

**KLASIFIKASI *PROBABILISTIC NEURAL NETWORK* (PNN)  
PADA DATA DIAGNOSA PENYAKIT DEMAM BERDARAH  
DENGUE (DBD) TAHUN 2018**

Reza Destiana Adyati<sup>1\*</sup>, Yuki Novia Nasution<sup>1</sup>, Sri Wahyuningsih<sup>1</sup>

<sup>1</sup>Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas  
Mulawarman, Indonesia

*Corresponding author:* rezadestiana@gmail.com

**Abstrak.** *Probabilistic Neural Network* (PNN) merupakan salah satu model pada Jaringan Syaraf Tiruan (JST) yang digunakan untuk pengklasifikasian. PNN sangat tergantung pada parameter *smoothing* ( $\sigma$ ) yang digunakan. Pada penelitian ini dilakukan pengacakan data dengan *Simple Random Sampling* (SRS) untuk mengetahui nilai  $\sigma$  dan proporsi data *training* dan data *testing* yang memberikan nilai akurasi tertinggi. Pada penelitian ini PNN akan diterapkan pada kasus Demam Berdarah Dengue (DBD). Dari klasifikasi PNN dengan pengacakan data *training* dan data *testing* pada pasien yang melakukan uji hematologi dengan menggunakan 6 faktor yaitu jenis kelamin, usia, kadar hemoglobin (hb), jumlah leukosit, jumlah trombosit, kadar hematokrit (ht), didapatkan hasil bahwa tingkat akurasi klasifikasi pasien DBD menggunakan metode PNN yang paling baik adalah pada proporsi data *testing* dan data *training* 90:10 dengan  $\sigma = 0,2$  dan  $\sigma = 0,3$ . Hasil akurasi antar data *training* sebesar 100% dan akurasi antara data *testing* terhadap data *training* sebesar 93%.

**Kata Kunci:** Demam berdarah dengue, klasifikasi, *simple random sampling*, *probabilistic neural network*.

## **1 PENDAHULUAN**

*Neutral Network* atau Jaringan Saraf Tiruan (JST) adalah suatu sistem informasi yang dimodelkan berdasarkan sistem saraf manusia, salah satu model *neural network* yang sering digunakan untuk pengklasifikasian adalah *probabilistic Neural Network* (PNN). PNN adalah suatu algoritma yang menggunakan fungsi probabilitas, tidak memerlukan set data yang besar dalam tahap pengerjaannya, serta memiliki kelebihan yang dapat mengatasi permasalahan yang ada pada *back propagation* yaitu dapat mengatasi waktu pelatihan yang lama, masalah pada global minimum, dan kesulitan perencanaan arsitektur jaringan. Pembelajaran pada PNN dikategorikan sebagai pembelajaran yang tidak terawasi sehingga tidak diperlukan proses pelatihan yang berulang (iterasi) untuk memperbaiki parameter-parameter yang nantinya akan digunakan dalam mengenali kelas dari suatu data. Parameter yang dimaksud adalah *smoothing parameter* [1].

Demam Berdarah Dengue (DBD) merupakan salah satu masalah kesehatan masyarakat utama di Indonesia. Perjalanan penyakit ini sangat cepat dan sering menjadi fatal karena banyak pasien yang meninggal akibat penanganannya yang terlambat. Saat ini, 40% populasi dunia tinggal di daerah yang memiliki resiko terinfeksi DBD. Asia menempati urutan pertama dalam jumlah penderita DBD setiap tahunnya. Sementara itu, sejak tahun 1968 hingga tahun 2009, *World Health Organization* (WHO) mencatat negara Indonesia sebagai negara dengan kasus DBD tertinggi di Asia Tenggara [2].

Pada tahun 2016, jumlah penderita DBD yang dilaporkan melalui Sistem Informasi Daerah (SIKDA) Samarinda sebanyak 2.814 kasus dengan jumlah kematian sebanyak 18 orang. Angka kesakitan DBD di Samarinda tergolong tinggi. Kalimantan Timur merupakan salah satu Provinsi dengan angka penderita DBD tertinggi kedua di Indonesia setelah Provinsi Bali. Kematian akibat DBD di Samarinda tergolong rendah karena angka kematian kurang dari 1% [3]. DBD merupakan penyakit yang disebabkan oleh virus Dengue yang ditularkan dari orang ke orang melalui gigitan nyamuk *Aedes aegypti* (Ae). Diperlukan deteksi dini apakah seorang terinfeksi atau tidaknya penyakit DBD, karena jika terlambat ditangani hal ini dapat berujung pada kematian [4]. Dengan demikian, penulis mengusulkan penelitian yang berjudul “Klasifikasi *Probabilistic Neural Network* (PNN) Pada Data Diagnosa Penyakit Demam Berdarah Dengue (DBD) di Puskesmas Temindung Tahun 2018”.

