

# Penerapan Metode Backpropagation Neural Network untuk Klasifikasi Penyakit Stroke

Mohd. Azhima, Iis Afrianty\*, Elvia Budianita, Siska Kurnia Gusti

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1</sup>12050113334@students.uin-suska.ac.id, <sup>2,\*</sup>iis.afrianty@uin-suska.ac.id, <sup>3</sup>elviabudianita@uin-suska.ac.id,

<sup>4</sup>siskakurniagusti@uin-suska.ac.id

Email Penulis Korespondensi: iis.afrianty@uin-suska.ac.id

**Abstrak**—Stroke merupakan penyakit tidak menular yang dapat terjadi secara tiba-tiba akibat gangguan fungsi otak secara lokal maupun global. Gejala awal penyakit stroke seringkali sulit dikenali sehingga menyebabkan banyak penderitanya tidak menyadari atau merasakan tanda-tandanya, sehingga angka kematiannya cukup tinggi. Penelitian ini bertujuan untuk mengetahui kemampuan metode Backpropagation Neural Network (BPNN) dalam melakukan klasifikasi pada penyakit stroke. Dataset yang digunakan terdiri dari 4891 rekam medis dengan kelas stroke dan non stroke yang melibatkan sepuluh variabel relevan (jenis kelamin, usia, hipertensi, riwayat penyakit jantung, BMI, kadar gula darah, dan sebagainya). Penelitian ini menjalankan tiga skenario dengan model arsitektur BPNN [19:25:1], [19:29:1], dan [19:35:1] dengan menggunakan kombinasi variabel tertentu yaitu perbandingan data pelatihan dan pengujian (90:10, 80 :20, 70:30), dan learning rate 0,1; 0,01; 0,001. Hasil pengujian dengan tingkat rata-rata akurasi tertinggi sebesar 96,14% dicapai dengan model arsitektur [19:29:1], learning rate 0,001, dan distribusi data pelatihan dan pengujian 80:20. Berdasarkan pengujian dapat disimpulkan bahwa BPNN dinilai mampu mengklasifikasikan penyakit stroke.

**Kata Kunci:** Backpropagation Neural Network; Klasifikasi; SMOTE; Stroke

**Abstract**—Stroke is a non-communicable disease that can occur suddenly due to local or global disruption of brain function. The early symptoms of stroke are often difficult to recognize, causing many sufferers not to realize or feel the signs, so the death rate is quite high. This research aims to determine the ability of the Backpropagation Neural Network (BPNN) method in classifying stroke. The dataset used consists of 4891 medical records with stroke and non-stroke classes which include ten relevant variables (gender, age, hypertension, history of heart disease, BMI, blood sugar levels, and so on). This research runs three scenarios with the BPNN architecture model [19:25:1], [19:29:1], and [19:35:1] using a certain combination of variables, namely the comparison of training and testing data (90:10, 80 :20, 70:30), and learning rate 0.1; 0.01; 0.001. Test results with the highest average accuracy level of 96.14% were achieved with an architectural model of [19:29:1], a learning rate of 0.001, and a training and testing data distribution of 80:20. Based on testing, it can be concluded that BPNN is considered capable of classifying stroke.

**Keywords:** Backpropagation Neural Network; Classification; SMOTE; Stroke

## 1. PENDAHULUAN

Penyakit Stroke adalah jenis penyakit tidak menular yang dapat terjadi secara tiba-tiba karena gangguan fungsi otak lokal atau global. Gejalanya berlangsung selama minimal 24 jam dan dapat menyebabkan kematian, tanpa ada penyebab lain selain gangguan vaskuler [1]. Gejala dini penyakit stroke sukar dideteksi, dengan banyak penderita tidak menyadari atau merasakannya. Hal ini mengakibatkan banyak orang mengabaikan gejala-gejala penyakit tersebut. Tingginya angka kematian yang disebabkan oleh penyakit ini seringkali dikaitkan dengan kurangnya pengetahuan masyarakat tentang stroke [2]. Kurangnya pengetahuan masyarakat terhadap penyakit stroke membuat mereka mengabaikan gejala-gejala penyakit ini, jika gejalanya diketahui lebih awal dan segera ditangani sedini atau secepat mungkin maka masih ada kemungkinan para pengidap penyakit stroke dapat diobati dan berangsur pulih [3]. Oleh karena itu, dibutuhkan metode teknologi informatika sebagai langkah untuk memprediksi gejala-gejala penyakit stroke.

Penyakit stroke telah banyak diteliti sebelumnya, seperti pada penelitian yang dilakukan oleh [4] menggunakan metode Fuzzy Backpropagation Neural Network (FBPNN). Penelitian tersebut menghasilkan tingkat akurasi sebesar 86,52%, yang menyimpulkan bahwa pola jaringan tidak mempengaruhi hasil akurasi dengan mengubah jumlah neuron hidden layer. Namun, *learning rate*, *epoch*, dan pembagian data berpengaruh pada tingkat akurasi yang dihasilkan. Pada penelitian [5], dilakukan dua jenis metode klasifikasi yaitu SVM dan ansambel (dikantongi) serta jaringan saraf tiruan (ANN), yang menghasilkan bahwa jaringan saraf tiruan yang dilatih dengan algoritma penurunan gradien stokastik memiliki akurasi tertinggi (95,3%). Penelitian lain yang dilakukan oleh [1] dengan performa algoritma KNN terbaik didapatkan dengan nilai  $k=5$  dan akurasi 93,54%. Selanjutnya, penelitian [6] mengklasifikasikan penyakit stroke menggunakan metode naïve bayes dan mendapatkan hasil akurasi tertinggi sebesar 80% dengan pembagian data 80:20. Terakhir, Backpropagation Neural Network (BPNN) telah digunakan dalam penelitian [7] untuk melakukan prediksi pada penyakit stroke. Model jaringan saraf tersebut mencapai akurasi pelatihan sebesar 98,13% dan akurasi pengujian sebesar 97,97%. Hasil ini menunjukkan bahwa model jaringan saraf tiruan dengan propagasi balik sangat layak digunakan untuk prediksi stroke. Oleh karena itu, berdasarkan penelitian sebelumnya, metode yang akan kita gunakan untuk melakukan klasifikasi penyakit stroke adalah Backpropagation Neural Network (BPNN).

