat ketepatan keseluruhan prediksi model, sedangkan precision dan recall menilai keseimbangan antara ketepatan hasil prediksi serta kemampuan model dalam mendeteksi data positif. F1-Score memberikan gambaran menyeluruh mengenai performa model, terutama saat menghadapi data

yang tidak seimbang, sementara AUC berfungsi untuk menilai sejauh mana model mampu membedakan antara kelas positif dan negatif. Melalui penggunaan metrik-metrik tersebut, peneliti dapat menentukan algoritma yang memiliki performa paling optimal dalam memprediksi risiko stroke dengan akurat dan efisien.

3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif eksperimental dengan tujuan membandingkan performa beberapa algoritma *machine learning* di bidang medis. Metode ini dipilih karena dapat menilai kinerja model secara objektif melalui indikator statistik, sehingga memungkinkan perbandingan langsung antaralgoritma. Pendekatan komparatif digunakan untuk menilai efektivitas tiga algoritma, yakni *Support Vector Machine (SVM), Naïve Bayes* , dan *Gradient Boosting* , dalam memprediksi risiko terjadinya stroke pada pasien (Azhar, Firdausy, & Amelia, 2022; Rahmawati et al., 2023)

Penggunaan Orange Data Mining dalam penelitian ini dipilih karena aplikasi tersebut memiliki antarmuka visual yang mudah digunakan serta mendukung proses machine learning secara efisien. Secara umum tahapan proses penelitian ditujukan pada (Gambar1) berikut:



Gambar 1. Diagram Alur penelitian.

Populasi dan Sampel Penelitian

Populasi dalam penelitian ini mencakup seluruh data pasien yang terdapat pada . Dataset ini berisi 5.109 data pasien dengan 11 variabel prediktor yang mencakup data demografis dan klinis seperti usia, jenis kelamin, hipertensi, kadar glukosa, indeks massa tubuh (BMI), kebiasaan merokok, dan status perkawinan.

Seluruh data pada dataset tersebut digunakan sebagai sampel penelitian dengan metode total sampling, karena setiap entri dianggap memiliki potensi representatif terhadap populasi pasien stroke.

Sumber dan Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang tersedia secara daring melalui repositori Kaggle. Dataset tersebut banyak dimanfaatkan dalam penelitian di bidang medis karena memiliki format yang sesuai untuk penerapan metode klasifikasi berbasis machine learning. Proses pengumpulan data dilakukan dengan mengunduh dataset dari Kaggle, kemudian mengimpornya ke dalam Orange Data Mining menggunakan *File Widget*, sebagaimana juga dilakukan oleh (Handayani et al., 2024) dalam studi klasifikasi stroke berbasis *Gradient Boosting*.

Tahapan Penelitian

Penelitian ini dilakukan secara sistematis melalui lima tahapan utama yang dilakukan di dalam lingkungan Orange Data Mining, yaitu sebagai berikut:

Pengumpulan Data (Data Collection)

Dataset Healthcare Stroke diunduh dari situs Kaggle dan diimpor ke Orange menggunakan File Widget. Data ini berisi informasi mengenai pasien dengan atau tanpa riwayat stroke yang digunakan sebagai dasar untuk pelatihan model klasifikasi.

Prapemrosesan Data (Data Preprocessing)

Tahap prapemrosesan data dilakukan untuk mempersiapkan dataset sebelum model dilatih. Langkah ini mencakup pembersihan data dari nilai yang kosong, penghapusan data duplikat, serta mengubah variabel kategorikal menjadi format numerik agar dapat diproses oleh algoritma machine learning.(Pasiolo et al., 2025) Menjelaskan bahwa teknik *One-Hot Encoding* digunakan secara efektif untuk mengubah data kategorikal agar dapat dikenali oleh algoritma machine learning. Selain itu, proses normalisasi pada atribut numerik seperti usia, BMI, dan kadar glukosa juga penting dilakukan supaya setiap variabel memiliki skala yang seimbang, sehingga model dapat belajar dengan lebih optimal.

