APPLICATION OF ADASYN TECHNIQUE IN CLASSIFICATION OF STROKE DISEASE USING BACKPROPAGATION NEURAL NETWORK

PENERAPAN TEKNIK ADASYN PADA KLASIFIKASI PENYAKIT STROKE MENGGUNAKAN BACKPROPAGATION NEURAL NETWORK

Said Rizki Zikrillah Aulia[1], Okfalisa Okfalisa[2*], Elin Haerani[3], Lola Oktavia[4]
[1,2,3,4]Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim RIAU
rizkyzikry7@gmail.com[1], okfalisa@gmail.com[2], elin.haerani@uin-suska.ac.id[3], lolaoktavia_89@yahoo.com[4]

Abstract *- The high prevalence of stroke in Indonesia and the challenge of imbalanced medical record data are major obstacles to the development of an accurate early detection system. This research aims to build a reliable stroke classification model by applying the ADASYN (Adaptive Synthetic Sampling) oversampling technique to address class imbalance before the data is processed using the Backpropagation Neural Network (BPNN) algorithm. The ADASYN technique is applied with the goal of reducing the bias that arises from the imbalanced data distribution between the majority and minority classes. Testing was conducted through various data splitting scenarios (70:30, 80:20, 90:10) and hyperparameter variations to find the optimal configuration. The best results were obtained with the 90:10 data split scheme, using an architecture of 29 neurons and a learning rate of 0.01, which successfully achieved peak performance with an accuracy of 90.46% and an F1-Score of 91.03%. This study demonstrates that the combination of ADASYN and BPNN is a highly effective approach for producing a stroke prediction model that is not only accurate but also sensitive to the minority class, thus having great potential as an early detection support tool in the healthcare sector.*

Keywords *- Stroke Disease, Classification, Backpropagation Artificial Neural Network, ADASYN*

Abstrak - Tingginya prevalensi stroke di Indonesia dan tantangan data rekam medis yang tidak seimbang menjadi penghambat utama dalam pengembangan sistem deteksi dini yang akurat. Penelitian ini bertujuan untuk membangun model klasifikasi stroke yang andal dengan menerapkan teknik oversampling ADASYN (Adaptive Synthetic Sampling) untuk mengatasi ketidakseimbangan kelas sebelum data dilatih menggunakan algoritma Backpropagation Neural Network (BPNN). Pengujian dilakukan melalui berbagai skenario pembagian data (70:30, 80:20, 90:10) dan variasi hiperparameter untuk menemukan konfigurasi optimal. Hasil terbaik diperoleh pada skema pembagian data 90:10 dengan arsitektur 29 neuron dan learning rate 0,01, yang berhasil mencapai performa puncak dengan akurasi 90,46% dan F1-Score 91,03%. Studi ini menunjukkan bahwa kombinasi ADASYN dan BPNN merupakan pendekatan yang sangat efektif untuk menghasilkan model prediksi stroke yang tidak hanya akurat tetapi juga sensitif terhadap kelas minoritas, sehingga berpotensi besar sebagai alat bantu deteksi dini di bidang kesehatan.

Kata Kunci - Penyakit Stroke, Klasifikasi, Backpropagation Neural Network, ADASYN.

I. PENDAHULUAN

Penyakit stroke merupakan masalah kesehatan global yang mendesak, bertanggung jawab atas tingginya angka kematian serta kecacatan fisik dan mental yang secara drastis menurunkan kualitas hidup penderitanya [1]. Beban ini terasa lebih berat di benua Asia, yang menanggung 70% dari total kasus stroke global dengan tingkat kematian mencapai 80% [2]. Secara spesifik di Indonesia, prevalensi stroke menunjukkan tren peningkatan yang mengkhawatirkan, naik dari 7% pada tahun 2013 menjadi 10,9% pada tahun 2018, menurut data dari American Heart Association (AHA).

Berdasarkan hasil Riset Kesehatan Dasar (Riskesdas) yang dilaksanakan di Indonesia tahun 2007 hingga 2018, terdapat tren peningkatan prevalensi penyakit tidak menular, termasuk pada stroke (Badan Penelitian dan Pengembangan Kesehatan, 2021). Terjadi peningkatan prevelensi stroke yang mulanya 7% pada tahun 2013 menjadi 10,9% di tahun 2018. Oleh karena itu, pencegahan dan penanganan stroke sangat penting untuk menurunkan angka kematian dan kecacatan akibat penyakit ini. Tetapi dikarenakan biaya pemeriksaan stroke secara medis yang cukup mahal, sehingga banyak masyarakat yang mengabaikan pemeriksaan kesehatan (Sandy et al., 2022.).