BPNN merupakan salah satu contoh algoritma dalam bidang jaringan syaraf tiruan [8]. Beberapa penelitian telah menunjukkan bahwa algoritma pembelajaran jaringan syaraf tiruan mampu menyelesaikan masalah pemodelan deret waktu nonlinier dengan menggunakan algoritma backpropagation [9], [10]. Ada 3 fase dalam pelatihan BPNN, yaitu fase maju (feed forward), fase mundur (back propagation), dan fase modifikasi bobot [11]. Terdapat beberapa penelitian yang menggunakan BPNN ini untuk tugas klasifikasi. Penelitian [12] menggunakan BPNN untuk mengklasifikasi penyakit

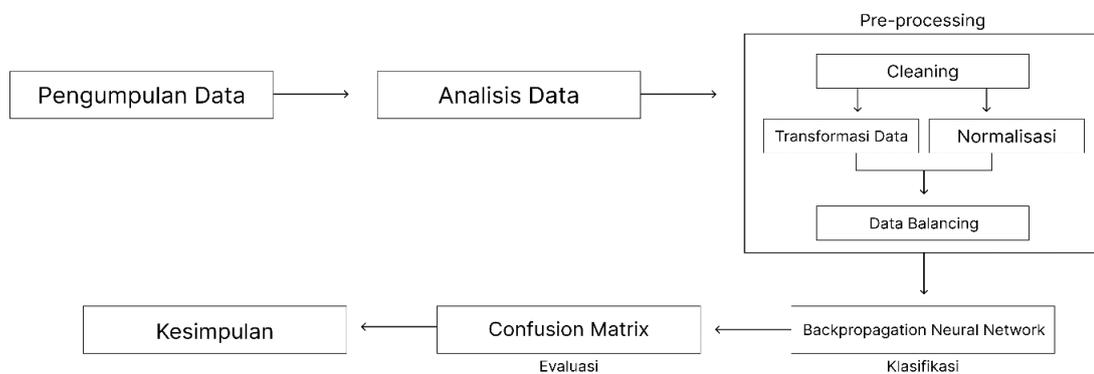
diabetes dan mendapatkan hasil akurasi prediksi sebesar 93% lebih tinggi dibandingkan teknik konvensional. Penelitian [13] klasifikasi jenis kelengkeng dengan menggunakan metode BPNN yang mendapat hasil akurasi sebesar 95%. Selanjutnya, penelitian stiqomatul Fajriyah Yulianti et al, melakukan tugas klasifikasi dengan membandingkan dua metode yaitu knn dan bpnn. Kedua-duanya menghasilkan akurasi yang sama2 tinggi dan menjanjikan yaitu 99,84% [14]. Studi yang dilakukan oleh [10] membahas perbandingan antara metode klasifikasi Backpropagation Neural Network (BPNN) dan Recurrent Neural Network (RNN). Dalam penelitian ini, metode klasifikasi Backpropagation Neural Network terbukti mencapai tingkat akurasi tertinggi dengan nilai sebesar 87,97%.

Meskipun berbagai penelitian telah dilakukan untuk klasifikasi penyakit stroke dengan menggunakan berbagai metode seperti FBPNN, SVM, KNN, Naïve Bayes, dan BPNN, ada beberapa hal yang masih perlu diteliti lebih lanjut. Sebagian besar penelitian sebelumnya belum mengeksplorasi secara mendalam pengaruh dari berbagai parameter BPNN pada performa model secara bersamaan dalam satu studi. Berdasarkan penjelasan sebelumnya maka penelitian ini bertujuan untuk mengetahui kemampuan metode Backpropagation Neural Network (BPNN) dalam melakukan klasifikasi pada dataset yang memiliki dua kelas target yaitu stroke dan tidak stroke.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian dilakukan seperti pada Gambar 1. Pertama melakukan pengumpulan data selanjutnya melakukan preprocessing pada dataset yang mana didalam tahap processing terdapat beberapa tahapan untuk melakukan perubahan dan perbaikan dataset. Setelah dilakukan processing, dilanjutkan dengan proses klasifikasi lalu melakukan evaluasi pada model, sehingga pada akhirnya dapat ditarik sebuah kesimpulan. Berikut tahapan penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

### 2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari studi sebelumnya [4], [15]. data yang digunakan mencakup jenis stroke dan faktor risiko stroke. Data jenis stroke dibagi menjadi dua kategori dari total 4981 data stroke, yaitu kategori non-stroke dengan 4733 data dan kategori stroke dengan 248 data. Data ini menampilkan informasi terkait penyakit stroke yang berisikan 10 variabel gejala stroke (jenis kelamin, usia, hipertensi, riwayat penyakit jantung, BMI, kadar gula darah, status pernikahan, jenis pekerjaan, jenis tempat tinggal, dan status merokok seperti pada Gambar 2 di bawah:

	A	B	C	D	E	F	G	H	I	J	K
1	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
2	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
3	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
4	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
5	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
6	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
7	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
8	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
9	Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
10	Female	81	1	0	Yes	Private	Rural	80.43	29.7	never smoked	1
11	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
12	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
13	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
14	Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
15	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
16	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1

