

KLASIFIKASI PENYAKIT KARDIOVASKULAR MENGGUNAKAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (SMOTE) DAN *BACKPROPAGATION NEURAL NETWORK* (BPNN)

Muhammad Yafie Anwary Rahman
(190441100052)

Dosen Pembimbing

1. Dr. Wahyudi Setiawan, S.Kom., M.Kom.
2. Muhammad Yusuf, S.T., M.MT., Ph.D.

ABSTRAK

Salah satu organ yang berperan penting dalam kelangsungan hidup manusia adalah organ jantung. Penyakit yang berkaitan dengan jantung salah satunya adalah penyakit kardiovaskular. Meninjau data yang tersedia dari Riset Kesehatan Dasar (RISKESDAS) pada tahun 2018, di Indonesia terdapat kasus penyakit kardiovaskular hingga mencapai 2.784.064 penduduk menderita penyakit jantung dan mengalami peningkatan setiap tahunnya. Penelitian ini bertujuan untuk membuat suatu model klasifikasi penyakit kardiovaskular yang optimal dengan menggunakan algoritma *Backpropagation Neural Network* (BPNN) . Selain itu penelitian ini juga menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) untuk menyeimbangkan data per kelas. Model BPNN menggunakan *learning rate* 0,01 , *epoch* 1.000, *batch size* 64, *node input layer* sebanyak 14, *node* pada *hidden layer* sebanyak 3 serta *node* pada *output layer* sebanyak 1 *node*. Data yang digunakan berjumlah 425.196 data sampel yang terdiri dari 14 fitur dan 2 kelas. Data dilakukan *preprocessing* melalui tahap seleksi fitur, *cleaning data*, transformasi data, normalisasi data lalu dilakukan *oversampling* menggunakan teknik SMOTE selanjutnya dilakukan klasifikasi. Pada proses *training* menggunakan *k-fold cross validation* dengan nilai $k = 10$. Penelitian ini menggunakan 5 skenario uji coba yang dapat disimpulkan bahwa penggunaan algoritma *Backpropagation Neural Network* dengan SMOTE menghasilkan akurasi klasifikasi tertinggi yaitu sebesar 79,99%, presisi sebesar 79,2%, nilai *recall* sebesar 83,3%, dan nilai AUC sebesar 79,99%.

Kata Kunci : Klasifikasi, Kardiovaskular, *Backpropagation Neural Network*, RISKESDAS, SMOTE

ABSTRACT

One of the organs that plays a crucial role in human survival is the heart. One of the diseases related to the heart is cardiovascular disease. Based on the data from the Basic Health Research (RISKESDAS) in 2018, there were cases of cardiovascular disease in Indonesia, affecting up to 2,784,064 people suffering from heart disease, and the numbers have been increasing annually. This study aims to create an optimal classification model for cardiovascular disease using the Backpropagation Neural Network (BPNN) algorithm. Additionally, the study employs the Synthetic Minority Oversampling Technique (SMOTE) to balance the data per class. The BPNN model is configured with a learning rate of 0.01, 1,000 epochs, a batch size of 64, 14 nodes in the input layer, 3 nodes in the hidden layer, and 1 node in the output layer. The dataset used in the study consists of 425,196 samples with 14 features and 2 classes. The data undergoes preprocessing, including feature selection, data cleaning, data transformation, and data normalization, followed by oversampling using the SMOTE technique before performing the classification. The training process utilizes k-fold cross-validation with a value of $k = 10$. The study explores 5 test scenarios, and it can be concluded that the utilization of the Backpropagation Neural Network algorithm with SMOTE yields the highest classification accuracy, which is 79.99%, precision of 79.2%, recall value of 83.3%, and an AUC value of 79.99%.