Untuk menjembatani kesenjangan ini, teknologi komputasi seperti Machine Learning menawarkan solusi yang sangat menjanjikan. Sebagai disiplin ilmu yang mampu belajar dari data untuk membuat prediksi, machine learning dapat meningkatkan efisiensi dan akurasi analisis data kesehatan, sehingga menyediakan dukungan keputusan yang lebih solid bagi praktisi medis. Di antara berbagai algoritmanya, Backpropagation Neural Network (BPNN) telah menunjukkan kinerja yang unggul dalam berbagai tugas klasifikasi. Keberhasilannya terbukti tidak hanya pada klasifikasi stroke dengan akurasi mencapai 96,14% [4]. Penelitian yang dilakukan oleh [5]mengenai *Prediksi Sirkulasi Uang Elektronik: Backpropagation dengan Adopsi Algoritma Genetika* menunjukkan bahwa integrasi Backpropagation dan Algoritma Genetika mampu meramalkan peredaran uang elektronik di Indonesia dengan baik. Berdasarkan data deret waktu Bank Indonesia periode Januari 2009–Desember 2019, model yang diusulkan berhasil memprediksi sirkulasi uang elektronik tahun 2020 dengan nilai mean square error (MSE) terendah sebesar 0,000035 pada skenario rasio data pelatihan 90%:10%, learning rate 0,8, serta probabilitas crossover dan mutasi 0,4:0,6. Secara keseluruhan, penelitian ini membuktikan bahwa Backpropagation yang dipadukan dengan Algoritma Genetika dapat memberikan performa yang lebih baik dibandingkan dengan BPNN murni dalam memprediksi sirkulasi uang elektronik.

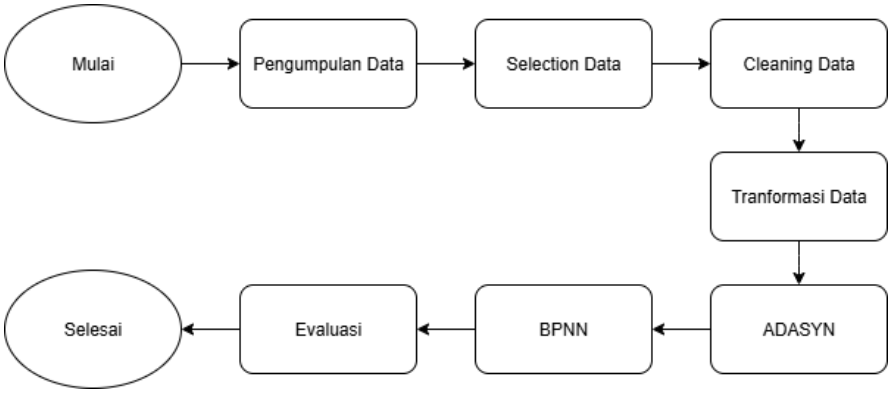
Berdasarkan tingginya prevalensi stroke di Indonesia, potensi prediktif dari algoritma BPNN, serta tantangan krusial mengenai data yang tidak seimbang, maka penelitian ini mengusulkan sebuah pendekatan yang lebih komprehensif. Penelitian ini bertujuan untuk merancang dan mengembangkan sistem klasifikasi stroke dengan mengimplementasikan teknik ADASYN untuk menangani ketidakseimbangan data sebelum melakukan pelatihan model menggunakan algoritma Backpropagation Neural Network. Dengan demikian, penelitian ini diharapkan dapat menghasilkan model klasifikasi yang tidak hanya akurat, tetapi juga lebih sensitif dan andal dalam mengidentifikasi individu berisiko tinggi, sebagai alat bantu deteksi dini yang efektif.

II. SIGNIFIKANSI STUDI

Bab ini menguraikan kerangka kerja dan tahapan metodologis yang diterapkan secara sistematis dalam penelitian ini, mulai dari persiapan data, implementasi model, hingga evaluasi akhir.

2.1. Tahapan Penelitian

Pada tahapan peneltian, akan diuraikan secara rinci tahapan-tahapan sistematis yang dilakukan dalam penelitian ini untuk membangun model klasifikasi stroke. Alur penelitian, yang digambarkan pada diagram di atas, dimulai dari tahap persiapan data yang mencakup pengumpulan, seleksi, pembersihan, dan transformasi. Selanjutnya, dilakukan penanganan data tidak seimbang menggunakan teknik ADASYN, diikuti oleh proses pemodelan dengan algoritma *Backpropagation Neural Network* (BPNN). Tahap akhir dari rangkaian proses ini adalah evaluasi performa model untuk mengukur keakuratan dan efektivitasnya.