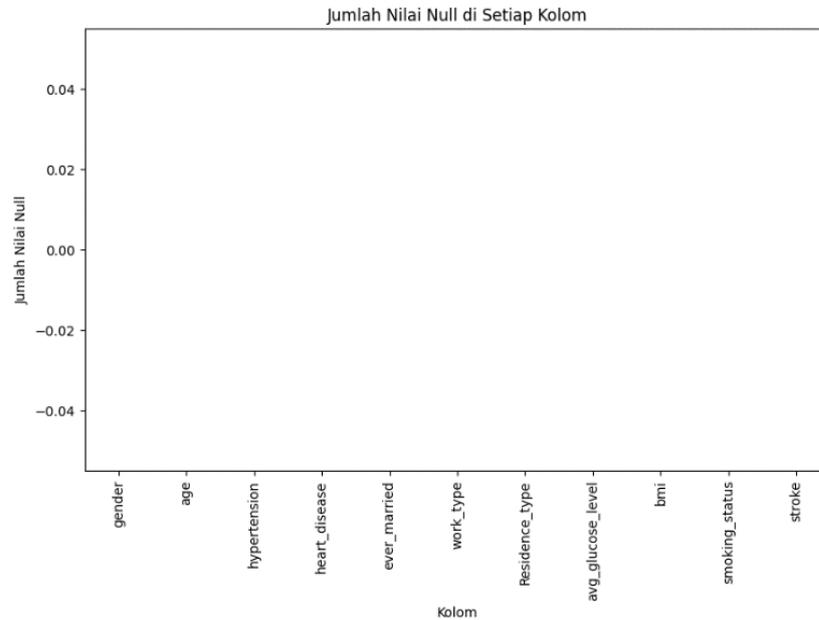
Gambar 2. Original Dataset

### 2.3 Preprocessing

Pre-processing adalah tahap mengubah data mentah menjadi data yang sesuai dengan prosedur [16]. Pada penelitian ini dilakukan tahap *cleaning* dan tranformasi data.

#### a. *Cleaning*

Sebelum dataset dapat digunakan pada penelitian ini, dilakukan pengecekan data untuk memastikan tidak ada data yang null atau NaN. Hasil pengecekan dapat dilihat pada Gambar 3.



Gambar 3. Pengecekan Nilai Null atau NaN

#### b. Transformasi Data

Transformasi Data merupakan tahap dalam skala pengukuran dimana data asli diubah menjadi bentuk yang sesuai untuk diproses pada tahap selanjutnya. Penelitian ini menggunakan metode one-hot encoding untuk melakukan transformasi data, setiap kategori unik dalam variabel kategorikal diubah menjadi kolom terpisah dalam bentuk representasi biner (0 atau 1) [17]. Adapun variabel yang diubah yaitu; jenis kelamin, status perkawinan, jenis pekerjaan, lingkungan tempat tinggal, dan status merokok. Dataset yang telah melalui proses One Hot Encoding dapat dilihat pada Gambar 4 berikut:

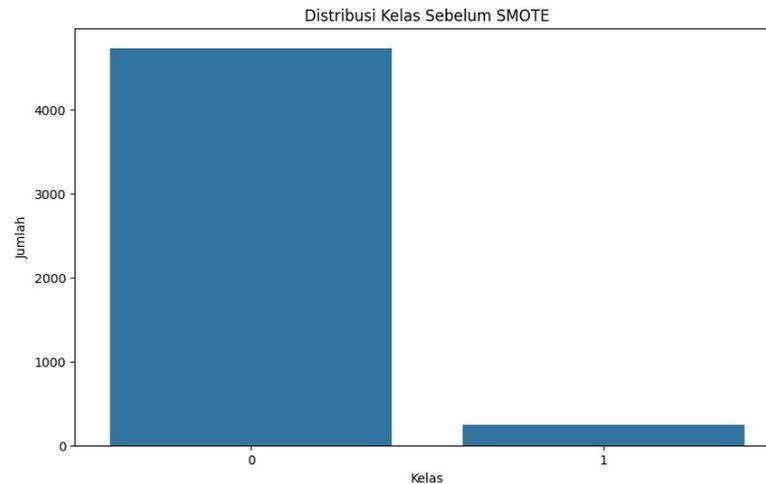
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	age	ypertensi	part_disea	glucose_l	bmi	stroke	nder_Fem	ender_Ma	r_married	r_married	_type_Gov	_type_Pri	pe_Self	_type_chi	nce_type	nce_type	_l_status	Uat	form	status_nev	ig_status	smokes
2	67	0	1	228,69	36,6	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0
3	80	0	1	105,92	32,5	1	0	1	0	1	0	1	0	0	1	0	0	0	0	1	0	0
4	49	0	0	171,23	34,4	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1
5	79	1	0	174,12	24	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0
6	81	0	0	186,21	29	1	0	1	0	1	0	1	0	0	0	1	0	0	1	0	0	0
7	74	1	1	70,09	27,4	1	0	1	0	1	0	1	0	0	1	0	0	0	0	1	0	0
8	69	0	0	94,39	22,8	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0
9	78	0	0	58,57	24,2	1	1	0	0	1	0	1	0	0	0	1	1	0	0	0	0	0
10	81	1	0	80,43	29,7	1	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0
11	61	0	1	120,46	36,8	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1
12	54	0	0	104,51	27,3	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1
13	79	0	1	214,09	28,2	1	1	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0
14	50	1	0	167,41	30,9	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0
15	64	0	1	191,61	37,5	1	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	1
16	75	1	0	221,29	25,8	1	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	1
17	60	0	0	89,22	37,8	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0
18	71	0	0	193,94	22,4	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1

Gambar 4. Dataset Setelah One Hot Encoding

Dapat dilihat pada Gambar 4. Dataset yang dilakukan one hot encoding mengalami penambahan kolom dalam bentuk representasi biner (0 dan 1), hal ini sekaligus menambah jumlah variabel pada dataset.

