

ABSTRAK

Hipertensi merupakan masalah kesehatan global dengan prevalensi tinggi yang berpotensi memicu komplikasi serius, sehingga memerlukan mekanisme deteksi dini yang cepat dan presisi. Pemanfaatan *machine learning* menawarkan solusi analitik untuk mendukung diagnosis penyakit berdasarkan rekaman data medis secara otomatis. Penelitian ini bertujuan untuk mengklasifikasikan penyakit hipertensi menggunakan algoritma *Decision Tree C4.5* berbasis kriteria *entropy* yang diterapkan pada dataset pasien dari UPT Puskesmas Modopuro, Kabupaten Mojokerto. Proses pemodelan melibatkan 1.537 data rekam medis dengan atribut utama yang dievaluasi meliputi tekanan darah Sistole, Diastole, dan Umur. Untuk mengevaluasi keandalan sistem, pengujian dilakukan melalui variasi skenario rasio pembagian data (70:30, 80:20, 90:10) serta penyesuaian tingkat kedalaman pohon (*max_depth* 3, 5, dan 7). Hasil penelitian membuktikan bahwa model mencapai performa paling optimal pada konfigurasi rasio data 90:10 dengan *max_depth* 3, yang menghasilkan tingkat akurasi tertinggi sebesar 92,98%. Berdasarkan interpretasi struktur pohon keputusan yang terbentuk, atribut Sistole teridentifikasi sebagai indikator klinis paling dominan (*root node*) dalam membedakan diagnosis kelas hipertensi dan normal. Secara keseluruhan, penerapan arsitektur algoritma C4.5 ini terbukti sangat efisien, akurat, serta mampu merepresentasikan aturan logis keputusan yang transparan, menjadikannya sangat potensial untuk diimplementasikan sebagai sistem pendukung keputusan medis.

Kata kunci : Hipertensi, Klasifikasi, *Decision Tree*

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
KATA PENGANTAR	iv
ABSTRAK	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR KODE.....	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.2.1 Permasalahan.....	3
1.2.2 Metode Usulan	3
1.2.3 Pertanyaan Penelitian	3
1.3 Tujuan	3
1.4 Manfaat Penelitian	4
1.4 Batasan Masalah.....	4
1.5 Sistematika Penelitian	4
BAB II KAJIAN PUSTAKA	6
2.1 Klasifikasi	6
2.2 Hipertensi	6
2.3 Diskretisasi.....	8
2.4 Label encoding	9
2.5 <i>Decision Tree</i>	9
2.5.1 Algoritma C4.5.....	10
2.6 <i>Confusion matrix</i>	12
2.7 Penelitian Terkait	13
BAB III METODE USULAN.....	17
3.1 Perancangan Sistem	17

3.2 Data	18
3.3 <i>PreProcessing</i>	20
3.3.1 Cleaning data.....	21
3.3.2 Transformasi Data.....	21
3.4 Pembagian Data	27
3.4.1 Data <i>Training</i>	28
3.4.2 Data <i>Testing</i>	28
3.5 Klasifikasi <i>Decision Tree C4.5</i>	28
3.6 Evaluasi Model.....	35
3.7 Skenario Pengujian.....	35
BAB IV HASIL DAN PEMBAHASAN	37
4.2 Lingkungan Uji Coba	37
4.2 Persiapan Data.....	37
4.2.1 Memuat Dataset	37
4.2.2 Data Cleaning.....	38
4.2.2.1 Penanganan Data Duplikat.....	39
4.2.3 Transformasi Data.....	40
4.2.3.1 Diskretisasi Data	40
4.2.3.2 <i>Label encoding</i>	42
4.2.3.3 Pembagian Data	43
4.3 Permodelan.....	44
4.4 Evaluasi	45
4.5 Implementasi User Interface	46
4.6 Hasil Uji Skenario	47
4.6.1 Hasil Uji Skenario 1 (Rasio Pembagian Data 70:30).....	48
4.6.2 Hasil Uji Skenario 2 (Rasio Pembagian Data 80:20).....	49
4.5.3 Hasil Uji Skenario 3 (Rasio Pembagian Data 90:10).....	51
4.5.4 Hasil Uji Coba Seluruh Skenario	52
4.7 Analisis Hasil Uji Skenario	54
4.7.1 Analisis Hasil Uji Skenario 1 (Rasio Pembagian Data 70:30).....	54
4.7.2 Analisis Hasil Uji Skenario 2 (Rasio Pembagian Data 80:20).....	56
4.7.3 Analisis Hasil Uji Skenario 3 (Rasio Pembagian Data 90:10).....	58
4.7.4 Analisis Semua Skenario Uji Coba	60
BAB V PENUTUP.....	64