Gambar 1. Tahapan Penelitian

2.2. Data Penelitian

Data untuk analisis ini diambil dari "Stroke Prediction Dataset" yang diakses melalui platform Kaggle. Kumpulan data tersebut mencakup 5.110 data dengan 12 atribut, termasuk data demografis (usia, jenis kelamin, status pernikahan), faktor risiko klinis (hipertensi, penyakit jantung, kadar glukosa, BMI), dan variabel gaya hidup (jenis pekerjaan, tipe tempat tinggal, status merokok) untuk memprediksi kejadian stroke. Struktur lengkap dataset awal dirangkum pada Tabel 2.1.

Tabel 1 Data Penelitian

id	Gender	Age	Hypertension	Heart_disease	Ever_married	Work_type	Residence_type	Avg_glucose_level	BMI	Smoking_status	Stroke
9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
51676	Female	61	0	0	Yes	Self-employed	Rural	202.21	N/A	never smoked	1
31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
…	…	…	…	…	…	…	…	…	…	…	…
19723	Female	35	0	0	Yes	Self-employed	Rural	82.99	30.6	never smoked	0
37544	Male	51	0	0	Yes	Private	Rural	166.29	25.6	formerly smoked	0
44679	Female	44	0	0	Yes	Govt_job	Urban	85.28	26.2	Unknown	0

2.3. Selection Data

Pada ada tahap seleksi data, dilakukan proses krusial untuk memilih atribut yang paling signifikan dari dataset mentah guna menentukan fitur yang akan digunakan untuk pemodelan. Tujuan utamanya adalah untuk mereduksi dimensionalitas dan mengurangi *noise* dengan membuang fitur yang tidak relevan, di mana pada penelitian ini kolom id dihilangkan karena hanya berfungsi sebagai pengenal unik yang tidak memiliki nilai prediktif. Hasil dari proses ini adalah dataset yang lebih ringkas dengan 11 variabel, yang terbagi menjadi satu variabel dependen (stroke) dan sepuluh variabel independen. Pemilihan fitur yang efektif ini tidak hanya meningkatkan akurasi dan kinerja model dengan memfokuskan analisis pada data yang paling berpengaruh, tetapi juga mengurangi kompleksitas komputasi dan risiko *overfitting*, sehingga berkontribusi pada hasil prediksi yang lebih andal dan dapat diinterpretasikan.

2.4. Cleaning Data

Cleaning data adalah langkah awal yang dilakukan pada set data yang tersedia, yang Pembersihan data bertujuan untuk mengeliminasi noise dan inkonsistensi dalam dataset [6]. Proses ini melibatkan pemilihan data yang relevan, mengoreksi ketidakakuratan, dan menghapus entri data yang tidak lengkap, tidak konsisten, atau tidak relevan. Dalam konteks penelitian ini, pembersihan data akan diterapkan pada dataset stroke untuk mengatasi masalah seperti nilai yang hilang, entri yang tidak konsisten, dan adanya data yang diberi label “NaN” atau “Tidak Diketahui”. Penanganan anomali ini sangat penting untuk menghindari bias atau kesalahan selama tahap pemodelan dan untuk memastikan keandalan dan validitas keseluruhan hasil yang dihasilkan oleh jaringan saraf.

2.5. Transformasi Data

Tahap transformasi data dilakukan untuk mengonversi fitur-fitur kategorikal ke dalam format numerik agar kompatibel dengan algoritma *machine learning* [7]. Proses yang dikenal sebagai *encoding* ini diterapkan pada beberapa fitur. Secara spesifik, teknik Label Encoding digunakan pada fitur Ever_married dan Residence_type untuk mengubahnya menjadi nilai numerik. Sementara itu, fitur Gender, Work_type, dan Smoking_status diubah menggunakan One-Hot Encoding, yang membuat kolom-kolom biner baru untuk setiap kategori (contohnya, gender_male dan work_type_private) untuk menghindari bias urutan. Hasil lengkap dari proses konversi ini, yang memastikan semua data dalam format numerik.