Pemodelan (Modeling)

Tahapan ini merupakan inti dari penelitian, di mana tiga algoritma machine learning diterapkan untuk membangun model prediksi risiko stroke, yaitu:

a. Support Vector Machine (SVM): bekerja dengan mencari hyperplane terbaik yang memisahkan dua kelas data secara optimal. Kernel yang digunakan adalah RBF (Radial Basis Function).
b. Naïve Bayes (NB): menggunakan prinsip probabilitas berdasarkan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen. Model ini dipilih karena kemampuannya yang cepat dan efisien.
c. Gradient Boosting (GB): merupakan metode ensemble yang membangun beberapa model decision tree secara bertahap. Setiap pohon memperbaiki kesalahan dari pohon sebelumnya. Parameter yang digunakan dalam penelitian ini yaitu n_estimators = 300, learning_rate = 0.005, dan max_depth = 3.

(Wisesty, Wirayuda, Sthevanie, & Rismala, 2024). Menunjukkan bahwa proses pengolahan fitur dan pemilihan algoritma yang sesuai dapat secara signifikan meningkatkan performa model pada data medis.Selain itu,(Ari Wibowo et al., 2024) menyebutkan bahwa kombinasi parameter yang optimal pada model *Gradient Boosting* mampu meningkatkan nilai akurasi secara konsisten pada dataset biomedis.

Evaluasi Model (Model Evaluation)

Evaluasi model dilakukan menggunakan metode 10-Fold Cross Validation, Merupakan teknik standar yang digunakan untuk menguji keandalan model dengan cara membagi data ke dalam beberapa subset. Metrik evaluasi yang digunakan mencakup Accuracy, Precision, Recall, F1-Score, dan AUC.Penelitian oleh (Sebastian & Juliane, 2023) Juga menerapkan metode serupa dalam prediksi stroke serta menekankan pentingnya nilai AUC sebagai ukuran kemampuan model dalam membedakan antara kelas positif dan negatif.

a. Accuracy: tingkat ketepatan model dalam memprediksi kelas data.
b. Precision: menunjukkan seberapa tepat model dalam memprediksi data positif.
c. Recall (Sensitivity): mengukur kemampuan model dalam mengenali seluruh data positif.
d. F1-Score: rata-rata harmonis antara precision dan recall.
e. AUC (Area Under Curve): mengukur kemampuan model membedakan antara kelas positif dan negatif.5. Analisis Hasil (Result Interpretation)

Hasil perbandingan dari ketiga algoritma menunjukkan variasi performa berdasarkan metrik evaluasi. Analisis ini dilakukan dengan bantuan *Confusion Matrix* untuk mengetahui kesalahan klasifikasi.Studi oleh (Herlistiono & Violina, 2024) Menyoroti pentingnya melakukan analisis terhadap *False Positive* dan *False Negative* dalam konteks diagnosis penyakit,sementara (Ananda Setyarini et al., 2024)

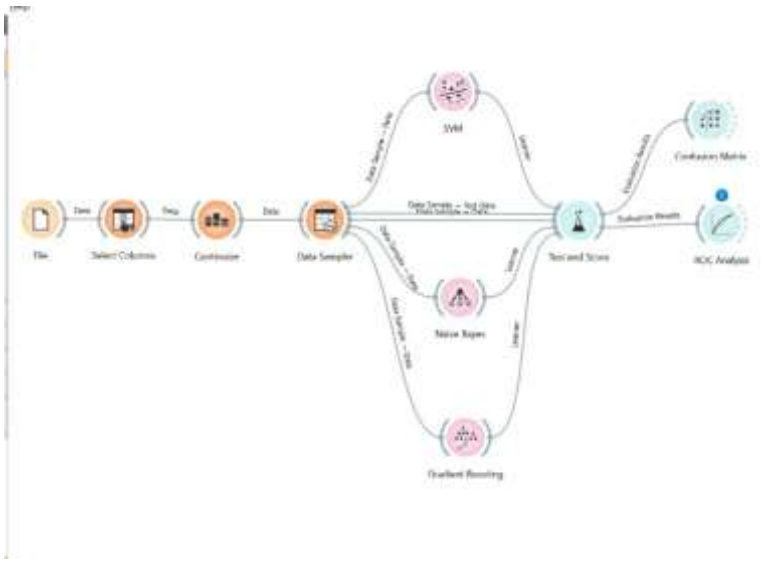
Selain itu, dilakukan analisis terhadap Confusion Matrix untuk mengetahui kesalahan klasifikasi (False Positive dan False Negative). Hasil analisis ini digunakan untuk menentukan algoritma terbaik dalam mendeteksi kemungkinan terjadinya stroke.

4. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil penelitian yang diperoleh dari proses analisis data menggunakan aplikasi Orange Data Mining, serta membahas keterkaitannya dengan teori dan penelitian terdahulu. Penelitian dilakukan dengan menerapkan tiga algoritma machine learning, yaitu Support Vector Machine (SVM), Naïve Bayes (NB), dan Gradient Boosting (GB) untuk memprediksi kemungkinan terjadinya penyakit stroke berdasarkan dataset Healthcare Stroke Data yang bersumber dari Kaggle. Dataset tersebut berisi data pasien dengan berbagai variabel prediktor seperti usia, jenis kelamin, riwayat hipertensi, kadar glukosa rata-rata, indeks massa tubuh (BMI), dan kebiasaan merokok.

Hasil Pengujian Model

Penelitian ini menggunakan aplikasi Orange Data Mining untuk membandingkan performa tiga algoritma machine learning, yaitu Support Vector Machine (SVM), Naïve Bayes (NB), dan Gradient Boosting (GB) dalam memprediksi kemungkinan terjadinya penyakit stroke. Dataset yang digunakan berasal dari Healthcare Stroke Dataset di platform Kaggle yang terdiri dari 5.109 data pasien dengan 11 atribut seperti usia, jenis kelamin, tekanan darah, kadar glukosa, serta indeks massa tubuh (BMI) yang merepresentasikan faktor-faktor risiko stroke.



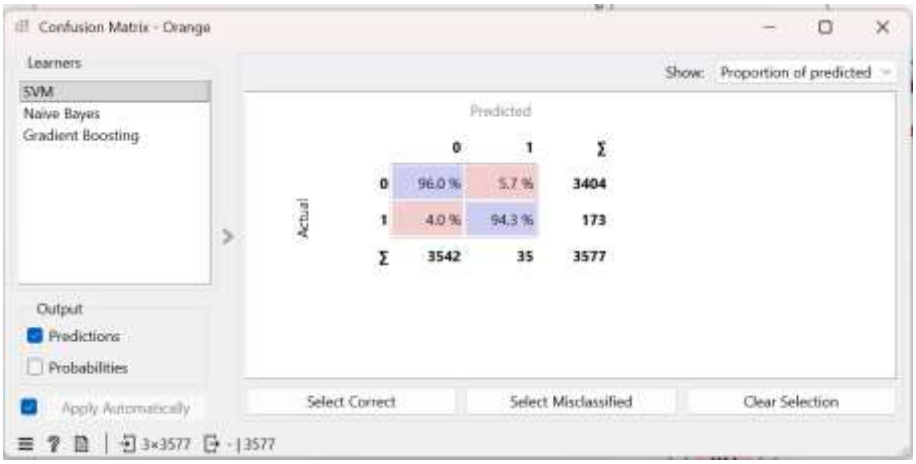
Gambar 2. Workflow Analisis Data Menggunakan Orange Data Mining.

Proses analisis dilakukan melalui beberapa tahapan. Data terlebih dahulu dimasukkan menggunakan File Widget, kemudian atribut-atribut penting dipilih melalui Select Columns. Selanjutnya, data kategorikal diubah menjadi bentuk numerik menggunakan Continuize dengan metode One-Hot Encoding agar bisa diproses oleh model. Data tersebut kemudian dibagi menjadi 70% data latih dan 30% data uji menggunakan Data Sampler. Setelah tahap ini, ketiga algoritma dijalankan secara bersamaan dan dievaluasi menggunakan metode 10-Fold Cross Validation untuk memastikan hasil pengujian tidak bias terhadap satu subset data tertentu.

