

Klasifikasi Data Pasien Penyakit Tuberkulosis Paru Menggunakan Metode Probabilistic Neural Network (Studi Kasus: Puskesmas Telaga Sari Kota Balikpapan)

Classification of Pulmonary Tuberculosis Patient Data Using the Probabilistic Neural Network Method (Case Study: Telaga Sari Public Health Center, Balikpapan City)

Laila Thalia Ramadhini¹, M. Fathurahman^{2a)}, Nariza Wanti Wulan Sari³

^{1,2,3} Program Studi Statistika, FMIPA Universitas Mulawarman, Indonesia

^{1,2,3} Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman, Indonesia

^{a)}Corresponding author: fathur@fmipa.unmul.ac.id

ABSTRACT

Pulmonary tuberculosis is an infectious disease that remains a major health problem in Indonesia. Early detection of this disease is very important to improve the effectiveness of treatment and prevention of its spread. The purpose of this study is to classify laboratory test data of pulmonary tuberculosis patients using the Probabilistic Neural Network method. The data used are medical records of patients with pulmonary tuberculosis disease at Puskesmas Telaga Sari, Balikpapan City in 2023-2024. The variables used are age, weight, systolic blood pressure, diastolic blood pressure, cough duration, fever duration, shortness of breath, and loss of appetite. The classification process involves the stages of encoding, data normalization, division of training data and testing data using a proportion of 80:20, and calculation of accuracy using confusion matrix. The results showed that classification using the Probabilistic Neural Network method was appropriate in classifying pulmonary tuberculosis disease and obtained the best smoothing parameter (α) value of 0.1 with an accuracy value of 82.95% for training data and 95.45% for testing data.

Keywords: *classification, confusion matrix, probabilistic neural network, pulmonary tuberculosis, smoothing parameter*

1. Pendahuluan

Perkembangan ilmu pengetahuan dan teknologi yang semakin pesat terutama meluasnya penggunaan internet oleh banyak pihak, maka semakin banyak pula data yang dihasilkan melalui penggunaan itu baik dari data pribadi maupun data transaksi yang dilakukan. *Database* yang dihasilkan menjadi sangat berharga bagi pihak yang berkepentingan (Srirahayu dan Pribadie, 2023). Salah satu proses pengolahan data yang dapat mengatasi masalah ini adalah *data mining*. Menurut Tarigan, dkk. (2022), *data mining* merupakan proses penggalian informasi untuk menemukan pola-pola penting (*pattern recognition*) pada tumpukan data dalam *database* sehingga menjadi pengetahuan (*discovery of knowledge*). Seiring dengan pesatnya perkembangan teknologi dan informasi, *data mining* mampu menghasilkan informasi dan pengetahuan di berbagai bidang seperti manajemen bisnis, manajemen produksi, dan kesehatan. Banyak metode yang dapat dilakukan menggunakan *data mining*, salah satunya adalah metode klasifikasi. Metode klasifikasi merupakan salah satu metode *data mining* yang mengelompokkan data dari sekelompok variabel dalam menggambarkan dan membedakan kelas label atau target yang bertujuan untuk mengklasifikasi kelas dari objek yang label kelasnya tidak diketahui (Cahyanti, dkk., 2023). Metode yang paling sering digunakan pada klasifikasi salah satunya adalah *Probabilistic Neural Network* (PNN).

Metode PNN merupakan salah satu model *neural network* yang menggunakan fungsi peluang. Metode ini pertama kali diperkenalkan oleh Donald Specht pada Tahun 1990. Metode PNN menggunakan proses *training supervised* untuk proses pembelajaran datanya dan sering digunakan untuk menyelesaikan masalah klasifikasi (Azizah, 2021). Menurut Wenando, dkk. (2023), metode PNN dinilai berpotensi baik dalam mengklasifikasikan dokumen dibandingkan dengan metode klasifikasi lainnya karena memungkinkan prediksi diagnosis penyakit berdasarkan data kesehatan dengan akurasi tinggi meskipun data yang diproses berisi data yang tidak seimbang dan tidak lengkap.

TB paru masih menjadi masalah kesehatan utama di seluruh dunia dan menyerang jutaan orang setiap tahunnya. TB paru merupakan penyakit menular yang disebabkan oleh *mycobacterium tuberculosis* dan ditularkan melalui percikan dahak. Penularan TB paru oleh patogen tuberkulosis ini menyerang paru-paru dan organ tubuh lainnya (Rosamawati, dkk., 2023). TB paru dapat menyebar dari satu orang ke orang lain melalui udara (dahak pada pasien). Jika seorang pasien yang terinfeksi TB paru kemudian bersin atau batuk, maka dapat menghasilkan *droplet* yang mengandung bakteri-bakteri basil TB paru (Lelewana, dkk., 2024). Berdasarkan laporan *World Health Organization* (WHO) Tahun 2023, diperkirakan sudah terdapat 10,6 juta kasus TB paru dan terus bertambah setiap tahunnya. Indonesia termasuk negara dengan penderita TB paru terbesar kedua setelah India dengan proporsi kasus baru sebesar 10% dibandingkan seluruh kasus di dunia.