2.6. Normalisasi Data

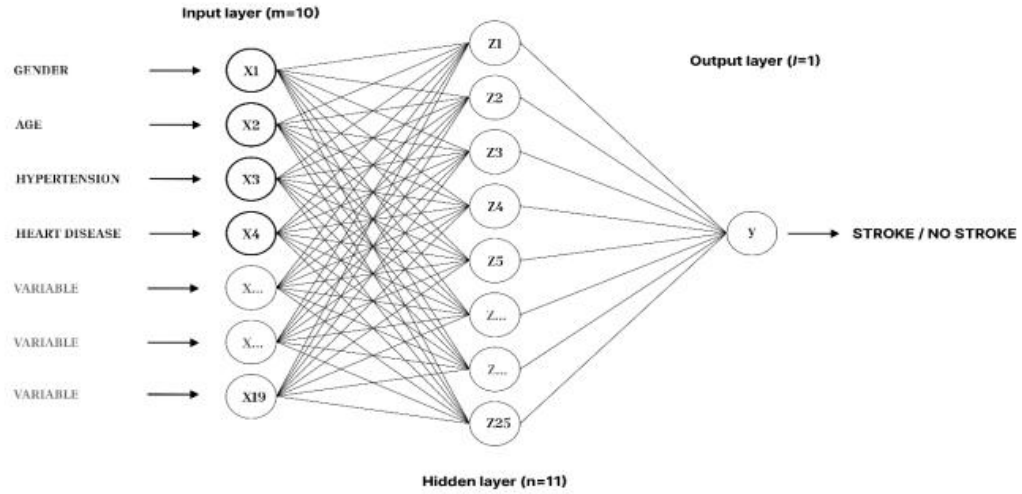
Untuk memastikan setiap fitur memberikan kontribusi yang seimbang pada model, dilakukan normalisasi data menggunakan teknik Min-Max Scaling. Metode ini bekerja dengan mengubah setiap nilai numerik dalam dataset ke dalam rentang seragam antara 0 dan 1[8]. Normalisasi sangat penting ketika dataset memiliki fitur dengan satuan dan skala yang berbeda, seperti usia dan kadar glukosa, yang tanpanya dapat menyebabkan fitur dengan rentang nilai lebih besar mendominasi proses pembelajaran mesin [9]. Dengan demikian, penerapan Min-Max Scaling dapat mencegah bias model dan berpotensi meningkatkan kecepatan konvergensi selama pelatihan.

2.7. ADASYN

Teknik ADASYN diterapkan dengan tujuan untuk mengurangi bias yang timbul dari ketidakseimbangan distribusi data antara kelas mayoritas dan minoritas (Hidayat et al., 2021). Sebelum proses ini, dataset menunjukkan ketidakseimbangan yang signifikan, dengan 4861 data untuk kelas '0' (mayoritas) dan hanya 249 data untuk kelas '1' (minoritas). Setelah penerapan ADASYN, distribusi kelas berhasil diseimbangkan melalui *oversampling*, di mana jumlah data untuk kelas '1' ditingkatkan menjadi 4882, sementara kelas '0' tetap sebanyak 4861 data.

2.8. Backpropagation Neural Network (BPNN)

Backpropagation Neural Network (BPNN) adalah algoritma pembelajaran terawasi yang terdiri dari beberapa lapisan, termasuk lapisan input, satu atau lebih lapisan tersembunyi, dan lapisan output, yang semuanya saling berhubungan dengan bobot yang dapat disesuaikan [11]. Jaringan ini beroperasi dalam dua fase utama: fase umpan maju dan fase umpan balik (backpropagation). Pada fase feedforward, data input diteruskan dari lapisan input melalui lapisan tersembunyi ke lapisan output, di mana output awal dihasilkan. Jika output yang dihasilkan tidak berada dalam rentang kesalahan yang dapat diterima dibandingkan dengan target output, kesalahan tersebut disebarkan ke belakang dari lapisan output ke lapisan input melalui lapisan tersembunyi. Selama fase mundur ini, bobot disesuaikan untuk meminimalkan kesalahan menggunakan gradient descent atau metode optimasi serupa. Proses berulang ini terus berlanjut hingga jaringan mencapai tingkat akurasi yang memuaskan. Arsitektur umum BPNN diilustrasikan pada Gambar [12].



