

Analisis Penyeimbangan Data pada Penyakit Stroke Menggunakan SMOTE

Muhammad Ubay Khoiruddin*, Nur Nafiiyah

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Lamongan, Lamongan, Indonesia

Email: ¹ubaykhoiruddin06@gmail.com, ²mynaff@unisla.ac.id

Email Penulis Korespondensi: ubaykhoiruddin06@gmail.com*

Submitted: 08/07/2025; Accepted: 14/07/2025; Published: 08/09/2025

Abstrak Stroke tidak hanya mempengaruhi kesehatan individu tetapi juga membebani sistem kesehatan nasional secara signifikan. Oleh karena itu, upaya deteksi dan pencegahan dini sangat penting. Penelitian ini bertujuan untuk membandingkan kinerja klasifikasi data penyakit stroke asli dan data berimbang menggunakan algoritma Support Vector Machine (SVM) dan Naive Bayes. Dataset yang digunakan adalah data publik dari Kaggle. Hasil klasifikasi pada data asli menunjukkan bahwa metode SVM dan Naive Bayes Bernoulli memiliki akurasi tertinggi sebesar 0,945. Namun, presisi tertinggi dan nilai recall masing-masing sebesar 0,584 dan 0,699 dicapai dengan metode Naive Bayes Gaussian. Sedangkan pada data berimbang, akurasi Naive Bayes Bernoulli turun menjadi 0,806, namun nilai presisi dan recall meningkat menjadi 0,808 dan 0,806. Hasil ini menunjukkan bahwa penyeimbangan data dapat meningkatkan kinerja klasifikasi dalam hal presisi dan sensitivitas.

Kata kunci: Menyamakan Kelas; Naive Bayes; Penyakit Stroke; SMOTE; SVM

Abstract Stroke not only affects individual health but also places a significant burden on the national health system. Therefore, early detection and prevention efforts are crucial. This study aims to compare the classification performance of original stroke disease data and balanced data using the Support Vector Machine (SVM) and Naive Bayes algorithms. The dataset used is public data from Kaggle. The classification results on the original data show that the SVM and Naive Bayes Bernoulli methods have the highest accuracy of 0.945. However, the highest precision and recall values of 0.584 and 0.699, respectively, were achieved by the Naive Bayes Gaussian method. Meanwhile, on the balanced data, the accuracy of Naive Bayes Bernoulli decreased to 0.806, but the precision and recall values increased to 0.808 and 0.806. These results indicate that data balancing can improve classification performance in terms of precision and sensitivity.

Keyword: Balanced Class; Naive Bayes; Stroke Disease; SMOTE; SVM

1. PENDAHULUAN

Stroke adalah penyebab utama orang meninggal di Indonesia, maupun di dunia karena salah satu penyakit kronis. Berdasarkan data dari Kementerian Kesehatan, prevalensi stroke di Indonesia mencapai 10,9% dari jumlah penduduk penyakit tidak menular. Penyakit ini dikarenakan oleh gangguan aliran darah ke otak akibat penyumbatan arteri, pecahnya pembuluh darah, atau kelainan dalam kondisi darah. Akibatnya, suplai oksigen dan nutrisi ke otak terhenti, menyebabkan kerusakan pada jaringan otak. Risiko stroke dapat dikategorikan menjadi dua kelompok: faktor yang tidak dapat diubah, meliputi usia, jenis kelamin, dan faktor genetik, serta faktor yang dapat diubah, meliputi hipertensi, diabetes, kebiasaan merokok, dan obesitas. Dampak stroke tidak hanya mencakup aspek kesehatan individu, tetapi juga memberikan beban besar bagi sistem kesehatan nasional. Oleh karena itu, upaya deteksi dini dan pencegahan penyakit ini menjadi sangat penting [1].

Seiring perkembangan teknologi, inovasi berbasis data telah banyak diterapkan untuk mendukung diagnosis dan pencegahan stroke. Salah satu pendekatan yang paling menjanjikan adalah teknologi klasifikasi data, yang memungkinkan analisis data pasien untuk mengidentifikasi individu dengan risiko tinggi stroke. Dengan proses klasifikasi, data yang mencakup berbagai variabel, seperti usia, indeks massa tubuh, kadar gula darah, dan kebiasaan hidup, dapat diolah menjadi informasi yang relevan untuk menentukan tingkat risiko stroke seseorang. Penelitian terkini menunjukkan bahwa penerapan teknologi machine learning [2], [3], [4] dan neural network [5] pada data kesehatan pasien mampu mencapai tingkat akurasi deteksi hingga 80%. Hal ini membuktikan bahwa teknologi memiliki potensi besar untuk membantu profesional kesehatan dalam menghadapi permasalahan medis yang kompleks. *Naive Bayes* adalah salah satu algoritma yang sering diterapkan dalam teknologi klasifikasi data, terutama di bidang kesehatan [6], [7], [8]. Metode Naive Bayes didasarkan pada teorema Bayes, yang memungkinkan perhitungan peluang berdasarkan data sebelumnya [9]. Keunggulan utama Naive Bayes adalah kesederhanaannya, namun tetap mampu memberikan hasil yang efektif, terutama dalam pengolahan dataset besar dengan asumsi independensi antar variabel [10], [11].