## **2 TINJAUAN PUSTAKA**

### **2.1 Simple Random Sampling (SRS)**

Metode *Simple Random Sampling* (SRS) adalah metode yang digunakan untuk memilih sampel dari populasi dengan cara sedemikian rupa sehingga setiap anggota populasi mempunyai peluang yang sama besar untuk diambil sebagai sampel [5].

### **2.2 Jaringan Syaraf Tiruan**

Jaringan saraf tiruan (JST) didefinisikan sebagai suatu sistem pemrosesan informasi yang mempunyai karakteristik menyerupai jaringan saraf manusia. JST

disebut juga *brain metaphor*, *computational neuroscience* atau *parallel distributed processing* serta *connection*. JST tersusun dari sejumlah besar elemen yang melakukan kegiatan yang analog dengan fungsi-fungsi biologis *neuron* yang paling elementer. Elemen-elemen ini terorganisasi sebagaimana layaknya anatomi otak, walaupun tidak persis.

Berbeda dengan metode lain, algoritma untuk JST beroperasi secara langsung dengan angka sehingga data yang tidak numerik harus diubah menjadi data numerik. Dibandingkan dengan cara perhitungan konvensional, JST tidak memerlukan atau menggunakan suatu model matematis atas permasalahan yang dihadapi. Oleh karena itu JST juga dikenal sebagai model *free-estimator* [6].

### 2.3 Probabilistic Neural Network (PNN)

PNN merupakan suatu metode JST yang menggunakan prinsip dari teori statistik yaitu *Bayesian Classification* untuk menggantikan prinsip heuristik yang digunakan oleh algoritma *back propagation* [1]. Proses yang dilakukan oleh PNN dapat berlangsung lebih cepat bila dibandingkan dengan JST *back propagation*. Hal ini dikarenakan PNN hanya membutuhkan satu kali iterasi pelatihan bila dibandingkan dengan JST *back propagation* yang membutuhkan beberapa kali iterasi pelatihan. Selain itu, PNN tidak membutuhkan set data yang besar dalam tahap pembelajarannya, serta memiliki kelebihan dapat mengatasi permasalahan yang ada pada JST *back propagation* yaitu dapat mengatasi waktu pelatihan yang lama dan terjebak pada global minimum. Namun, PNN memiliki masalah pada penentuan parameter *smoothing* yang biasanya digunakan dengan cara *trial and error* atau *user defined* [7].

### 2.4 Algoritma PNN

Algoritma pengklasifikasian pada PNN dengan memisalkan matriks V adalah data *training* sebagai berikut [8]:

1. Menetapkan matriks input  $\mathbf{X}$  dimana data *input* dapat berupa data *training* dan data *testing*, maka matriks  $\mathbf{X}$  terdiri dari dua yaitu matriks  $\mathbf{X}_{training}$  dan matriks

$$\mathbf{X}_{testing} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{k1} & X_{k1} & \dots & X_{kj} & \dots & X_{kp} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{q1} & X_{q2} & \dots & X_{qj} & \dots & X_{qp} \end{bmatrix} \quad \text{dimana } \begin{matrix} k = 1, 2, \dots, q \\ j = 1, 2, \dots, p \end{matrix}$$

2. Menetapkan matriks target  $\mathbf{W}$  yang bersesuaian dengan matriks  $\mathbf{V}$

$$\mathbf{V} = \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1j} & \dots & V_{1p} \\ V_{21} & V_{22} & \dots & V_{2j} & \dots & V_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ V_{i1} & V_{i2} & \dots & V_{ij} & \dots & V_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ V_{n1} & V_{n2} & \dots & V_{nj} & \dots & V_{np} \end{bmatrix}; \mathbf{W} = \begin{bmatrix} W_{1h} \\ W_{2h} \\ \vdots \\ W_{ih} \\ \vdots \\ W_{nh} \end{bmatrix}$$

dimana  $i = 1, 2, \dots, n$

$h = \text{kelas data yang bersesuaian dengan data } training (1, 2, \dots, r)$

2. Menetapkan bobot bias lapisan input

$$b_i = \sqrt{\frac{-\ln(0,5)}{\sigma}} \quad (1)$$

3. Masing-masing poin berikut dikerjakan untuk  $k = 1$  sampai  $q$  pada semua data *input* (data *training* dan data *testing*).