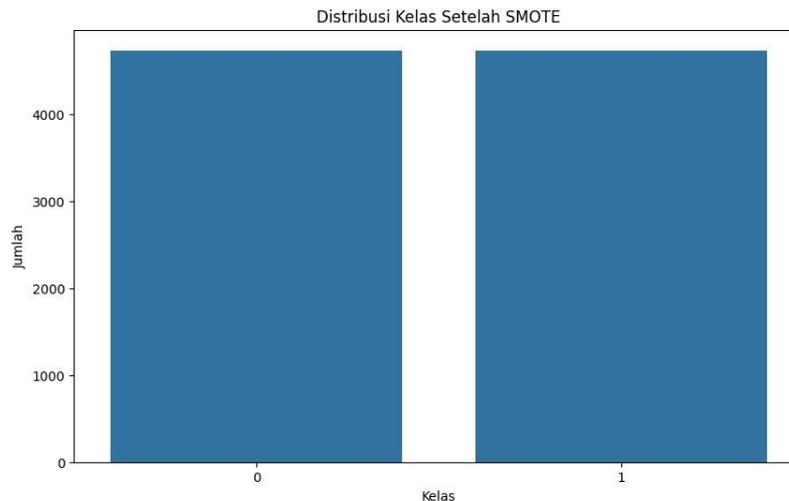
#### c. Data Balancing

Untuk mengatasi ketidakseimbangan data antara kelas stroke dan non-stroke digunakan metode SMOTE (Synthetic Minority Over-sampling Technique). SMOTE bertujuan untuk meningkatkan jumlah data pada kelas minoritas dengan membuat sampel sintetis baru berdasarkan sampel yang ada [18]. Metode ini membantu mengurangi ketidakseimbangan kelas dalam data dengan menambahkan variasi pada kelas minoritas, sehingga model dapat mempelajari pola dengan lebih baik [19] yang terdapat pada Gambar 5 dan Gambar 6.



**Gambar 5.** Sebelum SMOTE

Pada gambar diatas dapat dilihat bahwa dataset yang digunakan memiliki ketimpangan data atau data yang tidak seimbang cukup jauh, dimana kelas stroke hanya 248 data sementara kelas non-stroke berjumlah 4733 data.



**Gambar 6.** Setelah SMOTE

Setelah proses SMOTE dilakukan, jumlah dataset meningkat dari 4981 data menjadi 9466 data. Jumlah sampel dalam kelas stroke ditingkatkan hingga setara dengan jumlah sampel dalam kelas non-stroke.

#### d. Normalisasi Data

Normalisasi merupakan pendekatan reprocessing yang salah satu mengubah variabel dalam suatu dataset untuk memenuhi kondisi tertentu atau memiliki skala yang seragam [20]. Tujuan normalisasi data adalah untuk menghilangkan perbedaan skala atau rentang nilai antara variabel yang berbeda dalam dataset. Dalam penelitian ini digunakan metode min-max scaling untuk melakukan normalisasi data [4]. Berikut persamaan (1):

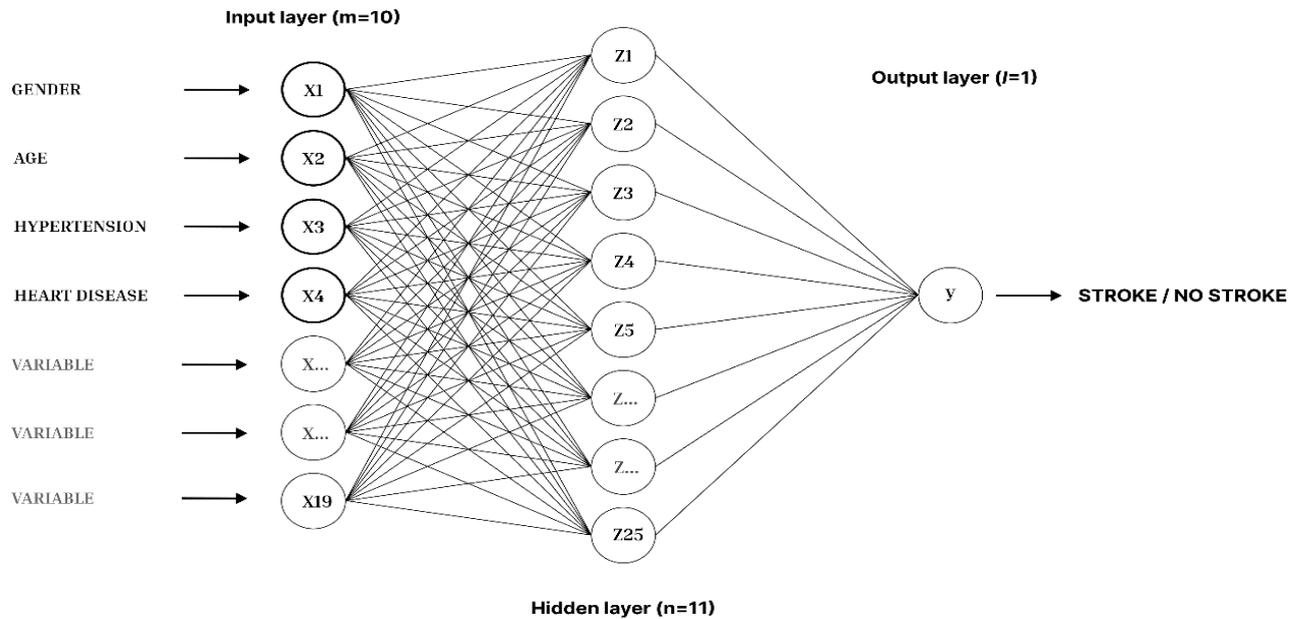
$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A \quad (1)$$

Teknik penskalaan nilai dengan menggunakan rentang spesifik, biasanya dari 0 hingga 1, diterapkan dalam proses ini. Langkah awalnya adalah dengan mengurangi setiap angka variabel dengan nilai terendah yang ada pada variabel tersebut.