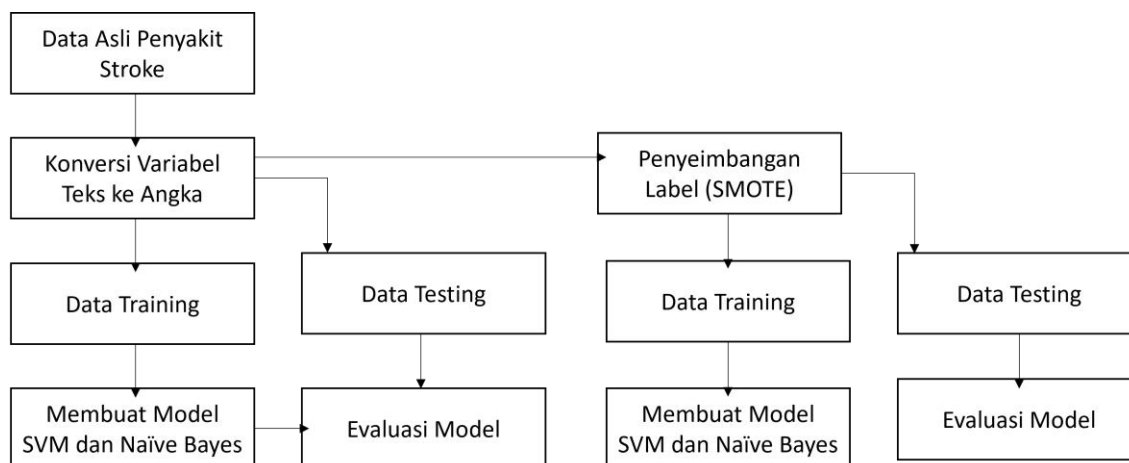
Beberapa penelitian melakukan perbandingan metode machine learning ataupun neural network dalam mengklasifikasi penyakit stroke, misalkan metode yang dibandingkan adalah Support Vector Machine dan Particle Swarm Optimization [12], [13]; *k*-Nearest Neighbor dan Naive Bayes [14], [15], [6]. Perbedaan penelitian ini dengan sebelumnya adalah penelitian ini membandingkan data asli dan setelah diseimbangkan kelas target, selanjutnya masing-masing data diklasifikasi menggunakan SVM dan Naive Bayes. Oleh karena itu penelitian ini membandingkan data penyakit stroke asli dan sudah diseimbangkan untuk diklasifikasi menggunakan SVM dan

Naive Bayes, berdasarkan penelitian [16] melakukan perbandingan SVM dan Naive Bayes dalam analisis sentimen.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian ini adalah: a) menelaah berbagai sumber ilmiah yang relevan, seperti jurnal ilmiah, dan buku-buku referensi. Tujuannya adalah untuk memperoleh landasan teori yang kuat dan memahami konteks permasalahan yang diteliti; b) setelah studi literatur, peneliti melakukan pengamatan dan analisis terhadap objek penelitian. Langkah ini bertujuan untuk mengidentifikasi permasalahan, yaitu dataset penyakit stroke kelas tidak seimbang, dan menganalisis data penyakit stroke yang sudah seimbang untuk klasifikasi penyakit stroke menggunakan machine learning; c) peneliti mengumpulkan data mentah dari *Kaggle* dengan judul “*Brain Stroke Prediction Dataset*”, yang dapat diakses melalui: <https://www.kaggle.com/datasets/zzettrkalkpakbal/full-filled-brain-stroke-dataset>. Data selanjutnya dianalisis klasifikasi menggunakan Naive Bayes dan SVM, kemudian diseimbangkan jumlah masing-masing kelas agar didapatkan perbedaan hasil akurasi; d) membandingkan analisis akurasi klasifikasi data penyakit stroke asli dan setelah diseimbangkan. Secara umum tahapan penelitian dalam Gambar 1.



Gambar 1. Alur Penelitian

Data penyakit stroke pada penelitian sebelumnya [17] menggunakan metode Naive Bayes. Penelitian ini mengusulkan menggunakan metode Naive Bayes Gaussian, Multinomial, dan Bernoulli, yang membedakan dengan penelitian [17] adalah penelitian ini membandingkan dengan data yang diseimbangkan dengan SMOTE seperti dalam penelitian [18], serta melakukan klasifikasi data penyakit stroke menggunakan Support Vector Machine (kernel linear, polynomial, dan Radial Basis Function) seperti penelitian [19]. Tujuan penelitian ini melakukan analisis algoritma SMOTE dalam menyeimbangkan data penyakit stroke dan mengklasifikasi data penyakit stroke menggunakan Naive Bayes dan SVM.

2.2 Rancangan Data

Dataset terdiri dari 4981 baris, dengan jumlah kelas 0 (normal) 4733 baris, dan kelas 1 (sakit) 248 baris serta terdapat sebelas kolom/atribut/variabel. Atribut dari data penyakit stroke adalah Gender, Age, Hypertension, Heart_Disease, Ever_Married, Work_Type, Residence_Type, Avg_Glucose_Level, BMI, dan Smoking_Status, serta satu variabel dependen sebagai variabel target, yaitu Stroke. Pembagian data penyakit stroke adalah 80% dari total data sebagai training dan 20% sebagai testing. Berdasarkan atribut dari data penyakit stroke ada yang isi dari kolom berupa data teks, sehingga perlu mengubah ke bentuk angka. Tabel 1 merupakan hasil konversi data teks ke angka.