Kalimantan Timur merupakan salah satu provinsi di Indonesia dengan peningkatan jumlah kasus baru yang cukup tinggi dan berada pada urutan ke 13 (Kementerian Kesehatan Republik Indonesia, 2024). Balikpapan merupakan salah satu kota yang berada di Provinsi Kalimantan Timur dengan kasus TB paru pada Tahun 2022 yaitu sebanyak 2.142 kasus. Meskipun penyakit ini dapat disembuhkan dan dicegah, hal itu tetap menjadikan penyakit TB paru sebagai ancaman global yang sangat serius termasuk salah satunya di Kota Balikpapan (Dinas Kesehatan Provinsi Kalimantan Timur, 2023).

Penelitian-penelitian sebelumnya telah menunjukkan keberhasilan penerapan PNN dalam bidang kesehatan. Misalnya, Azizah (2021) menggunakan PNN untuk mengklasifikasi berat bayi baru lahir dan memperoleh tingkat akurasi tinggi. Kebaruan dari penelitian ini terletak pada penerapan metode PNN untuk klasifikasi data pasien penyakit Tuberkulosis Paru (TB paru) dengan menggunakan data laboratorium sub-lokal dari Puskesmas Telaga Sari Kota Balikpapan. Dengan demikian, penelitian ini bertujuan untuk mengklasifikasikan data pasien penyakit Tuberkulosis Paru menggunakan metode PNN serta menentukan nilai *smoothing parameter* terbaik berdasarkan data pasien di Puskesmas Telaga Sari Kota Balikpapan. Hasil penelitian diharapkan dapat memberikan kontribusi ilmiah terhadap pengembangan metode klasifikasi berbasis PNN dalam konteks data kesehatan sub-lokal serta menjadi referensi bagi puskesmas lain dalam meningkatkan efektivitas diagnosis awal TB paru.

2. Metode Penelitian

2.1 Sumber Data

Penelitian ini menggunakan rancangan yang bersifat *ex post facto*, yakni data penelitian dikumpulkan setelah kejadian tersebut berlangsung atau lewat (Primantoro, dkk., 2024). Sampel yang digunakan pada penelitian ini adalah pasien TB paru di Puskesmas Telaga Sari Kota Balikpapan Tahun 2023-2024. Data yang digunakan merupakan data sekunder, dimana data sekunder merupakan data yang diperoleh dengan cara membaca, mempelajari, dan memahami melalui media lain yang bersumber dari literatur, buku-buku, serta dokumen perusahaan (Sawo, dkk., 2021). Adapun data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari data internal puskesmas berupa rekapitulasi medis pasien yang diambil dari hasil laboratorium.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari variabel target (*Y*) dan 8 variabel input (*X*) yang diberikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Simbol	Variabel	Definisi	Skala Pengukuran	Satuan
Y	Hasil Diagnosis	Hasil diagnosis pasien yang ditentukan berdasarkan hasil pemeriksaan laboratorium. Dengan kategori pasien TB paru dinotasikan sebagai 1 dan kategori hasil diagnosis pasien tidak TB paru dinotasikan sebagai 0	Nominal	—
X_1	Usia	Usia pasien pada saat melakukan pemeriksaan laboratorium di Puskesmas Telaga Sari	Rasio	Tahun
X_2	Berat Badan	Berat badan pasien pada saat melakukan pemeriksaan laboratorium di Puskesmas Telaga Sari	Rasio	Kg
X_3	Tekanan Darah Sistolik	Tekanan darah sistolik pasien pada saat melakukan pemeriksaan laboratorium di Puskesmas Telaga Sari	Rasio	mmHg
X_4	Tekanan Darah Diastolik	Tekanan darah diastolik pasien pada saat melakukan pemeriksaan laboratorium di Puskesmas Telaga Sari	Rasio	mmHg
X_5	Lama Batuk	Lama batuk yang dialami pasien sejak munculnya gejala hingga saat pemeriksaan di Puskesmas Telaga Sari	Rasio	Hari
X_6	Lama Demam	Lama demam yang dialami pasien sejak munculnya gejala hingga saat pemeriksaan di Puskesmas Telaga Sari	Rasio	Hari
X_7	Sesak Napas	Sesak napas yang dialami pasien pada saat melakukan pemeriksaan laboratorium di Puskesmas Telaga Sari. Dengan kategori	Nominal	—