## 2.5 BPNN

Jaringan Syaraf Tiruan Back Propagation (BPNN) merupakan algoritma pembelajaran terawasi yang terdiri dari beberapa bagian, yaitu *input layer*, *hidden layer*, dan *output layer* yang selanjutnya terjadi perubahan bobot yang menghubungkan masing-masing layer [13]. Jaringan ini bekerja berdasarkan dua fase, yaitu fase umpan maju dan fase umpan balik. Pada fase umpan maju, data dari lapisan masukan diteruskan ke lapisan keluaran melalui lapisan tersembunyi. Jika rentang kesalahan tidak berada dalam batas yang ditentukan sesuai dengan nilai target pada neuron lapisan keluaran, data keluaran akan dikirim kembali ke lapisan masukan melalui lapisan tersembunyi untuk penyesuaian lebih lanjut. Arsitektur umum BPNN dapat dilihat pada Gambar 7 [21].

BPNN memanfaatkan nilai kesalahan keluaran untuk menyesuaikan bobot-bobotnya secara terbalik (backward). Sebelum menghitung kesalahan ini, tahapan perambatan ke depan (forward propagation) harus dilakukan terlebih dahulu. Fungsi aktivasi yang digunakan dalam BPNN harus memenuhi kriteria kesinambungan, dapat dideferensiasi, dan tidak menurun. Contoh fungsi aktivasi yang memenuhi ketiga syarat ini adalah sigmoid, ReLU, logsig, tansig, dan purelin.



Gambar 7. Arsitektur BPNN

Jumlah neuron pada hidden layer ditentukan menggunakan persamaan yang dirujuk dari buku [22] Berikut persamaan (2) :

$$i < m < 2i \tag{2}$$

Dimana  $m$  merupakan neuron hidden dan  $i$  merupakan neuron input.

Berdasarkan gambar di atas jaringan menggunakan arsitektur 19-25-1, dengan 19 neuron di lapisan input, 19 neuron di hidden layer, dan 1 neuron di lapisan output. Neuron di lapisan input berasal dari variabel dalam dataset yang digunakan.

### 2.6 Confusion matrix

Confusion matrix digunakan dalam penelitian ini untuk mengukur performa dari model ini. Confusion matrix diperlukan untuk melakukan proses pengukuran pada model karena nilai kinerja klasifikasi pada dataset tidak semua hasilnya menyentuh akurasi 100% [23]. Pada **Tabel 1** confusion matrix memiliki 4 kategori hasil prediksi. True Positive (TP) adalah Jumlah observasi yang secara benar diklasifikasikan sebagai positif oleh model, True Negative (TN) adalah Jumlah observasi yang secara benar diklasifikasikan sebagai negatif oleh model, False Positive (FP) yaitu Jumlah observasi yang seharusnya diklasifikasikan sebagai negatif, tetapi salah diklasifikasikan sebagai positif oleh model juga dikenal sebagai kesalahan Type I, False Negative (FN) Jumlah observasi yang seharusnya diklasifikasikan sebagai positif, tetapi salah diklasifikasikan sebagai negatif oleh model. Juga dikenal sebagai kesalahan Type II.

Tabel 1. Confusion Matrix [21]

Kelas Benar	Stroke	Tidak Stroke
Stroke	True Positive	False Positive
Tidak Stroke	False Negative	True Negative

Dari confusion matrix lah didapatkan nilai *accuracy*, *precision*, *recall* dan *f1 score*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

*Precision* adalah rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dari semua kelas positif yang telah diprediksi dengan benar, berapa banyak data yang benar-benar positif.

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

*Recall* atau *sensivity* adalah akurasi yang diprediksi secara positif yang menjelaskan berapa banyak yang diberi label dengan benar

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

F1 Score adalah perbandingan rata-rata dari precision dan recall yang dibobotkan.

$$f1\ score = 2x \frac{Precision \times Recall}{(Precision+Recall)} \tag{6}$$

### 3.6 Parameter Uji

Penelitian ini dilakukan dengan beberapa pengujian model pola BPNN yang terdiri dari tiga lapisan yaitu lapisan input, tersembunyi, dan output lalu dibagi kedalam tiga skenario yaitu [19:25:1], [19:29:1], dan [19:35:1], learning rate ( $\alpha = 0.1$ ), ( $\alpha = 0.01$ ) dan ( $\alpha = 0.001$ ) dengan masing-masing skenario akan dilakukan epoch sebanyak 50 iterasi serta perbandingan data latih dan data uji yang digunakan dalam pengujian yakni 70%:30%, 80%:20% dan 90%:10% dari jumlah dataset serta validation test ialah 20% dari data latih.