Tabel 1. Hasil Konversi Data

No	Variabel	Sebelum	Sesudah
1	Gander	Female	0
		Male	1
2	Ever_married	No	0
		Yes	1

		Govt_Job	0
3	Work_type	Private	1
		Self_employed	2
		Children	3
4	Residence_Type	Rural	0
		Urban	1
		Unknown	0
5	Smoking_Status	Formerly_Smoked	1
		Never_Smoked	2
		Smokes	3

Tahap menyeimbangkan data kelas (penyakit stroke), metode untuk menyeimbangkan data variabel target adalah SMOTE (Synthetic Minority Oversampling Technique) [18] (Persamaan 1). Data hasil penyeimbangan dibagi 80% untuk membuat model/training, dan 20% untuk uji coba/testing. Hasil SMOTE total data berjumlah 9466 baris, dengan kelas 0 (normal) berjumlah 4733 baris, kelas 1 (sakit) berjumlah 4733 baris.

$$\Delta(x, y) = w_x w_y \sum \delta(x, y)^r \quad (1)$$

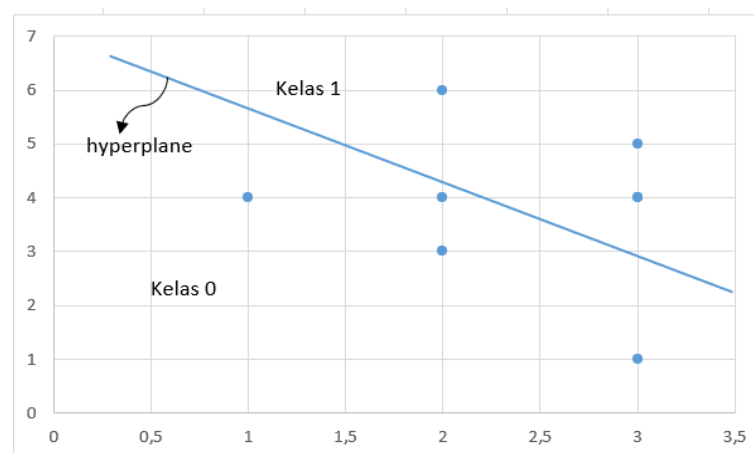
$\Delta(x, y)$ adalah jarak antar variabel x, y , w_x, w_y adalah bobot variabel x, y .

Metode klasifikasi Naive Bayes bekerja berdasarkan nilai probabilitas dari setiap atribut terhadap kelas target. Persamaan 2 adalah proses perhitungan probabilitas Naive Bayes [17]. $P(Y|A)$ merupakan probabilitas kelas target (y) terhadap atribut input (a). $P(A|Y)$ merupakan probabilitas atribut input data tes yang muncul terhadap kelas target. Naive Bayes Gaussian menghitung $P(A|Y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$. (2)

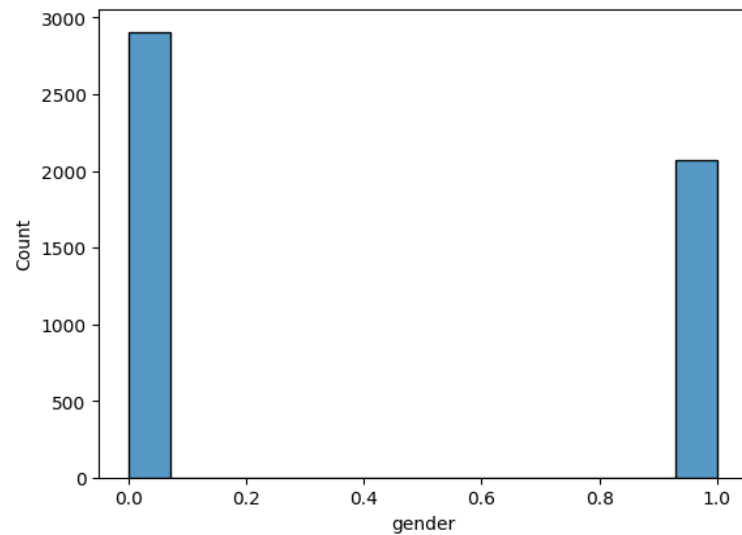
$$P(Y|A) = \frac{P(A|Y)P(Y)}{P(A)} \quad (3)$$

SVM merupakan metode klasifikasi berdasarkan hyperplane persamaan garis dari data training [20]. Cara mendapatkan persamaan garis hyperplane, yaitu: menentukan variabel/atribut input dan variabel target/output, setiap baris data melakukan perhitungan $\sum w_i \cdot x_i + b + t \geq 1$, w adalah bobot, x adalah variabel input, dan t adalah target/output, proses perhitungan ini berlangsung sampai mendapatkan hyperplane yang memisahkan kelas 0 dan 1 seperti Gambar 2. Evaluasi model menggunakan Akurasi (Persamaan 3).

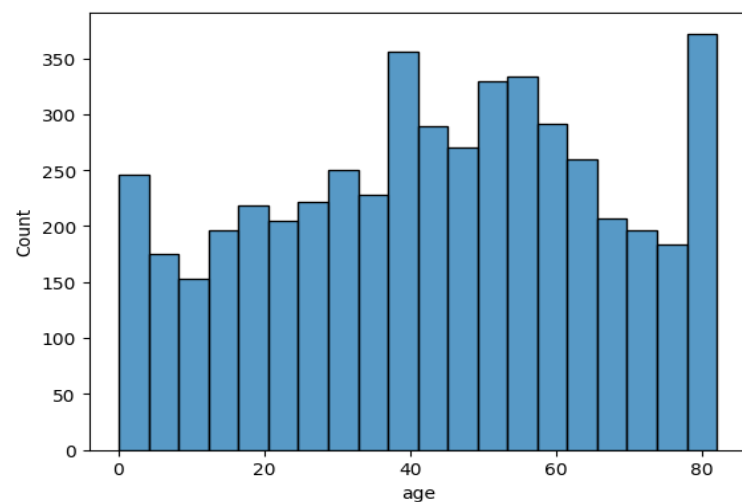
$$\text{Akurasi} = \frac{\text{True_Normal} + \text{True_Sakit}}{\text{True_Normal} + \text{False_Normal} + \text{False_Sakit} + \text{True_Sakit}} \quad (4)$$