5.1 Kesimpulan	64
5.2 Saran.....	64
REFERENSI	65
LAMPIRAN 1: <i>Dataset</i> Hipertensi.....	68

DAFTAR GAMBAR

Gambar 3. 1 Block Diagram Perancangan Sistem.....	17
Gambar 3. 2 Pohon Keputusan	34
Gambar 4. 1 10 Data Teratas	38
Gambar 4. 2 Hasil Cek Missing Value dan Data Duplikat	39
Gambar 4. 3 Contoh Data Duplikat	39
Gambar 4. 4 Hasil Data Setelah Hapus Duplikat.....	40
Gambar 4. 5 Hasil Diskretisasi Data.....	41
Gambar 4. 6 Hasil <i>Label encoding</i>	42
Gambar 4. 7 Tampilan Awal.....	46
Gambar 4. 8 Halaman Diagnosa	47
Gambar 4. 9 Halaman Hasil Diagnosa.....	47
Gambar 4. 10 <i>Confusion matrix</i> 70:30.....	49
Gambar 4. 11 <i>Confusion matrix</i> 80:20.....	50
Gambar 4. 12 <i>Confusion matrix</i> 90:10.....	52
Gambar 4. 13 Perbandingan Akurasi	54
Gambar 4. 14 Pohon keputusan	62

DAFTAR TABEL

Tabel 3. 1 Fitur-Fitur Dataset.....	18
Tabel 3. 2 Contoh Data	20
Tabel 3. 3 Pembagian Kategori Umur	22
Tabel 3. 4 Contoh Diskretisasi Umur	22
Tabel 3. 5 Pembagian Kategori Tinggi Badan.....	22
Tabel 3. 6 Contoh Diskretisasi Tinggi Badan.....	23
Tabel 3. 7 Pembagian Kategori Berat Badan.....	23
Tabel 3. 8 Contoh Diskretisasi Berat Badan	23
Tabel 3. 9 Pembagian Kategori IMT	23
Tabel 3. 10 Contoh Diskretisasi IMT	24
Tabel 3. 11 Pembagian Kategori Sistol.....	24
Tabel 3. 12 Contoh Diskretisasi Sistol.....	24
Tabel 3. 13 Pembagian Kategori Diastole	24
Tabel 3. 14 Contoh Diskretisasi Diastole	25
Tabel 3. 15 Pembagian Kategori Nafas	25
Tabel 3. 16 Contoh Diskretisasi Nafas	25
Tabel 3. 17 Pembagian Kategori Nadi.....	25
Tabel 3. 18 Contoh Diskretisasi Nadi.....	26
Tabel 3. 19 Contoh Diskretisasi Semua Atribut	26
Tabel 3. 20 Contoh <i>Label encoding</i>	27
Tabel 3. 21 Contoh Data Yang Akan Dihitung.....	28
Tabel 3. 22 Contoh Data <i>Training</i>	29
Tabel 3. 23 Hasil Perhitungan Masing-Masing Atribut.....	34
Tabel 3. 24 Contoh Data Yang Akan diuji	35
Tabel 4. 1 Lingkungan Uji Coba.....	37
Tabel 4. 2 Hasil Uji Skenario 1.....	48
Tabel 4. 3 Hasil Uji Skenario 2.....	50
Tabel 4. 4 Hasil Uji Skenario 3.....	51
Tabel 4. 5 Hasil Uji Seluruh Skenario	53

DAFTAR KODE

Kode 4. 1 Load Data	37
Kode 4. 2 Cek Missing Value dan Data Duplikat.....	38
Kode 4. 3 Hapus Data Duplikat	39
Kode 4. 4 Diskretisasi Data	41
Kode 4. 5 <i>Label encoding</i>	42
Kode 4. 6 Pembagian Data.....	43
Kode 4. 7 Permodelan.....	44
Kode 4. 8 <i>Confusion matrix</i>	45