**Tabel 2.** Pembagian Data

Pembagian Data	Data Train	Data Test	Learning Rate
90:10	8.519	947	0.1
80:20	7.573	1.893	0.1
70:30	6.626	2.940	0.1
80:20	7.573	1.893	0.01
70:30	6.626	2.940	0.01
80:20	7.573	1.893	0.01
80:20	7.573	1.893	0.001
70:30	6.626	2.940	0.001
80:20	7.573	1.893	0.001

## 3. HASIL DAN PEMBAHASAN

### 3.1 Implementasi BPNN

Pengujian menggunakan Library Tensorflow yang ada pada Keras dengan menggunakan bahasa pemrograman Python. Pengujian dilakukan pada Google Colab dengan memanggil library BPNN. Pengujian dilakukan pada lima model arsitektur berbeda yang diuji dengan parameter uji yang telah ditetapkan pada tahap sebelumnya. Library algoritma BPNN yang digunakan dari tensorflow ada pada Gambar 8.

```

1. import tensorflow as tf
2. from tensorflow.keras.models import Sequential
3. from tensorflow.keras.layers import Dense

# Step 3: Build and train the model
model = Sequential([
    Dense(19, input_shape=(x_train.shape[1],), activation='relu'),
    Dense(21, activation='relu'),
    Dense(1, activation='sigmoid')
])
    
```

**Gambar 8.** Implementasi BPNN

Model arsitektur yang digunakan pada Gambar 8 ialah model arsitektur BPNN [19;21;1] dengan 10 neuron pada layer input, 19 neuron pada hidden layer, dan 1 neuron pada layer output. Neuron di lapisan input berasal dari variabel dalam dataset yang digunakan.

### 3.5 Pembahasan

Pengujian dilakukan sebanyak 27 skema pengujian model pola BPNN yang dibagi kedalam 3 skenario yaitu [19:25:1], [19:29:1], dan [19:35:1] dan Learning rate ( $\alpha = 0.1$ ), ( $\alpha = 0.01$ ) dan ( $\alpha = 0.001$ ) dengan masing-masing skenario akan dilakukan pengujian sebanyak tiga kali masing-masing skema lalu diambil akurasi rata-rata dengan epoch sebanyak 50 iterasi serta perbandingan data latih dan data uji yang digunakan dalam pengujian yakni 70%:30%, 80%:20% dan 90%:10% dari jumlah dataset serta validation test ialah 20% dari data latih.

Hasil pengujian 27 skema dapat dilihat pada Tabel 3-5 berikut:

**Tabel 3.** Arsitektur Model 19 : 25 : 1

Latih : uji	Learning Rate	Akurasi (%)	Presisi	Recall	F1
70 : 30	0.1	95.18	95	95	95

80 : 20	0.1	95.58	96	95	96
<b>90 : 10</b>	<b>0.1</b>	<b>95.81</b>	<b>96</b>	<b>96</b>	<b>96</b>
70 : 30	0.01	95.45	95	96	95
80 : 20	0.01	95.71	95	95	95
90 : 10	0.01	95.07	95	95	95
70 : 30	0.001	95.49	95	95	95
80 : 20	0.001	95.81	96	96	96
90 : 10	0.001	95.81	96	96	96

Dari hasil pengujian pada Tabel 3, untuk model arsitektur 19-25-1. Akurasi tertinggi yang diperoleh yaitu 95.81% serta hasil evaluasi dengan nilai presisi 96%, Recall 96%, dan f-1 96% pada pembagian data latih 90:10 dengan learning rate (0.1). Sedangkan akurasi terendah yaitu 95.07% serta hasil evaluasi dengan nilai presisi 95%, Recall 95%, dan f-1 score 95% dengan menggunakan learning rate (0,01) dan pembagian data latih 90:10.

**Tabel 4.** Arsitektur Model 19 : 29 : 1

Latih : Uji	Learning Rate	Akurasi (%)	Presisi	Recall	F1
70 : 30	0.1	94.11	94	94	93
80 : 20	0.1	95.46	95	95	95
90 : 10	0.1	95.77	96	96	96
70 : 30	0.01	95.47	96	95	95
80 : 20	0.01	95.74	96	96	96
90 : 10	0.01	95.71	96	95	95
70 : 30	0.001	95.54	96	95	95
<b>80 : 20</b>	<b>0.001</b>	<b>96.14</b>	<b>96</b>	<b>96</b>	<b>96</b>
90 : 10	0.001	95.67	96	95	95

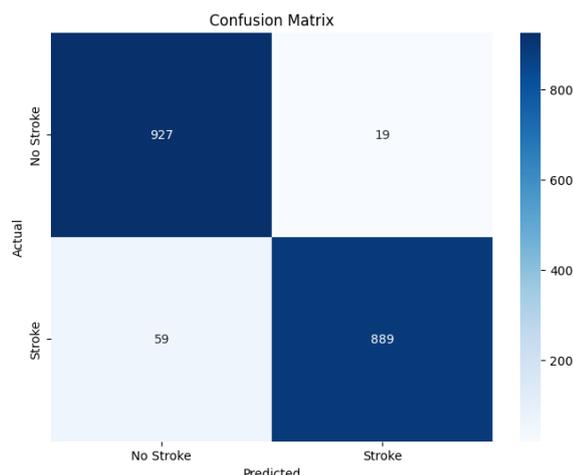
Dari hasil pengujian pada Tabel 4, untuk model arsitektur 19-29-1 Akurasi tertinggi yang diperoleh yaitu 96.14% serta hasil evaluasi dengan nilai presisi 96%, Recall 96%, dan f-1 96% pada pembagian data latih 80:20 dengan learning rate (0.001). Sedangkan akurasi terendah yaitu 94.11% serta hasil evaluasi dengan nilai presisi 94%, Recall 94%, dan f-1 score 93% dengan menggunakan learning rate (0,1) dan pembagian data latih 70:30.