Keyword : *Classification, Cardiovascular, Backpropagation Neural Network, RISKESDAS, SMOTE.*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN	iii
MOTTO	iv
HALAMAN PERSEMBAHAN	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	3
1.2.1 Permasalahan	4
1.2.2 Metode Penelitian	4
1.2.3 Pertanyaan Penelitian.....	4
1.3 Tujuan dan Manfaat	4
1.3.1 Tujuan	4
1.3.2 Manfaat	4
1.4 Batasan Masalah.....	5
1.5 Sistematika Penulisan Skripsi	6
BAB II KAJIAN PUSTAKA.....	7
2.1 Penyakit Kardiovaskular	7
2.1.1 Standar Operasional Prosedur (SOP) Penyakit Kardiovaskular	7
2.2 Klasifikasi.....	8
2.3 <i>Artificial Intelligence</i>	8
2.4 <i>Supervised Learning</i>	9
2.5 <i>Preprocessing</i>	10
2.5.1 <i>Cleaning Data</i>	10

2.5.2	Transformasi Data.....	11
2.5.3	Normalisasi Data.....	11
2.5.4	<i>Oversampling</i>	12
2.6	Fungsi Aktivasi	13
2.7	<i>Neural Network</i>	13
2.8	<i>Backpropagation Neural Network</i>	15
2.9	Evaluasi dan Validasi Hasil.....	18
2.9.1	<i>Cross Validation</i>	18
2.9.2	<i>Confusion Matrix</i>	18
2.9.3	<i>Receiver Operating Characteristic (ROC)</i>	19
2.10	Penelitian Terdahulu	21
BAB III METODE PENELITIAN		27
3.1	Jenis Penelitian.....	27
3.2	Alur Penelitian.....	27
3.2.1	Studi Pustaka.....	27
3.2.2	Analisa dan Perancangan Model.....	27
3.2.3	Implementasi.....	28
3.2.4	Skenario Uji Coba.....	28
3.2.5	Hasil dan Kesimpulan	29
3.2.6	Dokumentasi	29
3.3	Alur Sistem.....	29
3.2	Pengumpulan Data	31
3.3	<i>Preprocessing</i>	32
3.3.1	Seleksi Fitur	32
3.3.2	<i>Cleaning Data</i>	34
3.3.3	Transformasi Data.....	35
3.3.4	Normalisasi Data.....	36
3.3.5	<i>Oversampling</i>	37
3.4	<i>Backpropagation Neural Network</i>	40
3.4.1	Inisialisasi Bobot dan Bias.....	41
3.4.2	<i>Feed Forward</i>	42
3.4.3	Perhitungan <i>Error</i>	44
3.4.4	<i>Backward Pass</i>	45

3.4.5	<i>Update Bobot dan Bias</i>	46
3.4.6	<i>Iterasi</i>	48
3.4	Validasi dengan <i>k-fold</i>	48
3.5	Evaluasi dengan <i>Confusion Matrix</i>	50
BAB IV HASIL DAN PEMBAHASAN		52
4.1	Spesifikasi Perangkat	52
4.2	Pengumpulan Data	53
4.2.1	Mengunduh Data.....	53
4.2.2	Analisis Data.....	53
4.3	<i>Preprocessing</i>	55
4.3.1	Seleksi Fitur	56
4.3.2	<i>Cleaning Data</i>	57
4.3.3	Transformasi Data.....	57
4.3.4	Normalisasi Data.....	60
4.3.5	<i>Oversampling</i>	61
4.4	<i>Backpropagation Neural Network</i> dan <i>10-Fold Validation</i>	63
4.5	Evaluasi	66
4.6	Hasil Uji Coba.....	71
4.6.1	Skenario 1	71
4.6.2	Skenario 2	74
4.6.3	Skenario 3	77
4.6.4	Skenario 4	80
4.6.5	Skenario 5	82
4.6.6	Analisa Hasil.....	84
BAB V PENUTUP.....		86
5.1	Kesimpulan	86
5.2	Saran	86
DAFTAR PUSTAKA.....		88
LAMPIRAN -LAMPIRAN		93