- a. Mencari jarak matriks  $\mathbf{X}$  terhadap matriks sebagai berikut:

$$D_{ki} = \sqrt{\sum_{j=1}^p (X_{kj} - V_{ij})^2} \quad (2)$$

- b. Mencari nilai aktivasi  $a_{ki}$  sebagai berikut:

$$a_{ki} = e^{-(b_i D_{ki})^2} \text{ dengan } i = 1, 2, \dots, n \quad (3)$$

- c. Mencari nilai  $c_{kh}$  sebagai berikut :

$$c_{kh} = \sum_{i=1}^n (a_{ki} W_{ih}) \text{ dengan } h = 1, 2, \dots, r \quad (4)$$

- d. Mencari nilai maksimum ( $z$ ) hingga  $c_{kz} = \max \{c_{kh} \mid h = 1, 2, \dots, r\}$ .

- e. Menetapkan  $z$  sebagai kelas dari klasifikasi.

## 2.5 Akurasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar. Akan tetapi, tidak dapat dipungkiri bahwa kinerja suatu tidak bisa bekerja 100% benar. Oleh karena itu, sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya cara mengukur kinerja klasifikasi menggunakan *confusion matrix*. *Confusion matrix* merupakan tabel yang mencatat hasil kerja klasifikasi. Tabel 1 merupakan contoh matriks *confusion* yang melakukan klasifikasi masalah biner misalnya kelas 0 dan 1. Misalnya sel  $f_{11}$  adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, dan  $f_{10}$  adalah data dalam kelas 1 yang dipetakan secara salah ke kelas 0 [9].

Tabel 1: *Confusion Matrix* untuk Klasifikasi Dua Kelas

$f_{ij}$		Kelas Hasil Prediksi ( $j$ )	
		Kelas = 1	Kelas = 0
Kelas asli ( $i$ )	Kelas = 1	$f_{11}$	$f_{10}$
	Kelas = 0	$f_{01}$	$f_{00}$

Berdasarkan ini *confusion matrix*, dapat diketahui jumlah data dari masing-masing kelas yang diprediksi secara benar yaitu  $(f_{11} + f_{00})$  dan data yang di klasifikasikan secara salah yaitu  $(f_{10} + f_{01})$ . Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksi. Untuk menghitung akurasi digunakan formula sebagai berikut:

$$\begin{aligned}
 \text{akurasi} &= \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \\
 &= \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (5)
 \end{aligned}$$

### 3 DATA

Data yang digunakan pada penelitian ini adalah data sekunder dari Puskesmas Temindung Kota Samarinda. Data yang digunakan adalah pasien yang melakukan tes hematologi pada tahun 2018. Variabel penelitian yang digunakan pada penelitian ini terdiri dari 6 variabel sebagai berikut :

1. Jenis Kelamin ( $X_1$ ,  $X_1 = 0 = \text{Pria}$ ,  $X_1 = 1 = \text{Wanita}$ )
2. Usia ( $X_2$ )
3. Kadar Hemoglobin ( $X_3$ )
4. Jumlah Leukosit ( $X_4$ )
5. Jumlah Trombosit ( $X_5$ )
6. Kadar Hematokrit ( $X_6$ )

### 4 HASIL DAN PEMBAHASAN

Dari hasil perhitungan akurasi pada 100 kali pengacakan data untuk seluruh proporsi data *training* dan data *testing* dengan menggunakan sepuluh  $\sigma$  yang berbeda pada setiap proporsinya dilakukan perhitungan rata-rata untuk setiap  $\sigma$  yang berbeda dari setiap proporsinya. Hasil perhitungan rata-rata akurasi adalah sebagai berikut.