Simbol	Variabel	Definisi	Skala Pengukuran	Satuan
X_8	Hilang Nafsu Makan	pasien TB paru mengalami sesak napas sebagai 1 dan kategori pasien TB paru tidak mengalami sesak napas sebagai 0 Kehilangan nafsu makan yang dialami pasien pada saat melakukan pemeriksaan laboratorium di Puskesmas Telaga Sari. Dengan kategori pasien TB paru mengalami kehilangan nafsu makan sebagai 1 dan kategori pasien TB paru tidak mengalami kehilangan nafsu makan sebagai 0	Nominal	—

2.3 Teknik Sampling

Teknik sampling yang digunakan dalam penelitian ini adalah teknik *purposive sampling* dimana teknik ini didasarkan pada pertimbangan peneliti dalam menentukan sampel (Lenaini, 2021). Dalam hal ini, pertimbangannya adalah ketersediaan dan kelengkapan data terbaru rekam medis pasien TB paru pada puskesmas.

2.4 Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini terdiri dari delapan tahap, yaitu (Azizah, 2021):

1. Melakukan Analisis Statistika Deskriptif
Penyajian data secara deskriptif dilakukan dengan bantuan *software R* untuk 6 variabel *input* yaitu X_1 sampai dengan X_6 , kemudian untuk variabel target dan 2 variabel *input* lainnya yaitu X_7 dan X_8 , penyajian datanya menggunakan diagram lingkaran dengan tujuan agar lebih mudah memahami data yang ditampilkan.

2. Melakukan Normalisasi Data

Nilai normalisasi data ($x_{k,j}^*$) dapat ditentukan dengan menghitung nilai rata-rata dan variansi seperti pada Persamaan (1) dan (2) berikut:

$$\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{k,j} \quad (1)$$

$$s_j^2 = \frac{1}{n-1} \sum_{k=1}^n (x_{k,j} - \bar{x}_j)^2 \quad (2)$$

Berdasarkan perhitungan nilai rata-rata dan variansi pada Persamaan (1) dan (2), langkah selanjutnya adalah menghitung normalisasi data menggunakan Persamaan (3) berikut:

$$x_{k,j}^* = \frac{x_{k,j} - \bar{x}_j}{s_j} \quad (3)$$

dengan

n = banyak data

$x_{k,j}$ = data ke- k pada variabel ke- j dimana $j = 1, 2, \dots, p$

$\bar{x}_j$ = rata-rata variabel ke- j

s_j = simpangan baku variabel ke- j

s_j^2 = variansi variabel ke- j

$x_{k,j}^*$ = normalisasi data pada variabel ke- j .

3. Menentukan Nilai *Smoothing Parameter* (α)
4. Melakukan Pengacakan Data
5. Membagi Data *Training* dan Data *Testing*
6. Melakukan Perhitungan Klasifikasi Metode PNN Menggunakan Data *Training* dengan Langkah-langkah Sebagai Berikut:

- a. Menghitung bobot bias lapisan *input* menggunakan Persamaan (4) berikut:

$$b = \sqrt{\frac{-\ln(0,5)}{\alpha}} \quad (4)$$

dengan

b = bobot bias lapisan *input*

α = *smoothing parameter* dengan $0 < \alpha \leq 1$.

- b. Menghitung jarak Gower antar data *training* menggunakan Persamaan (5), (6), (7), dan (8) berikut.

$$\text{Nominal : } d_{k,i}(X_{k,j}, V_{i,j}) = \begin{cases} 0, & X_{k,j} = V_{i,j} \\ 1, & X_{k,j} \neq V_{i,j} \end{cases} \quad (5)$$

$$\text{Ordinal : } d_{k,i}(X_{k,j}, V_{i,j}) = \frac{|r_j(X_{k,j}) - r_j(V_{i,j})|}{\max\{r_j(X_{k,j})\} - \min\{r_j(X_{k,j})\}} \quad (6)$$

$$\text{Interval; Rasio : } d_{k,i}(X_{k,j}, V_{i,j}) = \frac{|(X_{k,j}) - (V_{i,j})|}{\max\{X_{k,j}\} - \min\{X_{k,j}\}} \quad (7)$$