Gambar 2. Ilustrasi Mencari Hyperplane

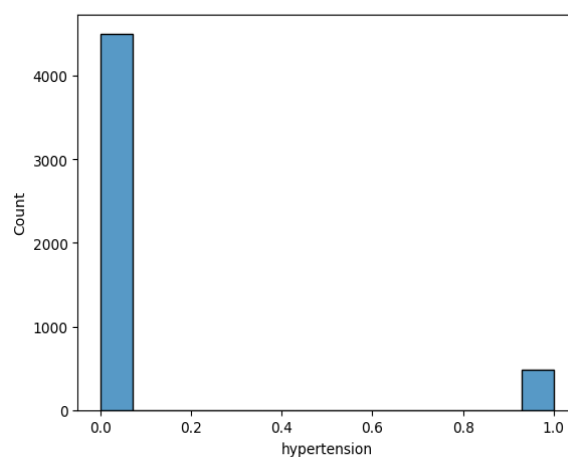


Gambar 3. Representasi Variabel Gender

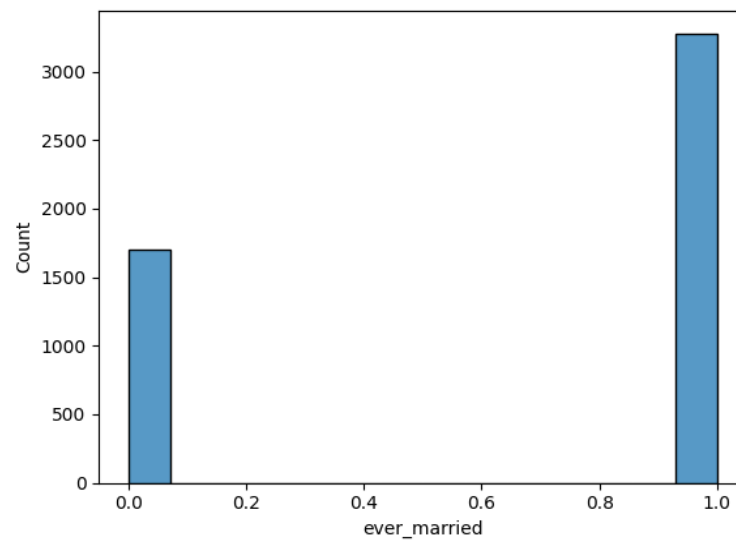


Gambar 4. Representasi Variabel Age

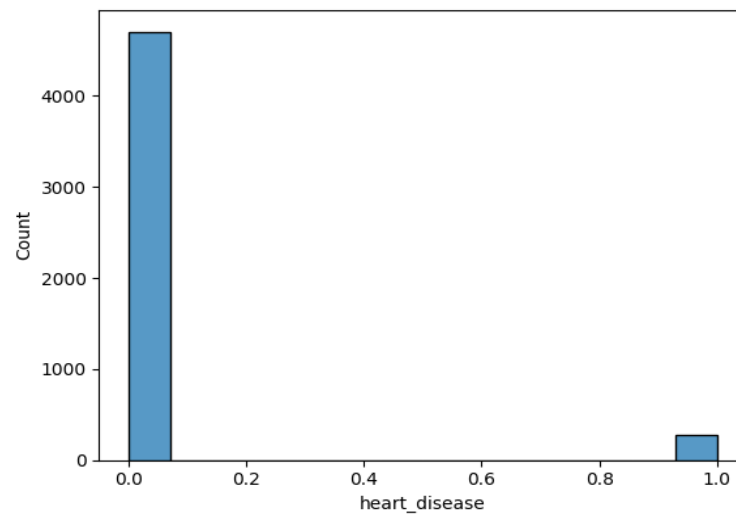
Gambar 3 sampai 12 merupakan representasi isi setiap variabel input penyakit stroke. Variabel *Gender* (Gambar 3), *Age* (Gambar 4), *Hypertension* (Gambar 5), *Ever_Meried* (Gambar 6), *Heart_Disease* (Gambar 7), *Work_Type* (Gambar 8), *Avg_Glucose_Level* (Gambar 9), *Residence_Type* (Gambar 10), *BMI* (Gambar 11), *Smoking_Status* (Gambar 12).



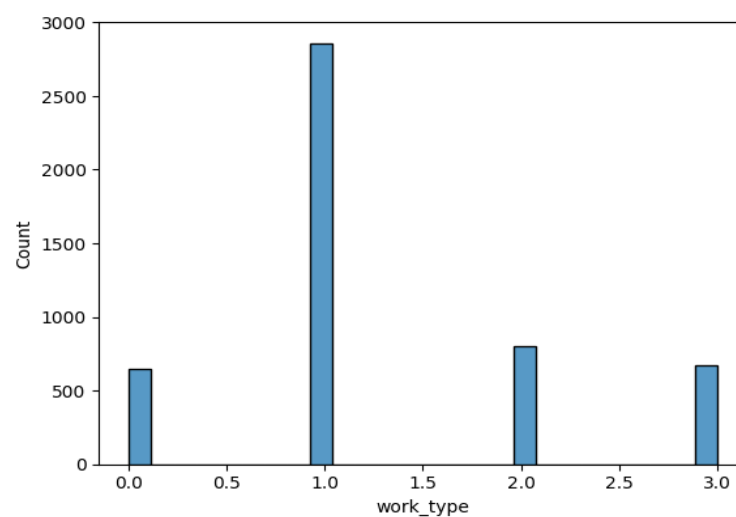
Gambar 5. Representasi Variabel Hypertension