**Tabel 5.** Arsitektur Model 19 : 35 : 1

latih : Uji	Learning Rate	Akurasi (%)	Presisi	Recall	F1
70 : 30	0.1	95.04	95	95	95
80 : 20	0.1	95.78	96	95	95
90 : 10	0.1	95.66	96	96	96
70 : 30	0.01	95.66	96	96	96
80 : 20	0.01	95.53	95	95	95
<b>90 : 10</b>	<b>0.01</b>	<b>96.06</b>	<b>96</b>	<b>95</b>	<b>95</b>
70 : 30	0.001	95.11	95	96	96
80 : 20	0.001	95.81	96	96	96
90 : 10	0.001	95.60	95	95	96

Dari hasil pengujian pada Tabel 5, untuk model arsitektur 19-35-1. Akurasi tertinggi yang diperoleh yaitu 96.06% serta hasil evaluasi dengan nilai presisi 96%, Recall 95%, dan f-1 95% pada pembagian data latih 90:10 dengan learning rate (0.1). Sedangkan akurasi terendah yaitu 95.07% serta hasil evaluasi dengan nilai presisi 95%, Recall 95%, dan f-1 score 95% dengan menggunakan learning rate (0,01) dan pembagian data latih 90:10.

Berdasarkan pengujian yang dilakukan sebanyak 27 skema model pola BPNN, dengan variasi arsitektur [19:25:1], [19:29:1], dan [19:35:1] serta learning rate ( $\alpha$ ) sebesar 0.1, 0.01, dan 0.001, ditemukan bahwa akurasi tertinggi sebesar 96.14% dicapai oleh model arsitektur [19:29:1] dengan learning rate 0.001 dan pembagian data latih dan data uji 80:20. Hasil evaluasi untuk model ini menunjukkan nilai presisi, recall, dan f-1 score masing-masing sebesar 96%. Sebaliknya, akurasi terendah sebesar 94.11% dengan nilai presisi, recall, dan f-1 score masing-masing sebesar 94% dicapai oleh model yang sama namun dengan learning rate 0.1 dan pembagian data latih dan data uji 70:30. Secara umum, semua arsitektur memiliki rentang akurasi antara 95% hingga 96% dengan satu-satunya pengecualian akurasi terendah yaitu 94% pada model [19:29:1] yang menggunakan learning rate (0,1) dan pembagian data latih 70:30, dengan pembagian data latih dan uji 90:10 menunjukkan tingkat akurasi yang lebih stabil dan rata-rata lebih tinggi dibandingkan skema lainnya. Pengaruh learning rate terhadap hasil akhir tidak konsisten, menunjukkan bahwa hasil lebih dipengaruhi oleh parameter lain seperti pembagian data model arsitektur lain yang diusulkan.



Gambar 9. Confusion Matrix

Pada Gambar 9 terdapat hasil confusion matrix dengan menggunakan model BPNN [19:29:1] dengan pembagian data 80:20 dan learning rate 0,001 diperoleh pada hasil kelas stroke yang diprediksi stroke sebanyak 889 serta kelas stroke yang diprediksi tidak stroke sebanyak 59, dan kelas tidak stroke diprediksi tidak stroke sebesar 927 kemudian kelas tidak stroke yang diprediksi stroke sebanyak 59.

#### 4. KESIMPULAN

Metode BPNN yang diusulkan dapat melakukan klasifikasi penyakit stroke dengan baik dan cepat. Dari 27 skema yang telah diuji, ditemukan bahwa learning rate dan model yang diusulkan menghasilkan tingkat akurasi yang tidak konsisten namun cukup tinggi. Selain itu, pembagian data latih yang optimal cenderung menghasilkan hasil yang lebih stabil dibandingkan dengan pembagian data latih yang lebih rendah. Penggunaan metode SMOTE (Synthetic Minority Over-sampling Technique) untuk mengatasi ketimpangan data juga memberikan dampak positif. Sebelumnya, pengujian dilakukan tanpa menggunakan SMOTE dan menghasilkan akurasi yang lebih rendah. Setelah menggunakan SMOTE, akurasi meningkat secara signifikan, menunjukkan bahwa teknik oversampling ini efektif dalam mengatasi masalah ketidakseimbangan kelas. Hasil penelitian menunjukkan bahwa variasi arsitektur model BPNN tidak mempengaruhi hasil akhir secara signifikan. Hal ini terlihat pada table 3, 4, dan 5 di mana akurasi yang diperoleh untuk setiap model BPNN yang diuji hampir memiliki rentang yang sama. Kombinasi yang tepat antara arsitektur, learning rate, dan pembagian data sangat penting untuk mencapai performa model yang optimal. Model terbaik dari semua skema menghasilkan akurasi sebesar 96,14%, yang dicapai pada skenario 2 dengan model arsitektur [19;29;1], penggunaan learning rate  $\alpha = 0,001$ , dan pembagian data latih 80% serta data uji 20%. Parameter ini juga menghasilkan presisi, recall, dan f-1 score sebesar 0,96. Penelitian ini menegaskan potensi BPNN sebagai alat bantu dalam diagnosis dini penyakit stroke, yang dapat digunakan oleh praktisi kesehatan untuk meningkatkan efektivitas penanganan dan pengelolaan pasien stroke. Penelitian ini berhasil dalam melakukan penerapan metode BPNN untuk klasifikasi penyakit stroke. Walau demikian, terdapat beberapa aspek yang dapat ditingkatkan untuk penelitian mendatang seperti melakukan meningkatkan model arsitektur dengan menambah jumlah lapisan pada hidden layer, penggunaan metode oversampling yang lain untuk memperoleh hasil yang lebih optimal, serta mengikutsertakan tenaga medis atau ahli terkait dalam menerjemahkan hasil klasifikasi penyakit stroke dapat memberikan sudut pandang medis yang lebih mendalam dan mengaitkan hasil klasifikasi dengan dampak medis yang lebih luas.