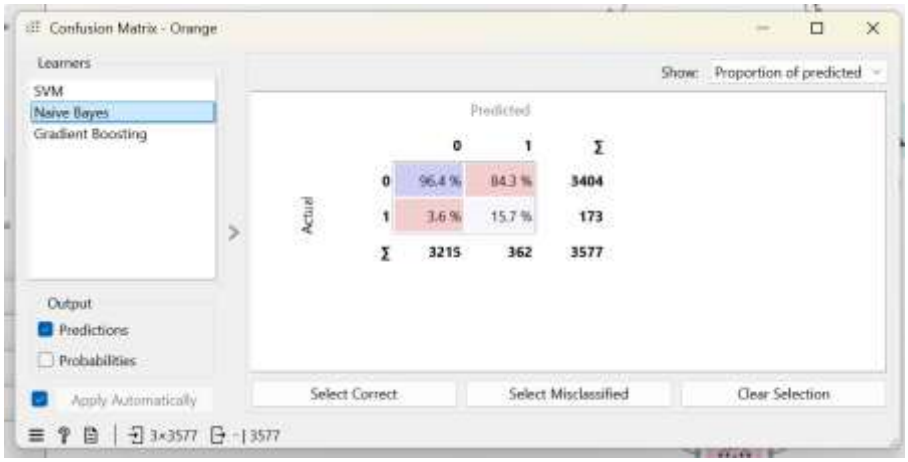
Tabel 1. Hasil Evaluasi Kinerja Model

Model	AUC	CA	F1	Precision	Recall	MCC
SVM	0.919	0.960	0.948	0.960	0.960	0.414
Naïve Bayes	0.803	0.882	0.901	0.925	0.882	0.171
Gradient Boosting	**0.885**	**0.952**	**0.929**	**0.954**	**0.952**	**0.105**

Hasil pengujian pada tabel di atas menunjukkan bahwa SVM memiliki performa paling tinggi dengan akurasi 0,960 dan nilai AUC 0,919, yang menandakan kemampuan model dalam membedakan antara data pasien stroke dan non-stroke secara akurat. Algoritma Gradient Boosting menempati urutan kedua dengan akurasi 0,952, sedangkan Naïve Bayes menempati posisi terakhir dengan akurasi 0,882 dan AUC 0,803, meskipun masih menunjukkan performa yang cukup baik.



Gambar 3. Confusion Matrix Support Vector Machine (SVM).

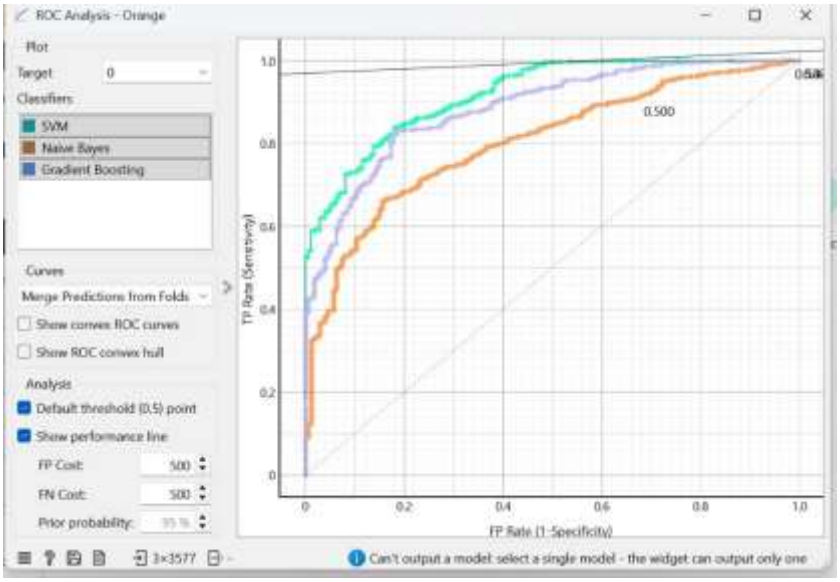


Gambar 4. Confusion Matrix Naive Bayes.

Gambar 5. Confusion Matrix Gradient Boosting.