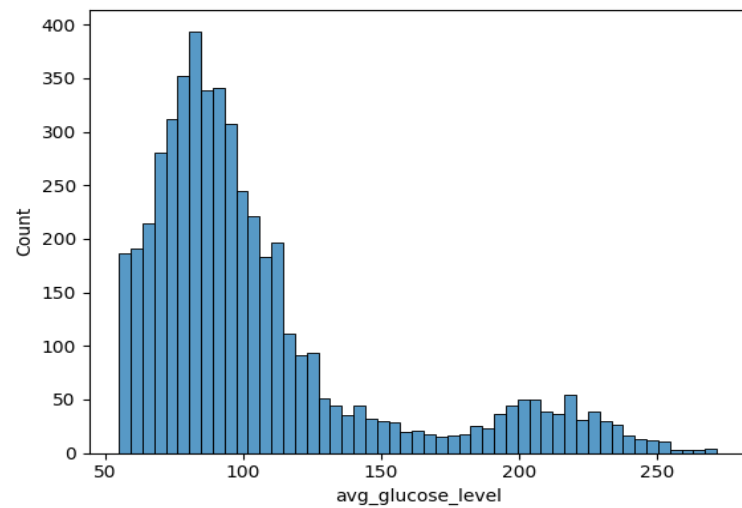
Gambar 6. Representasi Variabel Ever_Married



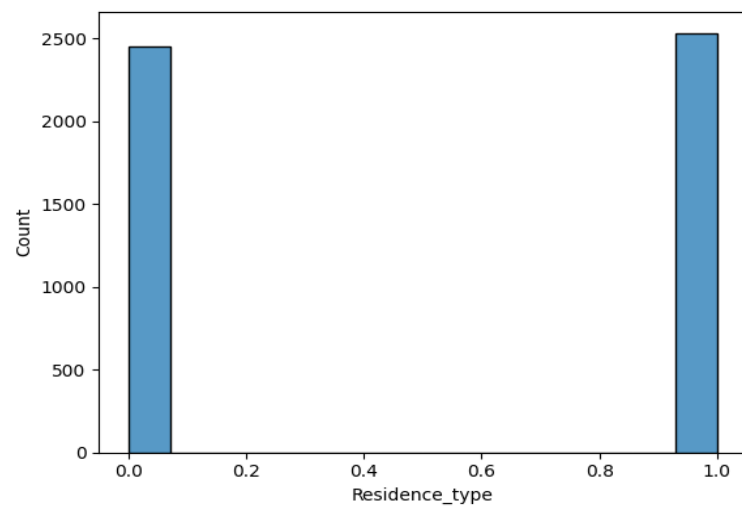
Gambar 7. Representasi Variabel Heart_Disease



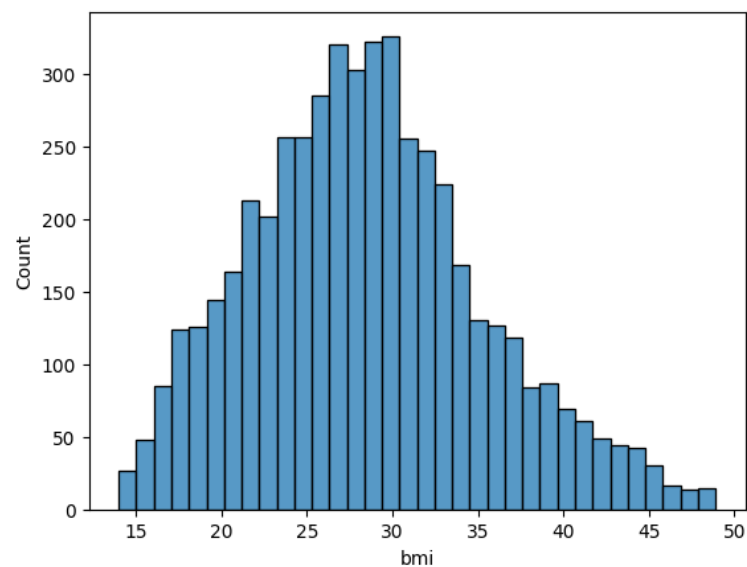
Gambar 8. Representasi Variabel Work_Type



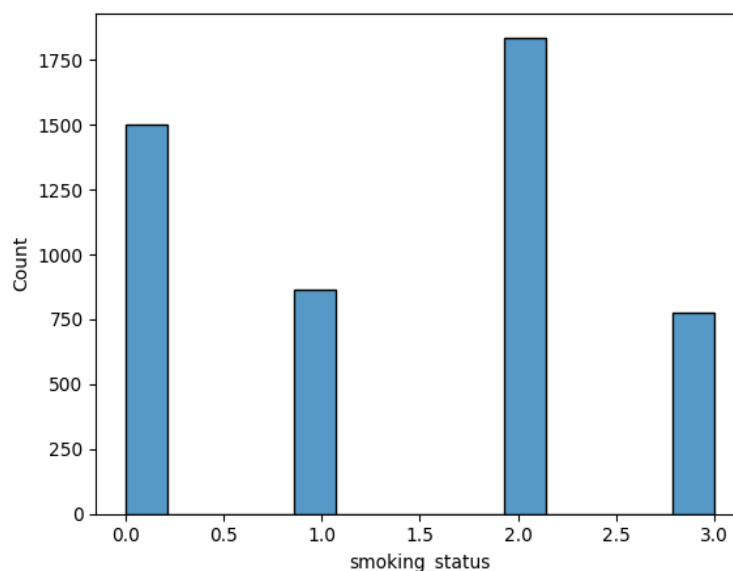
Gambar 9. Representasi Variabel Avg_Glucose_level