#### REFERENCES

- [1] Z. Zuriati and N. Qomariyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN)," *ROUTERS J. Sist. dan Teknol. Inf.*, vol. 1, no. 1, pp. 1–8, 2022, doi: 10.25181/rt.v1i1.2665.
- [2] N. A. Iskandar, I. Ernawati, and Y. Widiastwi, "Klasifikasi Diagnosis Penyakit Stroke Dengan Menggunakan Metode Random Forest," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 432–441, 2022, [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/2190>
- [3] H. Bugis, "Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 6, no. 1, pp. 8–14, 2022, doi: 10.47970/siskom-kb.v6i1.317.
- [4] K. Julita, I. Afrianty, S. Sanjaya, and F. Syafria, "Penerapan Fuzzy Backpropagation Neural Network dalam Klasifikasi Penyakit Stroke," vol. 8, no. 2, 2023.
- [5] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, "Classification of stroke disease using machine learning algorithms," *Neural Comput. Appl.*, vol. 32, no. 3, pp. 817–828, 2020, doi: 10.1007/s00521-019-04041-y.
- [6] 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/zsettrkalpakbal/full-filled->

- [7] B. Tan, "Back Propagation Neural Network Based Stroke Prediction," *CIBDA 2022 - 3rd Int. Conf. Comput. Inf. Big Data Appl.*, no. February 2022, pp. 56–59, 2022.
- [8] S. J. A. Sarosa, F. Utaminigrum, and F. A. Bachtiar, "Breast cancer classification using GLCM and BPNN," *Int. J. Adv. Soft Comput. its Appl.*, vol. 11, no. 3, pp. 157–172, 2019.
- [9] P. N. Napitupulu, A. R. Damanik, and J. E. Napitupulu, "Implementasi Algoritma Backpropagation Jaringan Syaraf Tiruan Untuk Prediksi Angka Harapan Hidup Di Kota Jambi," *J. JPILKOM ( J. Penelit. Ilmu Komput. )*, vol. 1, no. 1, pp. 10–15, 2023.
- [10] A. Fany Achmalia, S. Walid, I. Artikel, and S. Artikel, "Peramalan Penjualan Semen Menggunakan Backpropagation Neural Network Dan Recurrent Neural Network," *UNNES J. Math.*, vol. 9, no. 1, pp. 6–21, 2020, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [11] H. Elvaningsih, Elisawati, F. Tawakal, and Masrizal, "Seminar Nasional Sains dan Teknologi Informasi (SENSASI) Prediksi Stok Obat Menggunakan Metode Backpropagation (Studi Kasus: Puskesmas Dumai Barat)," *Semin. Nas. Sains dan Teknologi Inf.*, pp. 228–232, 2021, [Online]. Available: <http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage%7C228>
- [12] I. P. J. and A. A. Rao, "Prediction of Diabetes with a BPNN-NB ensemble classifier," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 5, pp. 1652–1657, 2019, doi: 10.26438/ijcse/v7i5.16521657.
- [13] M. Waail *et al.*, "Klasifikasi Jenis Kelengkeng Berdasarkan Morfologi Daun Dengan Ekstraksi Ciri RGB, GLCM, dan Bentuk Menggunakan Metode BPNN" vol. 4, no. 2, pp. 183–193, 2023.
- [14] Istiqomatul Fajriyah Yuliati, Septie Wulandary, and P. Sihombing, "Penerapan Metode SVM dan BPNN dalam Pengklasifikasian PUS di Jawa Barat," *J. Stat. dan Apl.*, vol. 4, no. 1, pp. 23–34, 2020, doi: 10.21009/jsa.04103.
- [15] F. Nabila, "Implementasi Algoritma C4.5 dalam Melakukan Klasifikasi Penyakit Stroke Otak," *J. Ekon. Vol. 18, Nomor 1 Maret 201*, vol. 2, no. 1, pp. 41–49, 2023.
- [16] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [17] C. Herdian, A. Kamila, F. Feiters Tampinongkol, A. S. Kembau, G. Agung, and M. Budidarma, "One-Hot Encoding Feature Engineering Untuk Label-Based Data Studi Kasus Prediksi Harga Mobil Bekas," *J. Inf. Interaktif*, vol. 9, no. 1, pp. 10–16, 2024.
- [18] M. Ibnu and C. Rachmatullah, "Penerapan SMOTE untuk Meningkatkan Kinerja Klasifikasi Penilaian Kredit," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 1, pp. 2407–389, 2023, doi: 10.30865/jurikom.v10i1.5612.
- [19] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 379, 2020, doi: 10.26418/jp.v6i3.42896.
- [20] B. Hauriza, M. Muladi, and I. M. Wirawan, "Prediksi Tingkat Inflasi Bulanan Indonesia Menggunakan Metode Jaringan Saraf Tiruan," *J. Teknol. dan Inf.*, vol. 11, no. 2, pp. 152–167, 2021, doi: 10.34010/jati.v11i2.4924.
- [21] S. Janani, R. Thenmozhi, and L. S. Jayagopal, "Theoretical Investigations for the Verification of Shear Centre and Deflection of Sigma Section by Back Propagation Neural Network Using Python," *Arch. Civ. Eng.*, vol. 65, no. 2, pp. 181–192, 2019, doi: 10.2478/ace-2019-0027.
- [22] S. Rajasekaran and G. A. V PAI, *NEURAL NETWORKS, FUZZY LOGIC AND GENETIC ALGORITHM: SYNTHESIS AND APPLICATIONS (WITH CD)*. PHI Learning, 2003. [Online]. Available: <https://books.google.co.id/books?id=bVbj9nhvHd4C>
- [23] M. A. Muslim *et al.*, *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer*. 2019.