Gambar 2. Arsitektur BPNN

Arsitektur jaringan syaraf tiruan Backpropagation (BPNN) terdiri dari tiga lapisan utama, dimulai dengan lapisan input yang berjumlah 10 neuron. Setiap neuron mewakili fitur atau atribut dari dataset yang digunakan, seperti gender (X1), umur (X2), hipertensi (X3), dan riwayat penyakit jantung (X4). Selain itu, terdapat beberapa variabel tambahan (X5 hingga X10) yang mencakup informasi penting lainnya seperti BMI, status merokok, kadar glukosa, dan fitur relevan lainnya yang berkaitan dengan stroke. Selanjutnya adalah lapisan tersembunyi (hidden layer) yang terdiri dari 11 neuron (Z1 hingga Z11). Tiap neuron pada lapisan ini menerima sinyal dari semua neuron di lapisan input melalui bobot tertentu. Masukan tersebut akan dijumlahkan dan dilewatkan ke dalam fungsi aktivasi seperti sigmoid atau ReLU. Proses ini menghasilkan keluaran non-linear yang memungkinkan jaringan untuk mengenali pola kompleks dalam data dan meningkatkan kemampuan prediksi jaringan terhadap kondisi stroke. Terakhir, lapisan keluaran terdiri dari satu neuron (y) yang memberikan hasil prediksi akhir berupa nilai biner, yaitu 1 untuk stroke dan 0 untuk tidak stroke. Nilai dari neuron ini dihasilkan berdasarkan hasil kalkulasi dari neuron-neuron di lapisan tersembunyi yang sudah disesuaikan bobotnya. Output ini juga melewati fungsi aktivasi (biasanya sigmoid) agar dapat diubah menjadi nilai probabilitas, sehingga dapat digunakan untuk pengambilan keputusan dalam klasifikasi stroke.

2.9. Evaluasi

Evaluasi performa model dilakukan menggunakan metrik yang diturunkan dari Confusion Matrix. Matriks ini mengukur kualitas klasifikasi dengan membandingkan hasil prediksi terhadap nilai aktual, yang dikategorikan ke dalam *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN)[13]. Berdasarkan nilai-nilai tersebut, dihitung beberapa metrik kunci: Akurasi, Presisi, Recall, dan F1-Score. Analisis komprehensif dari metrik ini memberikan gambaran menyeluruh tentang efektivitas model dalam mengklasifikasikan data secara akurat sekaligus mengidentifikasi potensi kelemahannya.

Tabel 2 Confusion Matrix

Kelas Benar	Stroke	Tidak Stroke
Stroke	True Positif	False Positif
Tidak Stroke	False Negatif	True Negatif

Akurasi adalah persentase dari prediksi yang benar terhadap keseluruhan data:

$$\boldsymbol{Akurasi = \frac{TP + TN}{TP + TN + FP + FN}} \quad \textbf{(1)}$$

Presisi mengukur proporsi prediksi positif yang benar-benar benar:

$$\boldsymbol{Presisi = \frac{TP}{TP + FP}} \quad \textbf{(2)}$$

Recall mengukur kemampuan model untuk menemukan semua contoh positif dalam data:

$$\boldsymbol{Recall = \frac{TP}{TP + FN}} \quad \textbf{(3)}$$

F1-Score adalah rata-rata harmonis antara presisi dan recall:

$$\boldsymbol{F1 = 2 . \frac{Presisi\ .\ Recall}{Presisi + Recall}} \quad \textbf{(4)}$$

III. HASIL DAN PEMBAHASAN

Bagian in menyajikan hasil dari serangkaian pengujian yang dilakukan untuk mengevaluasi performa model *Backpropagation Neural Network* (BPNN) dalam klasifikasi penyakit stroke. Pengujian dilakukan melalui tiga skenario utama yang dibedakan berdasarkan rasio pembagian data latih dan data uji, yaitu 70:30, 80:20, dan 90:10. Pada setiap skenario, dilakukan eksperimen dengan memvariasikan parameter arsitektur jaringan (jumlah neuron) dan *learning rate* untuk mengidentifikasi konfigurasi yang paling optimal. Kinerja setiap model diukur dan dibandingkan menggunakan metrik evaluasi standar: Akurasi, Presisi, Recall, dan F1-Score.

3.1. Hasil Pengujian

Berdasarkan Berdasarkan hasil pengujian dengan skema pembagian data 70:30, performa model dievaluasi pada tiga arsitektur berbeda (25, 29, dan 35 neuron) dengan tiga variasi *learning rate* (0,001, 0,01, dan 0,1). Hasil terbaik dicapai saat menggunakan arsitektur 35 neuron dengan learning rate 0,01. Konfigurasi ini menghasilkan kinerja paling seimbang dan unggul, dengan akurasi sebesar 89,88%, presisi 87,32%, recall 93,39%, dan F1-Score mencapai 90,25%, yang merupakan nilai F1-Score tertinggi di antara semua skenario yang diuji.