Gambar 10. Representasi Variabel Residence_type



Gambar 11. Representasi Variabel BMI



Gambar 12. Representasi Variabel Smoking_Status

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Penelitian ini melakukan pengujian klasifikasi data penyakit stroke menggunakan beberapa metode algoritma, yaitu Support Vector Machine (SVM) dengan tiga jenis kernel (linier, Radial Basis Function/RBF, dan polynomial), serta metode Naïve Bayes (Gaussian, Multinomial, dan Bernoulli). Pengujian dilakukan dalam dua skenario: (1) tanpa penyeimbangan data dan (2) dengan penyeimbangan data menggunakan metode Synthetic Minority Over-sampling Technique (SMOTE).

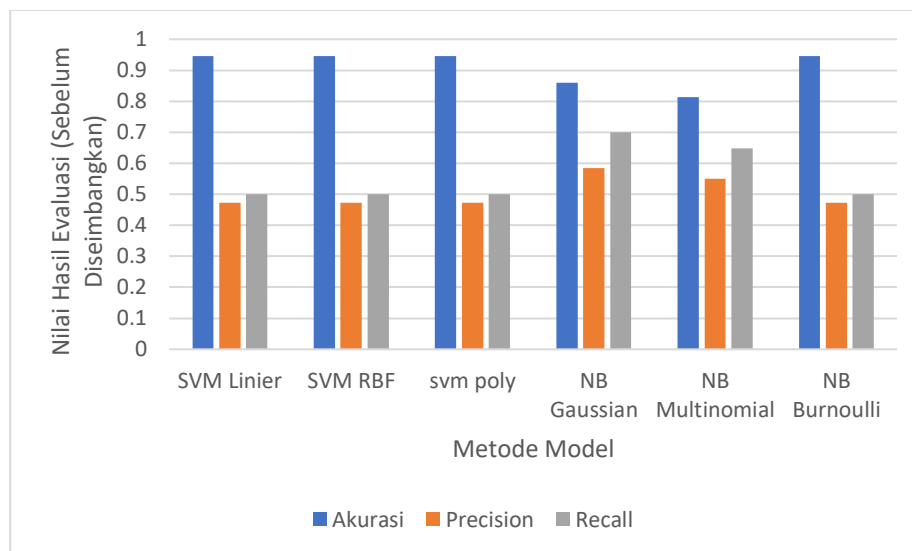
3.1.1 Hasil Uji Coba Tanpa Menyeimbangkan Data

Tabel 2. menunjukkan hasil uji akurasi terhadap data asli tanpa dilakukan penyeimbangan terlebih dahulu. Nilai akurasi, precision dan recall digunakan sebagai matrik evaluasi kinerja.

Tabel 2. Hasil Uji Data Asli

Model	Akurasi	Precision	Recall
SVM Linier	0,945	0,472	0,500
SVM RBF	0,945	0,472	0,500
SVM Polynomial	0,945	0,472	0,500
NB Gaussian	0,860	0,584	0,699
NB Multinomial	0,813	0,550	0,648
NB Bernoulli	0,945	0,472	0,500

Berdasarkan nilai recall, Naïve Bayes Gaussian memiliki performa terbaik dalam mengenali kelas stroke saat data belum diseimbangkan, meskipun nilai akurasinya tidak setinggi model lainnya. Hal ini menunjukkan bahwa akurasi yang tinggi pada data tidak seimbang belum tentu mencerminkan performa deteksi kelas minoritas yang baik.



Gambar 13. Grafik Hasil Uji Data Asli

Gambar 13 menampilkan grafik hasil uji coba tanpa penyeimbangan data. Grafik ini memperjelas perbedaan nilai precision dan recall antar model. Terlihat bahwa meskipun akurasi model SVM dan Naïve Bayes Bernoulli tinggi, nilai precision dan recall tetap rendah, menandakan keterbatasan dalam klasifikasi kasus stroke secara akurat.

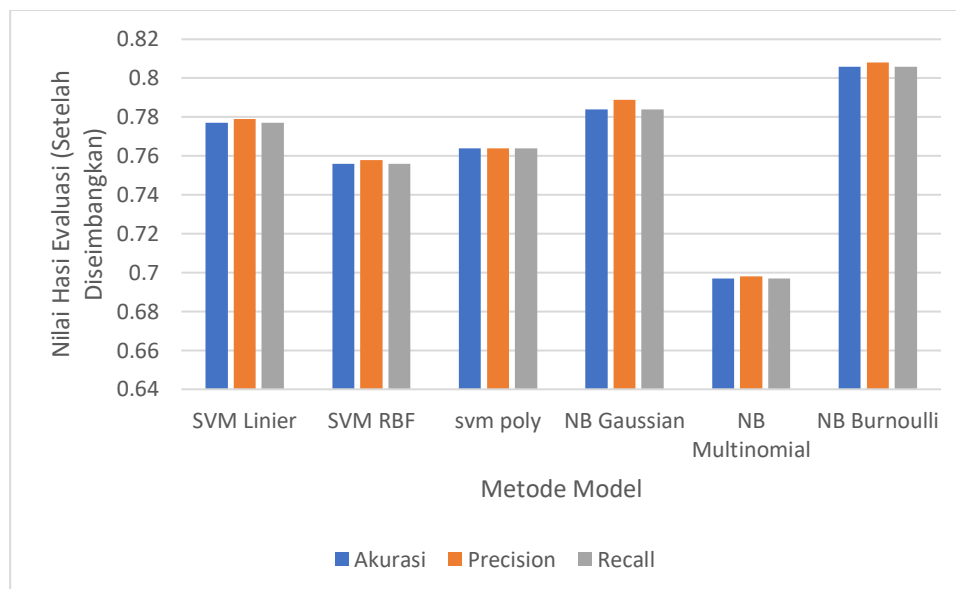
3.1.2 Hasil Uji Coba Setelah Menyeimbangkan Data