Berdasarkan Confusion Matrix, algoritma SVM menunjukkan jumlah klasifikasi benar paling tinggi dengan kesalahan prediksi yang sangat kecil. Hal ini memperlihatkan bahwa SVM lebih mampu mengenali pola data yang kompleks dibandingkan dua algoritma lainnya. Sementara itu, Naïve Bayes cenderung mengalami kesalahan prediksi karena sifatnya yang

mengasumsikan independensi antar fitur, yang pada data medis sering kali tidak terpenuhi secara sempurna.



Gambar 6. Kurva ROC Perbandingan Ketiga Algoritma.

Hasil kurva ROC memperlihatkan bahwa garis ROC milik SVM berada paling dekat dengan sudut kiri atas grafik, yang menandakan kemampuan model paling baik dalam membedakan antara kelas positif dan negatif. Secara keseluruhan, SVM memberikan hasil paling akurat dan stabil, diikuti oleh Gradient Boosting dan Naïve Bayes. Temuan ini membuktikan bahwa SVM merupakan algoritma yang paling efektif untuk menganalisis data kesehatan dengan karakteristik non-linear seperti pada kasus prediksi stroke.

5. KESIMPULAN DAN SARAN

Kesimpulan

Kesimpulan ditulis secara singkat yaitu mampu menjawab tujuan atau permasalahan penelitian dengan menunjukkan hasil penelitian atau pengujian hipotesis penelitian, **tanpa** mengulang pembahasan. Kesimpulan ditulis secara kritis, logis, dan jujur berdasarkan fakta hasil penelitian yang ada, serta penuh kehati-hatian apabila terdapat upaya generalisasi. Bagian kesimpulan dan saran ini ditulis dalam bentuk paragraf, tidak menggunakan penomoran atau *bullet.* Pada bagian ini juga dimungkinkan apabila penulis ingin memberikan saran atau rekomendasi tindakan berdasarkan kesimpulan hasil penelitian. Demikian pula, penulis juga sangat disarankan untuk memberikan ulasan terkait keterbatasan penelitian, serta rekomendasi untuk penelitian yang akan datang.

Saran

Penelitian ini masih dapat dikembangkan dengan menggunakan dataset yang lebih besar dan bervariasi agar hasil prediksi semakin akurat. Penggunaan algoritma tambahan seperti Random Forest atau Logistic Regression juga dapat menjadi pembanding yang menarik dalam studi selanjutnya. Selain itu, penerapan optimasi parameter seperti Grid Search dapat meningkatkan kinerja model, khususnya pada SVM dan Gradient Boosting. Hasil penelitian ini diharapkan dapat dimanfaatkan untuk membangun sistem pendukung keputusan yang membantu tenaga medis dalam mendeteksi potensi penyakit stroke secara lebih cepat dan tepat.

DAFTAR REFERENSI

Ananda Setyarini, D., Ayu Maharani Dyah Gayatri, A., Sri Kusuma Aditya, C., Rizki Chandranegara, D., Setyarini, D. A., M D Gayatri, A. A., & K Aditya, C. S. (2024). Stroke prediction with enhanced gradient boosting classifier and strategic hyperparameter. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer, 23*(2), 477-490. https://doi.org/10.30812/matrik.v23i2.3555

Ari Wibowo, S., Andriani, W., Informatika, T., YMI Tegal Jl Pendidikan No, S., Tegal, K., & Tengah, J. (2024). Evaluasi model deep learning pada pola dataset biomedis. *Jurnal Saintekom: Sains, Teknologi, Komputer Dan Manajemen, 14*(2), 195-207. https://doi.org/10.33020/saintekom.v14i2.738

Asadi, F., Rahimi, M., Daeechini, A. H., & Paghe, A. (2024). The most efficient machine learning algorithms in stroke prediction: A systematic review. *Health Science Reports, 7*(10), e70062. https://doi.org/10.1002/hsr2.70062

Ayuningtyas, Y., & And, I. S.-J. of I. (2023). Klasifikasi penyakit stroke menggunakan support vector machine (SVM) dan particle swarm optimization (PSO). *Ejournal.Unesa.Ac.Id*. Retrieved from https://ejournal.unesa.ac.id/index.php/jinacs/article/view/54156