Perhitungan jarak antara dua objek digunakan persamaan berikut:

$$d_{k,i} = \frac{\sum_{j=1}^p \delta_j(X_{k,j}, V_{i,j}) d_{k,i}}{\sum_{j=1}^p \delta_j(X_{k,j}, V_{i,j})} \quad (8)$$

dengan $\delta_j(X_{k,j}, V_{i,j}) = \begin{cases} 1, & \text{nilai } X_{k,j} \text{ dan } V_{i,j} \text{ tidak kosong atau hilang} \\ 0, & \text{lainnya} \end{cases}$

dimana

$d_{k,i}$ = jarak Gower pada matriks *input* **X** dan matriks **V**

$X_{k,j}$ = data ke k variabel ke j pada matriks **X**

$V_{i,j}$ = data ke i variabel ke j pada matriks **V**

$X_{k,j}$ = nilai seluruh observasi variabel j

$r_j(X_{k,j})$ = peringkat seluruh observasi variabel ordinal j

$\max\{X_{k,j}\}$ = nilai maksimum dari seluruh variabel j

$\min\{X_{k,j}\}$ = nilai minimum dari seluruh variabel j

$\max\{r_j(X_{k,j})\}$ = peringkat maksimum dari seluruh variabel ordinal j

$\min\{r_j(X_{k,j})\}$ = peringkat minimum dari seluruh variabel ordinal j

$\delta_j(X_{k,j}, V_{i,j})$ = pembobot jarak objek X dan V pada variabel j .

- c. Menghitung nilai aktivasi $a1_{ki}$ dengan menggunakan Persamaan (9).

$$a1_{k,i} = e^{-(b d_{k,i})^2}; \text{ dengan } i = 1, 2, \dots, n \quad (9)$$

dengan

$a1_{k,i}$ = nilai aktivasi

b = bobot bias lapisan *input*

e = eksponensial.

- d. Menjumlahkan nilai aktivasi yang dikalikan dengan vektor *input* **W** yang sesuai dengan kelas data *training* menggunakan Persamaan (10).

$$a2_{k,a} = \sum_{i=1}^n (a1_{k,i} * W_{i,a}); \text{ dengan } a = 1, 2 \quad (10)$$

dimana

$a2_{k,a}$ = hasil penjumlahan nilai aktivasi

$W_{i,a}$ = data ke i kelas ke a pada matriks target **W**.

- e. Menentukan kelas data *training* sesuai dengan penjumlahan nilai aktivasi yang terbesar.

7. Melakukan Perhitungan Klasifikasi Metode PNN Menggunakan Data *Testing* Berdasarkan Data *Training* dengan Langkah-langkah Sebagai Berikut:

- a. Menghitung jarak antara data *training* dan data *testing* menggunakan Persamaan (5), (6), (7), dan (8).

- b. Menghitung nilai aktivasi $a1_{ki}$ dengan menggunakan Persamaan (9).

- c. Menjumlahkan nilai aktivasi yang dikalikan dengan vektor *input* **W** yang sesuai dengan kelas data *training* menggunakan Persamaan (10).

- d. Menentukan kelas data *testing* sesuai dengan penjumlahan nilai aktivasi yang terbesar.

8. Melakukan Evaluasi Klasifikasi.

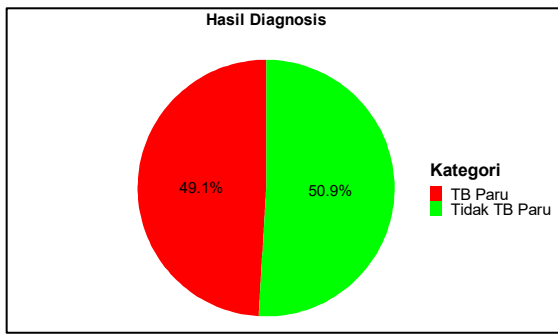
9. Mengambil Kesimpulan.

3. Hasil dan Pembahasan

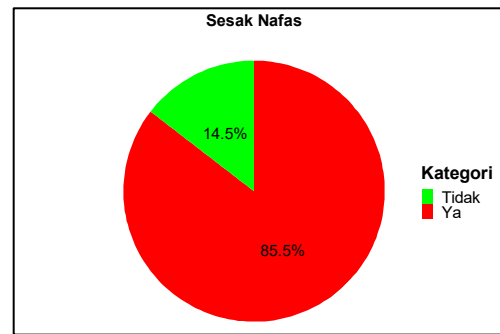
Data yang digunakan dalam penelitian ini adalah data pasien penyakit TB paru di Puskesmas Telaga Sari Kota Balikpapan Tahun 2023-2024. Jumlah data yang diperoleh pada penelitian ini sebanyak 110 orang.

3.1 Analisis Statistika Deskriptif

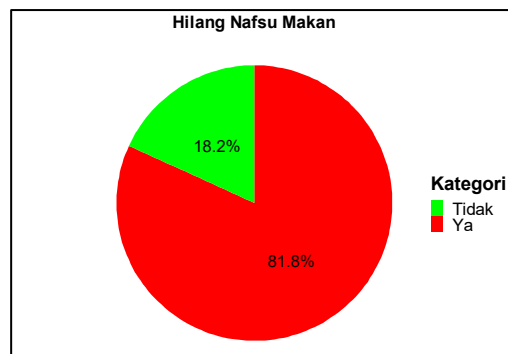
Menurut Martias (2021), analisis statistika deskriptif dilakukan untuk mengetahui gambaran umum pasien penyakit TB paru. Gambaran umum pada analisis statistika deskriptif meliputi persentase dalam bentuk diagram lingkaran dari variabel-variabel yang digunakan antara lain hasil diagnosis pasien penyakit TB paru, sesak napas, dan hilang nafsu makan. Sedangkan untuk variabel lainnya seperti usia, berat badan, tekanan darah sistolik, tekanan darah diastolik, lama batuk, dan lama demam akan dilakukan penyajian data secara deskriptif berupa tabel yang berisi informasi mengenai rata-rata, simpangan baku, nilai minimum, dan nilai maksimum dari variabel-variabel tersebut.