Setelah dilakukan penyeimbangan data dengan menggunakan metode SMOTE, diperoleh peningkatan yang signifikan pada hasil evaluasi model. Tabel 3 berikut menyajikan hasil akurasi, precision, dan recall setelah proses balancing data.

Tabel 3. Hasil Uji dengan Menyeimbangkan Data

Model	Akurasi	Precision	Recall
SVM Linier	0,777	0,779	0,777
SVM RBF	0,756	0,758	0,756
SVM Polynomial	0,764	0,764	0,764
NB Gaussian	0,784	0,789	0,784
NB Multinomial	0,697	0,698	0,697
NB Bernoulli	0,806	0,808	0,806

Model SVM kernel linier memberikan hasil akurasi sebesar 0,777, precision 0,779, dan recall 0,777, disusul oleh SVM Polynomial dan RBF yang memiliki performa sedikit lebih rendah. Sementara itu, Naïve Bayes Multinomial menunjukkan performa paling rendah dibandingkan model lain, dengan akurasi 0,697.



Gambar 14. Grafik Hasil Uji dengan Menyeimbangkan Data

Gambar 14 menggambarkan grafik hasil uji coba setelah data seimbang. Grafik ini memperlihatkan bahwa semua model mengalami peningkatan nilai precision dan recall yang signifikan. Naïve Bayes Bernoulli menonjol sebagai model dengan performa paling stabil dan tinggi di seluruh metrik evaluasi.

3.2 Pembahasan

Hasil pengujian menunjukkan perbedaan yang mencolok antara performa model klasifikasi pada data asli yang tidak seimbang dan data yang telah diseimbangkan menggunakan metode SMOTE. Pada data tidak seimbang, seluruh model SVM (kernel linier, RBF, dan polynomial) serta Naïve Bayes Bernoulli mencatatkan akurasi sebesar 0,945, precision 0,472, dan recall 0,500, yang menandakan kelemahan dalam mendeteksi kelas minoritas. Naïve Bayes Gaussian menunjukkan kemampuan yang lebih baik dalam mengenali kelas stroke dengan recall sebesar 0,699 dan precision 0,584, meskipun akurasinya hanya mencapai 0,860. Model Naïve Bayes Multinomial mencatatkan performa terendah pada data asli dengan akurasi 0,813, precision 0,550, dan recall 0,648. Temuan ini menegaskan bahwa nilai akurasi yang tinggi pada data tidak seimbang tidak selalu mencerminkan performa deteksi kelas minoritas yang optimal.

Setelah proses penyeimbangan data menggunakan SMOTE, seluruh model mengalami peningkatan performa yang signifikan. Naïve Bayes Bernoulli mencatatkan hasil terbaik dengan akurasi 0,806, precision 0,808, dan recall 0,806, menandakan kemampuannya mengenali kedua kelas secara seimbang. Naïve Bayes Gaussian menyusul dengan akurasi 0,784, precision 0,789, dan recall 0,784. Model SVM kernel linier juga menunjukkan peningkatan, dengan akurasi 0,777, precision 0,779, dan recall 0,777, sementara kernel polynomial dan RBF memiliki performa yang sedikit lebih rendah. Naïve Bayes Multinomial tetap menjadi model dengan performa terendah pada skenario data seimbang dengan akurasi 0,697. Hasil ini selaras dengan berbagai studi sebelumnya yang menyatakan bahwa metode oversampling seperti SMOTE efektif meningkatkan kinerja klasifikasi, terutama dalam mengenali kelas dengan distribusi minoritas, serta menegaskan pentingnya pemilihan model yang sesuai dengan karakteristik data.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa penerapan algoritma Naïve Bayes dan Support Vector Machine (SVM) efektif digunakan dalam klasifikasi data penyakit stroke. Namun, performa kedua metode tersebut sangat dipengaruhi oleh ketidakseimbangan distribusi kelas dalam dataset. Pada data yang belum diseimbangkan, seluruh model SVM (kernel linier, RBF, dan polynomial) serta Naïve Bayes Bernoulli menghasilkan akurasi tertinggi sebesar 94,5%, tetapi nilai precision (47,2%) dan recall (50%) relatif rendah, sehingga kurang optimal dalam mendeteksi kasus stroke. Setelah data diseimbangkan menggunakan metode SMOTE, performa model meningkat secara signifikan. Peningkatan paling menonjol terjadi pada model Naïve Bayes Bernoulli dengan akurasi 80,6%, precision 80,8%, dan recall 80,6%. Dengan demikian, dapat disimpulkan bahwa proses penyeimbangan data merupakan langkah krusial dalam meningkatkan nilai precision dan recall, serta mendukung terciptanya sistem klasifikasi penyakit stroke yang lebih presisi dan andal.