DAFTAR GAMBAR

Gambar 2. 1 Diagram <i>Artificial Intelligence</i>	9
Gambar 2. 2 Alur Kerja <i>Supervised Learning</i>	10
Gambar 2. 3 Fungsi Aktivasi Sigmoid Biner.....	13
Gambar 2. 4 Arsitektur Jaringan Saraf Tiruan	14
Gambar 2. 5 Arsitektur <i>Backpropagation Neural Network</i>	15
Gambar 2. 6 Komparasi Klasifikasi dengan Kurva ROC	20
Gambar 3. 1 Alur Penelitian	27
Gambar 3. 2 Alur Sistem	30
Gambar 3. 3 Alur Kerja SMOTE.....	37
Gambar 3. 4 Alur Kerja <i>Backpropagation Neural Network</i>	41
Gambar 3. 5 Skema 10-Fold Validation	48
Gambar 3. 6 <i>Training Set 10-Fold Validation</i>	49
Gambar 4. 1 Kode Program <i>Load Dataset</i>	53
Gambar 4. 2 Jumlah Baris dan Kolom Data	54
Gambar 4. 3 Jumlah Data Per Kelas	54
Gambar 4. 4 Visualisasi <i>Bar Chart</i> Data Per Kelas.....	54
Gambar 4. 5 Informasi Tipe Data	55
Gambar 4. 6 Kode Program <i>Cleaning Data</i>	57
Gambar 4. 7 <i>Split</i> Data Fitur dan Target.....	58
Gambar 4. 8 Kode Program <i>One Hot Encoding “Gender”</i>	58
Gambar 4. 9 Kode Program <i>One Hot Encoding “Race”</i>	59
Gambar 4. 10 Kode Program Normalisasi Data	61
Gambar 4. 11 Kode Program SMOTE.....	62
Gambar 4. 12 Visualisasi Data Setelah SMOTE	63
Gambar 4. 13 <i>Split</i> Data Fitur dan Target #2.....	63
Gambar 4. 14 Kode Program <i>Backpropagation Neural Network & 10 Fold Validation</i>	64
Gambar 4. 15 Kode Program Evaluasi Matriks	66
Gambar 4. 16 Perhitungan Matriks Evaluasi	67
Gambar 4. 17 Hasil Evaluasi Matriks Masing-Masing <i>Fold</i>	68
Gambar 4. 18 <i>Chart</i> Perbandingan Evaluasi pada Masing-Masing <i>Fold</i>	69
Gambar 4. 19 Visualisasi <i>Confusion Matrix</i>	69

Gambar 4. 20 Hasil Evaluasi Matriks Terbaik.....	70
Gambar 4. 21 Visualisasi Kurva ROC.....	70
Gambar 4. 22 <i>Chart</i> Perbandingan Hasil Evaluasi Skenario 1a.....	72
Gambar 4. 23 <i>Chart</i> Perbandingan Hasil Evaluasi Skenario 1b.....	73
Gambar 4. 24 <i>Chart</i> Hasil Perbandingan Evaluasi Skenario 2a.....	75
Gambar 4. 25 <i>Chart</i> Perbandingan Hasil Evaluasi Skenario 2b.....	76
Gambar 4. 26 <i>Chart</i> Perbandingan Hasil Evaluasi Skenario 3a.....	78
Gambar 4. 27 <i>Chart</i> Perbandingan Hasil Evaluasi Skenario 3b.....	79
Gambar 4. 28 <i>Chart</i> Perbandingan Hasil Evaluasi Skenario 4a.....	81
Gambar 4. 29 <i>Chart</i> Perbandingan Hasil Evaluasi Skenario 4b.....	82
Gambar 4. 30 Hasil Perbandingan Evaluasi Matriks dari Keseluruhan Uji Coba.....	85