Gambar 1. Persentase Hasil Diagnosis Pasien



Gambar 2. Persentase Pasien Sesak Napas



Gambar 3. Persentase Pasien Hilang Nafsu Makan

Berdasarkan Gambar 1, dapat dilihat dari 110 pasien penyakit TB paru di Puskesmas Telaga Sari Kota Balikpapan Tahun 2023-2024, sebanyak 49,1% pasien didiagnosis TB paru, sedangkan 50,9% tidak TB paru. Hasil ini menunjukkan bahwa prevalensi pasien dengan hasil diagnosis positif TB paru masih cukup tinggi pada populasi puskesmas. Fenomena ini sejalan dengan data *World Health Organization* (2023) yang menyebutkan bahwa Indonesia menempati urutan kedua tertinggi kasus TB paru di dunia setelah India, dengan proporsi kasus baru sekitar 10% dari total global. Selanjutnya, pada Gambar 2, diketahui bahwa 85,5% pasien mengalami gejala sesak napas dan 14,5% tidak mengalami sesak napas. Gejala sesak napas merupakan salah satu indikator klinis umum pada pasien TB paru akibat kerusakan jaringan paru-paru dan penurunan kapasitas pernapasan (Lelewana, dkk., 2024). Tingginya proporsi pasien dengan gejala ini memperkuat pentingnya deteksi dini melalui pemeriksaan fisik dan laboratorium di tingkat puskesmas. Pada Gambar 3, terlihat bahwa 81,8% pasien mengalami kehilangan nafsu makan, sedangkan 18,2% pasien tidak mengalaminya. Hilangnya nafsu makan sering dikaitkan dengan respon imun tubuh terhadap infeksi bakteri *Mycobacterium tuberculosis* serta efek sistemik dari proses inflamasi (Rosamawati, dkk., 2023). Temuan ini mendukung hasil penelitian serupa oleh (Lelewana, dkk., 2024) yang menunjukkan bahwa gejala klinis seperti demam, sesak napas, dan kehilangan nafsu makan merupakan faktor signifikan dalam penegakan diagnosis TB paru.

Tabel 2. Statistika Deskriptif

Variabel	Rata-rata	Simpangan Baku	Minimum	Maksimum
Usia	45	20	9	84
Berat Badan	57	16	20	95
Tekanan Darah Sistolik	132	21	93	184
Tekanan Darah Diastolik	82	12	54	120
Lama Batuk	19	13	0	46
Lama Demam	7	6	0	26

Berdasarkan Tabel 2, dapat diketahui bahwa rata-rata usia pasien adalah 45 tahun, dengan simpangan baku 20 tahun. Hal ini menunjukkan bahwa sebagian besar pasien TB paru berada pada kelompok usia produktif. Temuan ini sejalan dengan penelitian Rahayu, dkk. (2023) yang menyebutkan bahwa kelompok usia produktif (15–55 tahun) memiliki risiko tinggi terpapar TB akibat aktivitas sosial dan mobilitas yang tinggi.

Rata-rata berat badan pasien sebesar 57 kg, menunjukkan adanya kemungkinan penurunan berat badan akibat infeksi kronis TB paru yang mengganggu metabolisme tubuh (Rosamawati, dkk., 2023). Nilai rata-rata tekanan darah sistolik dan diastolik masing-masing sebesar 132 mmHg dan 82 mmHg, yang masih berada pada rentang normal. Namun, variasi tekanan darah menunjukkan adanya perbedaan kondisi fisiologis antar pasien

yang dapat dipengaruhi oleh tingkat keparahan penyakit maupun faktor komorbiditas..

Rata-rata lama batuk selama 19 hari dan lama demam selama 7 hari mengindikasikan bahwa sebagian besar pasien telah mengalami gejala yang cukup lama sebelum melakukan pemeriksaan laboratorium. Kondisi ini menggambarkan keterlambatan dalam pencarian pengobatan yang juga ditemukan dalam penelitian Lelewana, dkk. (2024), di mana faktor keterlambatan pemeriksaan menjadi hambatan utama dalam pengendalian penularan TB paru.