Tabel 3 Pembagian Data 70 : 30

Arsitektur	Learning Rate	Akurasi	Presisi	Recall	F1-Score
25	0,001	84,75	81,81	90,73	85,65
	0,01	86,43	84,25	89,71	86,89
	0,1	85,71	81,81	91,96	86,59
29	0,001	87,25	83,27	93,32	88,01
	0,01	89,5	87,56	92,16	89,8
	0,1	88,07	81,83	97,96	89,17
35	0,001	86,22	83,42	90,52	86,83
	0,01	89,88	87,32	93,39	90,25
	0,1	85,5	78,54	97,82	87,13

Pada pengujian dengan skema pembagian data 80% latih dan 20% uji, hasil terbaik diperoleh saat model menggunakan arsitektur 35 neuron dengan learning rate 0,01. Konfigurasi ini menunjukkan

performa paling unggul dibandingkan kombinasi lainnya, dengan capaian akurasi 90,21%, presisi 87,87%, recall 93,35%, dan F1-Score 90,53%. Nilai F1-Score tertinggi pada skenario ini mengindikasikan bahwa model mencapai keseimbangan performa yang paling optimal antara presisi dan recall.

Tabel 4 Pembagian Data 80 : 20

Arsitektur	Learning Rate	Akurasi	Presisi	Recall	F1-Score
25	0,001	85,18	81,46	91,21	86,06
	0,01	86,92	85,9	88,45	87,15
	0,1	85,85	80,15	95,4	87,11
29	0,001	86,31	82,53	92,23	87,11
	0,01	88,87	87,56	90,7	89,1
	0,1	87,44	83,47	93,46	88,18
35	0,001	87,08	83,61	92,33	87,76
	0,01	90,21	87,87	93,35	90,53
	0,1	87,03	81,28	96,32	88,16

Dalam skenario pengujian dengan pembagian data 90% latih dan 10% uji, performa puncak model tercapai saat menggunakan arsitektur 29 neuron dan learning rate 0,01. Konfigurasi ini menghasilkan nilai-nilai metrik tertinggi secara keseluruhan, dengan akurasi mencapai 90,46%, presisi 86,13%, recall 96,52%, dan F1-Score 91,03%. Hasil ini menunjukkan bahwa dengan proporsi data latih yang lebih besar, arsitektur 29 neuron memberikan keseimbangan performa terbaik antara presisi dan recall dibandingkan arsitektur lainnya.

Tabel 5 Pembagian Data 90 : 10

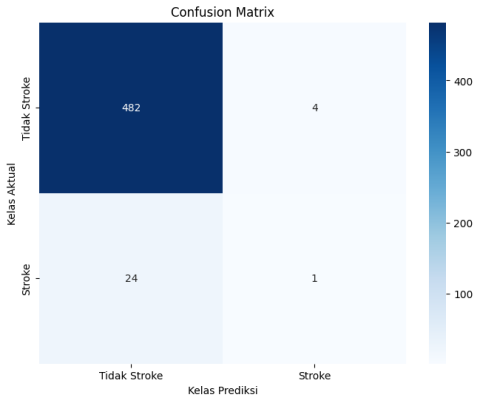
Arsitektur	Learning Rate	Akurasi	Presisi	Recall	F1-Score
25	0,001	88,1	85,12	92,43	88,63
	0,01	88,62	83,16	96,93	89,52
	0,1	89,23	84,04	96,93	90,03
29	0,001	87,79	84,64	92,43	88,37
	0,01	90,46	86,13	96,52	91,03
	0,1	89,74	84,18	97,96	90,55
35	0,001	86,97	84,94	89,98	87,39
	0,01	90,36	88,5	92,84	90,62
	0,1	88,51	84,97	93,66	89,11

Dari keseluruhan hasil pengujian pada tiga skenario pembagian data, dapat disimpulkan bahwa kinerja model terbaik secara keseluruhan dicapai pada skema pembagian data 90:10. Pada skenario ini, konfigurasi optimal diperoleh dengan menggunakan arsitektur 29 neuron dan learning rate 0,01, yang berhasil mencatatkan F1-Score tertinggi sebesar 91,03% dan akurasi 90,46%. Temuan ini mengindikasikan bahwa dengan memberikan porsi data latih yang lebih besar, model mampu mencapai keseimbangan presisi dan recall yang paling efektif, menjadikannya konfigurasi paling andal untuk klasifikasi stroke dalam penelitian ini.

