

ABSTRAK

Penyakit stroke menduduki peringkat pertama penyebab kecacatan dan kematian peringkat tiga di dunia sesudah penyakit kanker dan jantung. Masyarakat Indonesia banyak yang belum mengetahui dan menyadari tanda-tanda atau gejala stroke sedini mungkin karena gejalanya beragam. Apalagi masyarakat enggan memeriksakan diri ke rumah sakit untuk berkonsultasi terkait gejala yang dialami. Oleh karena itu, untuk memudahkan dalam mendiagnosa pasien penyakit stroke dilakukan klasifikasi dengan menggunakan metode *K-Nearest Neighbor* (K-NN). Atribut yang digunakan pada skripsi ini adalah *id*, *gender*, *age*, *hypertension*, *heart_disease*, *ever_married*, *work_type*, *residence_type*, *avg_glucose_level*, *bmi*, *smoking_status*, *stroke*. Dalam mengetahui suatu tingkatan akurasi nilai dari data yang dihasilkan di lanjutkan dengan proses perhitungan *recall*, akurasi, dan presisi, dan dengan persamaan. Pengujian dilakukan dengan menggunakan nilai tetangga $k=3$, $k=5$, $k=7$, $k=9$, dan $k=11$ untuk setiap pembagian data *testing* dan data *training*, penelitian mendapatkan nilai evaluasi terbaik dengan nilai presisi = 95.59%, *recall* = 100.00%, *f-1 score* = 97.74%, dan akurasi = 95.59% dengan pembagian data sebanyak $k=5$ pada *k-folding* 5.

Kata Kunci: Penyakit Stroke, *K-Nearest Neighbor* (K-NN)

ABSTRACT

Stroke is the third leading cause of disability and death in the world after cancer and heart disease. Many Indonesian people do not know and are aware of the signs or symptoms of a stroke as early as possible because the symptoms vary. Moreover, people are reluctant to go to the hospital to consult regarding the symptoms they are experiencing. Many approaches have been taken to predict disease by comparing the technology and performance of predictive data mining. Classification is a technique used to create data models that have not been classified. The purpose of this system is to produce a classification system for early diagnosis of stroke using the K-Nearest Neighbor (K-NN) method which can help and facilitate health workers in diagnosing stroke patients. The attributes used in this thesis are id, gender, age, hypertension, heart_disease, ever_married, work_type, residence_type, avg_glucose_level, bmi, smoking_status, stroke. In knowing a level of accuracy of the value of the data generated, it is continued with the process of calculating recall, accuracy, and precision, and with equations. Tests were carried out using neighboring values $k=3$, $k=5$, $k=7$, $k=9$, and $k=11$ for each distribution of testing data and training data, the study obtained the best evaluation value with precision = 95.59%, recall = 100.00 %, f-1 score = 97.74%, and accuracy = 95.59% by dividing the data by $k=5$ in k-folding 5.

Keywords: Stroke Disease, *K-Nearest Neighbor* (K-NN)

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN ORISINALITAS.....	iv
ABSTRAK	v
KATA PENGANTAR	v
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.2.1 Permasalahan Penelitian.....	2
1.2.2 Rencana Usulan.....	3
1.2.3 Pertanyaan Penelitian	3
1.3 Tujuan dan Manfaat Penelitian.....	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian	3
1.4 Batasan Masalah.....	3
1.5 Metode Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB II KAJIAN PUSTAKA	6
2.1 Pengertian Stroke	6
2.2 Faktor Resiko Penyakit Stroke	6
2.3 Pengertian Sistem	7
2.4 Pengertian Klasifikasi.....	8
2.5 K-Nearest Neighbor (K-NN).....	8
2.5.1 Algoritma K-Nearest Neighbor (K-NN).....	9
2.6 Evaluasi	10
2.7 Penelitian Terkait	11
BAB III METODE USULAN.....	13
3.1 Metodologi Penelitian	13

3.2	Dataset Penelitian	14
3.3	Perhitungan Algoritma Klasifikasi K-NN	15
3.4	Skenario Pengujian	20
3.5	Akurasi	20
BAB IV HASIL DAN PEMBAHASAN		22
4.1	Lingkungan Uji Coba	22
4.2	Tampilan Sistem	22
4.2.1	Tampilan Beranda	22
4.2.2	Tampilan Dataset	23
4.2.3	Tampilan Pengujian	23
4.3	Tahapan Proses	24
4.3.1	Membaca File Excel	24
4.3.2	Normalisasi Data	24
4.3.3	Pembagian Data Training dan Testing	24
4.3.4	KNN	25
4.3.5	Evaluasi Sistem	25
4.4	Hasil Pengujian	26
4.4.1	Pengujian tetangga k=3	26
4.4.2	Pengujian tetangga k=5	27
4.4.3	Pengujian tetangga k=7	28
4.4.4	Pengujian tetangga k=9	29
4.4.5	Pengujian tetangga k=11	30
BAB V Kesimpulan dan Saran		32
5.1	Kesimpulan	32
5.2	Saran	32
REFERENSI		33

DAFTAR TABEL

Tabel 3. 1 Data Fitur	14
Tabel 3. 2 Matriks <i>Confusion</i> untuk klasifikasi 2 kelas	15
Tabel 3. 3 Contoh Dataset Penyakit Stroke	16
Tabel 3. 4 Perubahan Dataset menjadi numeric	16
Tabel 3. 5 Perhitungan Normalisasi Dataset	17
Tabel 3. 6 Perhitungan nilai <i>eulicedan</i>	18
Tabel 3. 7 Perhitungan nilai k dan hasil <i>voting</i>	18
Tabel 3. 8 Contoh Dataset uji	19
Tabel 3. 9 Scenario Percobaan Pada Data	20
Tabel 3. 10 Confusion Matrix	21
Tabel 4. 1 Lingkungan Uji Coba	22
Tabel 4. 2 Pengujian tetangga k=3	26
Tabel 4. 3 Pengujian tetangga k=5	27
Tabel 4. 4 Pengujian tetangga k=7	28
Tabel 4. 5 Pengujian tetangga k=9	29
Tabel 4. 6 Pengujian tetangga k=11	30

DAFTAR GAMBAR

Gambar 3. 1 Rancangan Diagram Sistem	13
Gambar 4. 1 Tampilan beranda	22
Gambar 4. 2 Tampilan dataset	23
Gambar 4. 3 Tampilan pengujian.....	23
Gambar 4. 4 Grafik pengujian tetangga $k=3$	27
Gambar 4. 5 Grafik pengujian tetangga $k=5$	28
Gambar 4. 6 Grafik pengujian tetangga $k=7$	29
Gambar 4. 7 Grafik pengujian tetangga $k=9$	30
Gambar 4. 8 Grafik pengujian tetangga $k=11$	31