Secara keseluruhan, hasil analisis deskriptif ini menunjukkan bahwa pola gejala dan karakteristik pasien di Puskesmas Telaga Sari sejalan dengan gambaran epidemiologis TB paru di wilayah perkotaan yang memiliki tingkat mobilitas tinggi. Temuan ini memperkuat relevansi penggunaan data lokal puskesmas sebagai dasar penerapan metode data mining seperti PNN untuk mendukung diagnosis dini secara lebih efisien.

3.2 Normalisasi Data

Normalisasi data dilakukan agar semua variabel berada dalam jangkauan yang sama sehingga proporsi pengaruh pada fungsi di dalam klasifikator dapat seimbang (Permana & Salisah, 2022). Pada penelitian ini data tidak berada dalam jangkauan yang sama, oleh karena itu perlu dilakukan normalisasi data. Variabel-variabel yang akan dilakukan normalisasi data pada penelitian ini adalah variabel yang memiliki skala rasio, yaitu usia, berat badan, tekanan darah sistolik, tekanan darah diastolik, lama batuk, dan lama demam. Hasil normalisasi data dapat dilihat pada Tabel 3.

Tabel 3. Data Hasil Normalisasi

No	X_1	X_2	X_3	X_4	X_5	X_6
1	0,692	0,503	1,140	-0,715	-0,157	-0,016
2	-0,295	0,818	0,691	-0,214	-0,237	-0,998
3	0,536	0,692	1,538	0,871	-0,883	-1,195
⋮	⋮	⋮	⋮	⋮	⋮	⋮
110	2,044	1	0,890	-1,800	-1,125	-0,409

Berdasarkan Tabel 4, dapat dilihat bahwa semua variabel sudah berada dalam jangkauan yang sama sehingga dapat digunakan dalam tahapan PNN selanjutnya. Data hasil normalisasi ini akan digunakan dalam tahapan pengacakan data.

3.3 Menentukan Nilai Smoothing Parameter

Klasifikasi PNN sangat bergantung pada nilai α yang digunakan. Pada penelitian ini nilai *smoothing parameter* yang digunakan adalah 0,1;0,2;0,3;0,4; 0,5;0,6;0,7;0,8;0,9; dan 1.

3.4 Melakukan Pengacakan Data

Pengacakan data pada penelitian ini menggunakan *software R* sebanyak satu kali di mana data yang diacak adalah data yang telah dinormalisasi. Pengacakan data bertujuan agar semua data memiliki kesempatan yang sama untuk menjadi data *training* dan data *testing* (Rahayu, dkk., 2023).

3.5 Membagi Data Training dan Data Testing

Setelah melalui proses pengacakan data, selanjutnya dilakukan pembagian data *training* dan data *testing*. Peneliti menggunakan 1 proporsi dalam pembagian data *training* dan data *testing* yaitu 80:20. Dari 110 data dengan proporsi 80:20, 88 data pertama akan digunakan menjadi data *training* dan 22 data lainnya akan menjadi data *testing*.

3.6 Klasifikasi Menggunakan Metode PNN

Berikut adalah langkah-langkah klasifikasi metode PNN menggunakan data *training* dan data *testing* dengan proporsi 80:20.

1. Perhitungan bobot bias lapisan *input* dilakukan menggunakan Persamaan (4). Hasil perhitungan nilai bobot lapisan *input* (b) dapat dilihat pada Tabel 4.

Tabel 4. Nilai Bobot Bias Lapisan *Input*

α	b
0,1	2,633
0,2	1,862
0,3	1,520

0,4	1,316
0,5	1,177
0,6	1,075
0,7	0,995
0,8	0,931
0,9	0,878
1	0,833

Hasil perhitungan bobot bias lapisan *input* pada Tabel 5 akan digunakan dalam perhitungan nilai aktivasi.

- Perhitungan jarak dilakukan dengan menggunakan Persamaan (5), (6), (7), dan (8). Adapun hasil perhitungan jarak Gower antar data *training* dapat dilihat pada Tabel 5.

Tabel 5. Matriks Jarak Gower antar Data *Training*

$\begin{smallmatrix} k \\ i \end{smallmatrix}$	1	2	3	4	...	88
1	0	0,262	0,262	0,259	...	0,230
2	0,262	0	0,458	0,102	...	0,170
3	0,262	0,458	0	0,508	...	0,442
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
88	0,230	0,170	0,442	0,226	...	0

Perhitungan jarak yang sama juga dilakukan antara data *testing* dengan data *training* yang hasilnya dapat dilihat pada Tabel 6.

Tabel 6. Matriks Jarak Gower antara Data *Testing* dengan Data *Training*

$\begin{smallmatrix} k \\ i \end{smallmatrix}$	1	2	3	4	...	88
1	0,138	0,381	0,299	0,357	...	0,361
2	0,205	0,138	0,360	0,165	...	0,152
3	0,345	0,290	0,405	0,346	...	0,127
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
22	0,341	0,243	0,488	0,304	...	0,120

Perhitungan jarak Gower selanjutnya akan digunakan dalam menghitung nilai aktivasi.