REFERENCES

- [1] Ary Prandika Siregar, Dwi Priyadi Purba, Jojo Putri Pasaribu, and Khairul Reza Bakara, "Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke," *J. Penelit. Rumpun Ilmu Tek.*, vol. 2, no. 4, pp. 155–164, 2023, doi: 10.55606/juprit.v2i4.3039.
- [2] Y. Azhar, A. K. Firdausy, and P. J. Amelia, "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," *SINTECH (Science Inf. Technol. J.)*, vol. 5, no. 2, pp. 191–197, 2022, doi: 10.31598/sintechjournal.v5i2.1222.
- [3] N. D. Saputri, K. Khalid, and D. Rolliawati, "Komparasi penerapan metode Bagging dan Adaboost pada Algoritma C4. 5 untuk prediksi Penyakit Stroke," *Sist. J. Sist. Inf.*, vol. 11, no. 3, pp. 567–577, 2022, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [4] F. A. H. Airi, T. Suprpti, and A. Bahtiar, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," *E-Link J. Tek. Elektro dan Inform.*, vol. 18, no. 1, p. 73, 2023, doi: 10.30587/e-link.v18i1.5271.
- [5] S. Rahayu and Y. Yamasari, "Klasifikasi Penyakit Stroke dengan Metode Support Vector Machine (SVM)," *J. Informatics Comput. Sci.*, vol. 5, no. 03, pp. 440–446, 2024, doi: 10.26740/jinacs.v5n03.p440-446.
- [6] K. Akmal, A. Faqih, and F. Dikananda, "Perbandingan Metode Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 470–477, 2023, doi: 10.36040/jati.v7i1.6367.
- [7] A. P. Wijaya, "Perbandingan Algoritma Klasifikasi Random Foresst dengan Naïve Bayes Classifier pada Studi Penyakit Berdasarkan Pola Nutrisi," vol. 9, no. 1, pp. 429–438, 2025.
- [8] E. Martantoh, N. Yanieh, T. Informatika, U. Panca, and S. Bekasi, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa ' adah Menggunakan PHP MySQL," vol. 3, no. 2, pp. 166–175, 2022.
- [9] J. Dermawan, M. Fikry, S. Agustian, and L. Oktavia, "Klasifikasi Sentimen Terhadap Topik Pindah Ibu Kota Negara Pada Twitter Menggunakan Metode Naïve Bayes Classifier," vol. 5, pp. 600–610, 2024, doi: 10.30865/json.v5i3.7475.
- [10] J. K. Wororomi, *Data Mining*.
- [11] G. Ayu Ria Widiani and I. M. Mahardika Yasa, "Korelasi Tingkat Pengetahuan Terhadap Kemampuan Deteksi Dini Gejala Stroke Dengan Sikap Keluarga Terhadap Penanganan Pre Hospital," *Bina Gener. J. Kesehat.*, vol. 14, no. 2, pp. 25–30, 2023, doi: 10.35907/bgjk.v14i2.255.
- [12] S. Bumbungan *et al.*, "Penerapan Particle Swarm Optimization (PSO) dalam Pemilihan Parameter Secara Otomatis pada Support Vector Machine (SVM) untuk Prediksi Kelulusan Mahasiswa Politeknik Amamapare Timika," vol. 04, no. 1, pp. 81–93, 2023.
- [13] Y. Ayuningtyas and I. Made Suartana, "Klasifikasi Penyakit Stroke Menggunakan Support Vector Machine (SVM) dan Particle Swarm Optimization (PSO)," *JINACS (Journal Informatics Comput. Sci.)*, vol. 04, no. 04, pp. 452–457, 2023.
- [14] B. Delvika, S. Nurhidayarnis, and P. D. Rinada, "Comparison of Classification Between Naive Bayes and K-Nearest Neighbor on Diabetes Risk in Pregnant Women Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil," vol. 2, no. October, pp. 68–75, 2022.
- [15] F. Zahri, F. Insani, and L. Oktavia, "Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN) Untuk Klasifikasi Penyakit Gagal Jantung," *JOSH J. Inf. Syst. Res.*, vol. 6, no. 2, pp. 997–1005, 2025, doi: 10.47065/josh.v6i2.6480.
- [16] N. S. Ramadan and D. Darwis, "Perbandingan Metode Naïve Bayes Dan Svm Untuk Sentimen Analisis Masyarakat Terhadap Serangan Ransomware Pada Data Kip-K," *J. Sist. Inf. dan Inform.*, vol. 8, no. 1, pp. 12–23, 2024, doi: 10.47080/simika.v8i1.3621.
- [17] Y. N. Paramitha, A. Nuryaman, A. Faisol, E. Setiawan, and D. E. Nurvazly, "Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes," *J. Siger Mat.*, vol. 04, no. 01, pp. 11–16, 2023, [Online]. Available: <https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled->
- [18] S. Mutmainah, "Penanganan Imbalance Data Pada Klasifikasi Kemungkinan Penyakit Stroke," *J. Sains, Nalar, dan Apl. Teknol. Inf.*, vol. 1, no. 1, pp. 10–16, 2021, doi: 10.20885/snati.v1i1.2.
- [19] S. Kenny Riva, S. Casi, and S. Randy Erfa, "Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke," *e-Proceeding Eng.*, vol. 9, no. 3, p. 922, 2022.
- [20] V. Kecman, "Support Vector Machines – An Introduction Support Vector Machines – An Introduction," no. May 2005, 2014, doi: 10.1007/10984697.