3.2. Pembahasan

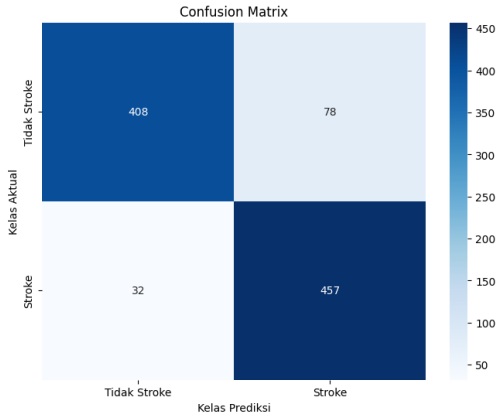
Gambar 3 tersebut menampilkan hasil dari *confusion matrix* untuk model yang diuji tanpa menggunakan teknik ADASYN yaitu pada performa terbaik pada saat menggunakan ADASYN, dengan skema pembagian data 90:10 pada arsitektur 29 neuron dan *learning rate* 0.01. Meskipun model ini mencapai akurasi yang tampak tinggi sebesar 94,52%, metrik lainnya menunjukkan kinerja yang sangat buruk, dengan presisi 20%, recall hanya 4%, dan F1-score 6,67%. Hasil *confusion matrix* menjelaskan mengapa hal ini terjadi: akurasi yang tinggi disebabkan oleh kemampuan model menebak kelas mayoritas "Tidak Stroke" dengan benar (486 *True Negative*). Namun, model ini

hampir sepenuhnya gagal mengenali kelas minoritas, di mana ia hanya berhasil mengidentifikasi 13 kasus "Stroke" (*True Positive*) dan salah mengklasifikasikan 35 kasus "Stroke" lainnya sebagai "Tidak Stroke" (*False Negative*). Performa yang sangat tidak seimbang ini menunjukkan bahwa tanpa ADASYN, model menjadi bias dan tidak efektif untuk deteksi dini.



Gambar 3. Confusion Matrix Tanpa Adasyn

Pada Gambar 3 menampilkan *confusion matrix* dari model dengan performa terbaik, yang diperoleh dari skenario pembagian data 90:10 dengan arsitektur 29 neuron dan *learning rate* 0,01. Matriks tersebut merinci kinerja klasifikasi model, di mana model berhasil memprediksi dengan benar sebanyak 457 kasus "Stroke" (*True Positive*) dan 408 kasus "Tidak Stroke" (*True Negative*). Namun, model juga melakukan kesalahan prediksi dengan mengklasifikasikan 32 kasus "Stroke" sebagai "Tidak Stroke" (*False Negative*) dan 78 kasus "Tidak Stroke" sebagai "Stroke" (*False Positive*). Tingginya jumlah prediksi yang benar (TP dan TN) dibandingkan dengan jumlah kesalahan (FP dan FN) inilah yang menghasilkan nilai akurasi dan F1-Score yang tinggi.



Gambar 4. Confusion Matrix Menggunakan Adasyn

Perbandingan hasil antara model dengan dan tanpa ADASYN secara jelas menunjukkan dampak krusial dari penanganan data yang tidak seimbang. Tanpa ADASYN, model menunjukkan performa yang sangat bias, meskipun akurasinya tampak tinggi (94,52%), nilai recall-nya hanya 4%, yang berarti model gagal mengidentifikasi 35 dari 48 kasus stroke aktual. Sebaliknya, setelah penerapan ADASYN, model mengalami transformasi kinerja yang drastis. Ia mampu mengidentifikasi 457 kasus stroke dengan benar, yang melejitkan recall menjadi 93,45% dan menghasilkan F1-Score yang seimbang sebesar 89,26%. Hal ini menegaskan bahwa penggunaan ADASYN merupakan langkah esensial yang berhasil mengubah model dari yang tidak efektif menjadi alat klasifikasi yang andal dan sensitif untuk deteksi dini penyakit stroke.