- Menghitung nilai aktivasi $a1_{ki}$ dengan menggunakan Persamaan (9). Perhitungan nilai aktivasi dilakukan dengan semua nilai *smoothing parameter* dan berlaku untuk nilai aktivasi antar data *training* maupun nilai aktivasi antara data *testing* dengan data *training*. Adapun sebagai contoh Peneliti menggunakan $\alpha = 0,1$ yang hasil perhitungannya dapat dilihat pada Tabel 7 dan Tabel 8.

Tabel 7. Matriks Nilai Aktivasi antar Data *Training*

$\begin{smallmatrix} k \\ i \end{smallmatrix}$	1	2	3	4	...	88
1	1	0,621	0,621	0,628	...	0,693
2	0,621	1	0,234	0,930	...	0,819
3	0,621	0,234	1	0,167	...	0,258
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
88	0,693	0,819	0,258	0,702	...	1

Hasil perhitungan nilai aktivasi antar data *training* akan digunakan dalam tahapan selanjutnya untuk mendapatkan hasil klasifikasi antar *training*.

Tabel 8. Matriks Nilai Aktivasi antara Data *Testing* dengan Data *Training*

$\begin{smallmatrix} k \\ i \end{smallmatrix}$	1	2	3	4	...	88
1	1	0,621	0,621	0,628	...	0,693
2	0,621	1	0,234	0,930	...	0,819

3	0,439	0,559	0,320	0,436	...	0,894
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
22	0,693	0,819	0,258	0,702	...	1

Setelah didapat nilai aktivasi antara data *testing* dengan data *training*, langkah selanjutnya adalah menjumlahkan nilai aktivasi yang dikalikan dengan vektor *input* **W**.

- Menjumlahkan nilai aktivasi yang dikalikan dengan vektor *input* **W** yang sesuai dengan kelas data menggunakan Persamaan (10). Kemudian menentukan kelas dengan cara membandingkan nilai kolom TB paru dengan tidak TB paru. Jika nilai kolom TB paru bernilai maksimum maka hasil prediksi adalah kelas data 1 dan jika sebaliknya maka nilai hasil prediksi adalah 0. Adapun hasil klasifikasi antar data *training* dengan $\alpha = 0,1$ dapat dilihat pada Tabel 9.

Tabel 9. Hasil Klasifikasi antar Data *Training*

No	TB Paru	Tidak TB Paru	Prediksi	Hasil Diagnosis
1	25,239	30,179	0	0
2	32,975	27,964	1	1
3	11,407	22,048	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
88	33,909	31,069	1	1

Hal yang sama juga dilakukan antara data *testing* dengan data *training* yang hasilnya dapat dilihat pada Tabel 10.

Tabel 10. Hasil Klasifikasi Data *Testing* dengan Data *Training*

No	TB Paru	Tidak TB Paru	Prediksi	Hasil Diagnosis
1	17,071	23,999	0	0
2	32,987	35,469	0	1
3	26,413	29,375	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
22	28,120	27,262	1	1

Berdasarkan Tabel 9 dan Tabel 10, kolom hasil prediksi diperoleh dengan membandingkan nilai kolom TB paru dengan tidak TB paru. Jika nilai kolom TB paru bernilai maksimum maka hasil prediksi adalah kelas data 1 dan jika sebaliknya maka nilai hasil prediksi adalah 0.

3.7 Evaluasi Klasifikasi

Setelah diperoleh hasil prediksi klasifikasi, selanjutnya dilakukan perhitungan akurasi untuk menggambarkan sejauh mana metode PNN dapat melakukan pengklasifikasian hasil diagnosis pasien penyakit TB paru. Berikut adalah hasil akurasi dengan menggunakan 10 nilai *smoothing parameter* (α).

Tabel 12. Nilai Akurasi dengan 10 Nilai α

Proporsi	α	Data <i>Training</i> (%)	Data <i>Testing</i> (%)
80:20	0,1	82,95	95,45
	0,2	82,09	90,90
	0,3	82,95	77,27
	0,4	76,13	77,27
	0,5	72,72	68,18
	0,6	64,77	68,18
	0,7	59,09	59,09
	0,8	55,68	54,54
	0,9	53,40	54,54
	1	53,40	54,54

Berdasarkan Tabel 12, dapat diketahui terdapat 10 nilai akurasi untuk data *training* dan data *testing*. Berdasarkan hasil akurasi dari tiap proporsi dipilih dengan cara melihat proporsi yang mendapatkan hasil akurasi tertinggi, tahap pertama yang dilakukan yaitu melihat nilai akurasi pada data *training*. Diketahui nilai akurasi pada data *training* tertinggi adalah 82,95% dengan $\alpha = 0,1$ dan $\alpha = 0,3$. Setelah dipilih nilai akurasi tertinggi pada data *training*, dilanjutkan dengan melihat nilai akurasi data *testing* tertinggi dari data *training*

yang telah dipilih sebelumnya. Diketahui nilai akurasi paling tertinggi adalah 95,45% dengan $\alpha = 0,1$.