Tabel 2: Rata-Rata akurasi

No	Proporsi Data	$\Sigma$	Rata-Rata Akurasi (%)		No	Proporsi Data	$\sigma$	Rata-Rata Akurasi (%)	
			Data Training	Data Testing				Data Training	Data Testing
			1	50:50				0,1	100
2	50:50	0,2	100	92,18	27	70:30	0,7	99,19	92,77
3	50:50	0,3	100	92,36	28	70:30	0,8	99,03	92,73
4	50:50	0,4	99,8	92,54	29	70:30	0,9	98,84	92,83
5	50:50	0,5	99,7	92,54	30	70:30	1	98,57	92,87
6	50:50	0,6	99,62	92,56	31	80:20	0,1	100	91,85
7	50:50	0,7	99,48	92,58	32	80:20	0,2	100	92,1
8	50:50	0,8	99,4	92,74	33	80:20	0,3	100	92,3
9	50:50	0,9	99,2	92,7	34	80:20	0,4	99,53	92,65
10	50:50	1	98,96	92,44	35	80:20	0,5	99,4	92,75
11	60:40	0,1	100	91,85	36	80:20	0,6	99,21	92,8
12	60:40	0,2	100	92,18	37	80:20	0,7	99,04	92,7
13	60:40	0,3	100	92,33	38	80:20	0,8	98,9	92,8
14	60:40	0,4	99,78	92,6	39	80:20	0,9	98,75	92,7
15	60:40	0,5	99,67	92,78	40	80:20	1	98,44	92,55
16	60:40	0,6	99,58	92,65	41	90:10	0,1	100	92,6
17	60:40	0,7	99,43	92,78	42	90:10	0,2	100	93
18	60:40	0,8	99,27	92,75	43	90:10	0,3	100	93
19	60:40	0,9	98,98	92,83	44	90:10	0,4	99,33	93
20	60:40	1	98,82	92,7	45	90:10	0,5	99,24	93,1
21	70:30	0,1	100	92,13	46	90:10	0,6	99,04	93
22	70:30	0,2	100	92,33	47	90:10	0,7	99,01	92,9
23	70:30	0,3	100	92,47	48	90:10	0,8	98,88	92,6
24	70:30	0,4	99,66	92,8	49	90:10	0,9	98,8	92,5
25	70:30	0,5	99,54	92,73	50	90:10	1	98,5	92,6

Dari tabel 2 dapat dilihat bahwa hasil akurasi untuk data *training* dan data *testing* yang paling baik terletak pada proporsi 90:10 dengan nilai akurasi adalah 100% untuk data *training* dan 93% untuk data *testing* dengan  $\sigma = 0,2$  dan  $\sigma = 0,3$ .

## 5 KESIMPULAN

Berdasarkan dari hasil analisis dan pembahasan yang telah dilakukan didapatkan kesimpulan dari klasifikasi PNN pada pasien diagnosa penyakit DBD dengan menggunakan 6 variabel yaitu jenis kelamin, usia, kadar hemoglobin, jumlah trombosit, jumlah leukosit, dan kadar hematokrit didapatkan hasil bahwa tingkat akurasi klasifikasi diagnosa penyakit DBD menggunakan metode PNN yang paling baik yaitu pada proporsi 90:10 dengan  $\sigma = 0,2$  dan  $\sigma = 0,3$ . Hasil rata-rata akurasi antar data *training* sebesar 100% dan akurasi data *testing* terhadap data *training* sebesar 93%.

**DAFTAR PUSTAKA**

- [1] Specht, D. F. (1990). Probabilistic Neural Network. *International Conference on Neural Network*, 3, 109-118.
- [2] Rasyada, A., Nasrul, E. dan Edward, Z. (2014). Hubungan nilai Hematokrit Terhadap Jumlah Trombosit pada penderita Demam Berdarah Dengue. *Jurnal Kesehatan Andalas*, 3(3), 343-347.
- [3] Dinas Kesehatan Kota Samarinda.(2016). *Profil Kesehatan Kota Samarinda Tahun 2016*. Samarinda:Dinas Kesehatan Kota Samarinda.
- [4] Hasibuan, C. A., Mukid, M.A. dan Prahutama A. (2017). Klasifikasi Diagnosa Penyakit Demam Berdarah (DBD) Menggunakan Support Vector Machine (SVM) Berbasis GUI Matlab. *Jurnal Gaussian*, 6(2), 171-180.
- [5] Nurhayati. (2008). Studi Perbandingan Metode Sampling Antara Simple Random Sampling Dengan Stratified Random. *Jurnal Basis Data*, 3(1), 18-32.
- [6] Hermawan, A. dan Suyanto.(2006). *Jaringan Saraf Tiruan Teori dan Aplikasi*. Yogyakarta: Andi.
- [7] Kusmadewi, Sri. (2004b). Klasifikasi Pola Menggunakan Jaringan Probabilistik. *Seminar Nasional Aplikasi Teknologi Informasi 2004*, 1(1), 59-64.
- [8] Wicaksana, A. J., Yasin, Hasbi dan Sudarno. (2016). Probabilistic Neural Network Berbasis GUI Matlab untuk Klasifikasi Data Rekam Medis. *Jurnal Gaussian*, 5(3), 427-436.
- [9] Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: Andi.