KESIMPULAN

Berdasarkan analisis dan pengujian yang telah dilakukan, dapat disimpulkan bahwa penerapan teknik *oversampling* ADASYN berhasil secara signifikan meningkatkan kinerja algoritma Backpropagation Neural Network (BPNN) dalam mengatasi masalah data tidak seimbang pada klasifikasi risiko stroke. Model mencapai performa optimal pada skema pembagian data 90:10 dengan arsitektur 29 neuron dan *learning rate* 0,01, yang menghasilkan akurasi sebesar 90,46% dan F1-Score 91,03%. Hasil ini membuktikan bahwa kombinasi ADASYN dan BPNN merupakan pendekatan yang efektif dan potensial untuk dikembangkan sebagai sistem pendukung keputusan dalam deteksi dini risiko stroke. Untuk penelitian selanjutnya, disarankan agar dilakukan eksplorasi dengan algoritma *deep learning* lain seperti Long Short-Term Memory (LSTM) atau membandingkan beberapa teknik *oversampling* berbeda untuk menemukan metode yang paling optimal.

REFERENSI

[1] R. E. Pambudi, Sriyanto, and Firmansyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.45," *Jurnal Teknika*, vol. 16, no. 02, pp. 221–226, 2022.

[2] P. Melani Almahmuda Batubara, I. Afrianty, S. Sanjaya, and F. Syafria, "Klasifikasi Penyakit Stroke Jaringan Syaraf Tiruan Menerapkan Metode Learning Vector Quantization," *Jurnal Informatika Universitas Pamulang*, vol. 8, no. 2, pp. 223–228, 2023.

[3] L. A. P. Sandy, E. Kusumawardhani, P. W. Nugraheni, L. T. N. Meleiva, and A. V. Gunawan, "Sistem Identifikasi Dini Penyakit Stroke Dengan Menggunakan Jaringan Syarag Tiruan Perambatan Balik," *Jurnal Keilmuan dan Aplikasi Bidanf Teknik Informatika*, vol. 16, no. 2, pp. 145–157, 2022.

[4] M. Azhima, I. Afrianty, E. Budianita, and S. Kurnia Gusti, "Penerapan Metode Backpropagation Neural Network untuk Klasifikasi Penyakit Stroke," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 6, pp. 3013–3021, 2024, doi: 10.30865/klik.v4i6.1956.

[5] E. Budianita, O. Okfalisa and M. R. Assiddiki, "The Prediction of E-Money Circulation: Backpropagation with Genetic Algorithm Adoption," *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, Taiz, Yemen, 2021, pp. 1-6, doi: 10.1109/ICOTEN52080.2021.9493468.

[6] I. Romli, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASIFIKASI PENYAKIT ISPA," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 4, no. 1, p. 10, Jun. 2021, doi: 10.21927/ijubi.v4i1.1727.

[7] E. Saputro and D. Rosiyadi, "Bianglala Informatika Penerapan Metode Random Over-Under Sampling Pada Algoritma Klasifikasi Penentuan Penyakit Diabetes," *Bianglala Informatika*, vol. 10, no. 1, pp. 42–47, 2022, [Online]. Available: https://archive.ics.uci.edu/ml/machine-learning-

[8] A. Toha, P. Purwono, and W. Gata, "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 4, no. 1, pp. 12–21, May 2022, doi: 10.12928/biste.v4i1.6079.

[9] A. Harmain, H. Kurniawan, Kusrini, and D. Maulina, "Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wila-yah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas," *TEKNIMEDIA*, vol. 2, no. 2, pp. 83–89, 2021.

[10] W. Hidayat, M. Ardiansyah, and A. Setyanto, "Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb," *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 1, pp. 11–20, Jun. 2021, doi: 10.29408/edumatic.v5i1.3125.

[11] M. R. L. M. Waail, A. Syahputra, R. Hidayat, "Klasifikasi Jenis Kelengkeng Berdasarkan Morfologi Daun Dengan Ekstraksi Ciri RGB, GLCM, dan Bentuk Menggunakan Metode BPNN," *Jurnal Sistem Informasi dan Komputer Akuntansi*, vol. 4, no. 2, pp. 183–193, 2023.

[12] M. Azhima, I. Afrianty, E. Budianita, and S. Kurnia Gusti, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Metode Backpropagation Neural Network untuk Klasifikasi Penyakit Stroke," *Media Online)*, vol. 4, no. 6, pp. 3013–3021, 2024.

[13] Tedyyana, Agus, Osman Ghazali, and Onno Purbo. "Model Design of Intrusion Detection System on Web Server Using Machine Learning Based." *Proceedings of the 11th International Applied Business and Engineering Conference, ABEC 2023, September 21st, 2023, Bengkalis, Riau, Indonesia*. 2024.