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	18
Tabel 2. 2 Tabel Keakuratan AUC	20
Tabel 2. 3 Penelitian Terdahulu	23
Tabel 3. 1 Skenario Uji Coba.....	29
Tabel 3. 2 Informasi Detail <i>Dataset</i>	31
Tabel 3. 3 Distribusi <i>Dataset</i>	32
Tabel 3. 4 Sebelum Seleksi Fitur	33
Tabel 3. 5 Setelah Seleksi Fitur	33
Tabel 3. 6 Sebelum <i>Cleaning Data</i>	34
Tabel 3. 7 Sesudah <i>Cleaning Data</i>	34
Tabel 3. 8 Sebelum Transformasi Data.....	35
Tabel 3. 9 Sesudah Transformasi Data	36
Tabel 3. 10 Sebelum Normalisasi Data.....	36
Tabel 3. 11 Setelah Normalisasi Data.....	37
Tabel 3. 12 Contoh <i>Dataset</i> Sebelum SMOTE.....	38
Tabel 3. 13 Contoh <i>Dataset</i> Setelah Dilakukan SMOTE	40
Tabel 3. 14 Inisialisasi Awal.....	41
Tabel 3. 15 Bobot Acak <i>Input Layer</i> ke <i>Hidden Layer</i>	41
Tabel 3. 16 Bias Acak <i>Input Layer</i> ke <i>Hidden Layer</i>	42
Tabel 3. 17 Bobot Acak <i>Hidden Layer</i> ke <i>Output Layer</i>	42
Tabel 3. 18 Bias Acak dari <i>Hidden Layer</i> ke <i>Output Layer</i>	42
Tabel 3. 19 <i>Input Data</i>	43
Tabel 3. 20 Hasil <i>Feed Forward</i>	44
Tabel 3. 21 Nilai <i>Error Output Layer</i>	44
Tabel 3. 22 Nilai <i>Input Kesalahan Hidden Layer</i>	45
Tabel 3. 23 Nilai <i>Kesalahan Hidden Layer</i>	46
Tabel 3. 24 Selisih Bobot <i>Hidden</i> ke <i>Output Layer</i>	46
Tabel 3. 25 Bobot <i>Hidden</i> dan Bias Baru	47
Tabel 3. 26 Bobot Baru.....	47
Tabel 3. 27 Tabel Hasil Prediksi.....	50
Tabel 3. 28 Tabel <i>Confusion Matrix</i>	50

Tabel 4. 1 Spesifikasi <i>Hardware</i>	52
Tabel 4. 2 Spesifikasi <i>Software</i>	52
Tabel 4. 3 Daftar <i>Library</i>	52
Tabel 4. 4 Sebelum Seleksi Fitur	56
Tabel 4. 5 Setelah Seleksi Fitur	56
Tabel 4. 6 Hasil <i>10-Fold Cross Validation</i> Skenario 1a.....	71
Tabel 4. 7 Hasil <i>Fold</i> Terbaik Skenario 1a.....	72
Tabel 4. 8 Hasil <i>10-Fold Cross Validation</i> Skenario 1b.....	73
Tabel 4. 9 Hasil <i>Fold</i> Terbaik Skenario 1b.....	74
Tabel 4. 10 Hasil <i>10-Fold Cross Validation</i> Skenario 2a.....	74
Tabel 4. 11 Hasil <i>Fold</i> Terbaik Skenario 2a.....	75
Tabel 4. 12 Hasil Perbandingan <i>10-Fold Cross Validation</i> Skenario 2b.....	75
Tabel 4. 13 Hasil <i>Fold</i> Terbaik Skenario 2b.....	76
Tabel 4. 14 Hasil Perbandingan <i>10-Fold Cross Validation</i> Skenario 3a.....	77
Tabel 4. 15 Hasil <i>Fold</i> Terbaik Skenario 3a.....	78
Tabel 4. 16 Hasil Perbandingan <i>10-Fold Cross Validation</i> Skenario 3b.....	78
Tabel 4. 17 Hasil <i>Fold</i> Terbaik Skenario 3b.....	79
Tabel 4. 18 Hasil Perbandingan <i>10-Fold Cross Validation</i> Skenario 4a.....	80
Tabel 4. 19 Hasil <i>Fold</i> Terbaik Skenario 4a.....	81
Tabel 4. 20 Hasil Perbandingan <i>10-Fold Cross Validation</i> Skenario 4b.....	81
Tabel 4. 21 Hasil <i>Fold</i> Terbaik Skenario 4b.....	82
Tabel 4. 22 Hasil Evaluasi Matriks Skenario 1a – 4b.....	83
Tabel 4. 23 Hasil Uji Skenario 5a.....	83
Tabel 4. 24 Hasil Uji Skenario 5b.....	83
Tabel 4. 25 Hasil Uji Coba Skenario 1 - 5.....	84