4. Kesimpulan

Berdasarkan hasil dan pembahasan, maka didapat kesimpulan bahwa nilai *smoothing parameter* (α) terbaik hasil klasifikasi menggunakan metode PNN pada data pasien penyakit TB paru di Puskesmas Telaga Sari Kota Balikpapan adalah $\alpha = 0,1$ dengan nilai akurasi terbaik sebesar 82,95% untuk data *training* dan 95,45% untuk data *testing*.

5. Referensi

- Azizah, P. D. I. (2021). Penerapan Probabilistic Neural Network pada Klasifikasi Berat Bayi Baru Lahir. *Jurnal Riset Statistika*, 1(2), 152-159.
- Cahyanti, F. L. D., Sarasati, F., Astuti, W., dan Firasari, E. (2023). Klasifikasi Data Mining dengan Algoritma Machine Learning untuk Prediksi Penyakit Liver. *Technologia*, 14(2), 134-139.
- Dinas Kesehatan Provinsi Kalimantan Timur. (2023). *Profil Kesehatan Tahun 2022*. Kalimantan Timur: Dinas Kesehatan Provinsi Kalimantan Timur.
- Kementerian Kesehatan Republik Indonesia. (2024). *Profil Kesehatan Indonesia 2023*. Jakarta: Kementerian Kesehatan RI.
- Lelewana, A. Y. P., Suarnianti, dan Restika, I. (2024). Determinan Perilaku Pencarian Pengobatan Pada Penderita TB Paru Di Puskesmas Tamalanrea. *Jurnal Ilmiah Mahasiswa & Penelitian Keperawatan*, 4(2), 142-151.
- Lenaini, I. (2021). Teknik Pengambilan Sampel Purposive dan Snowball Sampling. *Jurnal Kajian, Penelitian dan Pengembangan Pendidikan Sejarah*, 6(1), 33-39.
- Martias, L. D. (2021). Statistika Deskriptif Sebagai Kumpulan Informasi. *Jurnal Ilmu Perpustakaan dan Informasi*, 16(1), 40-59.
- Permana, I. dan Salisah, F. N. (2022). Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation. *Indonesian Journal of Informatic Research and Software Engineering*, 2(1), 67-72.
- Primantoro, S., Goejantoro, R., dan Prangga, S. (2024). Clustering Titik Panas Bumi Pada Potensi Kebakaran Hutan Menggunakan K-Affinity Propagation. *Jurnal Ekspansional*, 15(2), 73-79.
- Rahayu, N., Wasono, Amijaya, F. D. T. (2023). Klasifikasi Batubara Berdasarkan Jenis Kalori dengan Menggunakan Perbandingan Jarak Euclid, Jarak Manhattan, dan Jarak Minkowski Pada Algoritma Modified K-Nearest Neighbor (Studi Kasus: PT. XYZ Samarinda). *Jurnal Ilmiah Matematika*, 2(1), 1-10.
- Rosamawati, Sartika, dan Hasan, C. (2023). Faktor Risiko Kejadian Tuberkulosis Paru di Wilayah Kerja Puskesmas Kaluku Bodoa Kota Makassar. *Window of Public Health Journal*, 4(6), 1028-1040.
- Sawo, M. K., Rogi, O. H. A., dan Lakat, R. S. M. (2021). Analisis Pengembangan Kawasan Permukiman Berdasarkan Kemampuan Lahan Di Distrik Muara Tami. *Jurnal Spasial*, 8(3), 311-325.
- Srirahayu, A. dan Pribadie, L. S. (2023). Review Paper Data Mining Klasifikasi Data Mining. *Jurnal Ilmiah Informatika Global*, 14(1), 7-12.
- Tarigan, P. M. S., Hardinata, J. T., Qurniawan, H., Safii, M., dan Winanjaya, R. (2022). Implementasi Data Mining Menggunakan Algoritma Apriori dalam Menentukan Persediaan Barang (Studi Kasus: Toko Sinar Harahap). *Jurnal UMJ*, 12(2), 51-61.
- Wenando, F. A., Fatma, Y., Ulfa, A., Salma, dan Taurin, J. S. (2023). Aplikasi dan Kerentanan Algoritma Probabilistic Neural Network (PNN): Systematic Literature Review. *Jurnal Computer Science and Information Technology*, 4(2), 491-499.