Azhar, Y., Firdausy, A. K., & Amelia, P. J. (2022). Perbandingan algoritma klasifikasi data mining untuk prediksi penyakit stroke. *SINTECH (Science and Information Technology) Journal, 5*(2), 191-197. https://doi.org/10.31598/sintechjournal.v5i2.1222

Guhdar, M., Ismail Melhum, A., & Luqman Ibrahim, A. (2023). Optimizing accuracy of stroke prediction using logistic regression. *Journal of Technology and Informatics (JoTI), 4*(2), 41-47. https://doi.org/10.37802/joti.v4i2.278

Handayani, F., & Taufiq, R. M. (2024). Komparasi algoritma menggunakan teknik SMOTE dalam melakukan klasifikasi penyakit stroke otak. *Ejurnal.Umri.Ac.Id*, *5*(2), 367-372. https://doi.org/10.37859/coscitech.v5i2.7439

Hendro Martono, G., & Sulistianingsih, N. (2024). Enhancing stroke diagnosis with machine learning and SHAP-based explainable AI models. *Knowbase: International Journal of Knowledge in Database, 04*(02), 189-203.

Herlistiono, I. O., & Violina, S. (2024). Model prediksi risiko stroke menggunakan machine learning. *INTECOMS: Journal of Information Technology and Computer Science, 7*(4), 1230-1238. https://doi.org/10.31539/intecoms.v7i4.10942

M, S. L. J., & Subbulakshmi, P. (2024). Unveiling the potential of machine learning approaches in predicting the emergence of stroke at its onset: A predicting framework. *Scientific Reports*, 1-21. https://doi.org/10.1038/s41598-024-70354-1

Melnykova, N., Patereha, Y., Skopivskyi, S., Farion, M., Fedushko, S., & Drohomyretska, K. (2025). Machine learning for stroke prediction using imbalanced data. *Scientific Reports, 15*(1), 1-20. https://doi.org/10.1038/s41598-025-01855-w

Mizwar, A., Rahim, A., Sunyoto, A., & Arief, M. R. (2022). Stroke prediction using machine learning method with extreme gradient boosting algorithm. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan*, 21(3), 595-606. https://doi.org/10.30812/matrik.v21i3.1666

Muhamad Indra, Siti Ernawati, & Ilham Maulana. (2024). Machine learning for stroke prediction: Evaluating the effectiveness of data balancing approaches. *Jurnal Riset Informatika, 6*(4), 211-222. https://doi.org/10.34288/jri.v6i4.344

Pasiolo, L., Afrianty, I., Budianita, E., Abdillah, R., Studi, P., Informatika, T., & Baru, S. (2025). Penerapan teknik SMOTE pada klasifikasi penyakit stroke dengan algoritma support vector machine. *ZONAsi: Jurnal Sistem Informasi, 7*(1). Retrieved from https://journal.unilak.ac.id/index.php/zn/article/download/24731/7189

Rahmawati, L., Hafid, M., & Sunandar, M. A. (2023). Analisis data mining untuk memprediksi penyakit stroke dengan algoritma Naïve Bayes. *Jurnal Aplikasi Dan Teori Ilmu Komputer, 6*(2), 55-60. https://doi.org/10.17509/jatikom.v6i2.49051

Sari, K., Fadli, M., Fahmi Fudholi, M., Redy Susanto, E., Ilmu Komputer, M., Teknokrat Indonesia, U., & Negeri Lampung, P. (2025). Deteksi dini stroke menggunakan machine learning. *INSOLOGI: Jurnal Sains dan Teknologi, 4*(4), 706-720. https://doi.org/10.55123/insologi.v4i4.5590

Sebastian, R., & Juliane, C. (2023). Comparison of data mining classification algorithms for stroke disease prediction using the SMOTE upsampling method. *JUITA: Jurnal Informatika, 11*(2), 311-321. https://doi.org/10.30595/juita.v11i2.17348

Wisesty, U. N., Wirayuda, T. A. B., Sthevanie, F., & Rismala, R. (2024). Analysis of data and feature processing on stroke prediction using wide range machine learning models. *Jurnal Online Informatika, 9*(1), 29-40. https://doi.org/10.15575/join.v9i1.1249