

ABSTRAK

M. Bima Sakti Admaja

(15.04.111.00080)

Dosen Pembimbing I : Dr. Fika Hastarita Rachman, S.T., M.Eng

Dosen Pembimbing II : Dr. Yeni Kustiyahningsih, S.Kom., M.Kom

Data dari World Health Organization (WHO) tentang *Non Communicable Disease* merupakan penyebab utama kematian secara global. Faktor penyebabnya pun karena pola hidup yang tidak sehat, sehingga dapat mengindikasikan meningkatnya risiko penyakit stroke. Banyak pasien meninggal karena baru dikenali pada saat tahap akhir, sehingga prediksi penyakit stroke perlu diterapkan dalam membantu menangani masalah tersebut. Penelitian ini fokus pada klasifikasi menggunakan metode kecerdasan komputasional dalam prediksi penyakit stroke. Metode yang digunakan adalah *Logistic Regresion*, metode ini akan digunakan untuk mengklasifikasi dataset penyakit stroke dengan fitur-fitur yang ada sehingga ditemukan target kelasnya yaitu terkena stroke atau tidak pada dataset pasien yang di proses. Dataset yang digunakan pada penelitian ini adalah *Stroke Prediction Dataset*. Sistem akan dicoba menggunakan *Microsoft Excel* dalam perhitungan manualnya dan bahasa pemrograman *python* dengan *google colabs research* sebagai platformnya. Dari hasil uji coba penggunaan metode *Logistic Regresion* memperoleh nilai *accuracy* sebesar 95%, *recall* sebesar 100%, dan *f1-score* mencapai 97%.

Kata kunci – Penyakit Stroke, Klasifikasi, Logistic Regresion, Identifikasi

ABSTRACT

Data from the World Health Organization (WHO) on *Non Communicable Disease* is the main cause of death globally. The causative factor is due to an unhealthy lifestyle, so it can indicate an increased risk of stroke. Many patients die because they are only recognized at a late stage, so stroke prediction needs to be applied to help deal with this problem. This study focuses on classification using computational intelligence methods in the prediction of stroke. The method used is *Logistic Regresion*, this method will be used to classify the stroke dataset with existing features so that the target class is found, namely stroke or not in the processed patient dataset. The dataset used in this study is the *Stroke Prediction Dataset*. The system will be tested using *Microsoft Excel* in manual calculations and using *python* programming language with *google colabs research* as the platform. From the results of trials using the *Logistic Regresion* method data that we get is : *accuracy* value 95%, *recall* value 100%, and *f1-score* reaches 97%.

Keyword – *Stroke Disease, Classification, Logistic Regression, Identification*

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PERSEMBAHAN.....	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.2.1 Permasalahan.....	2
1.2.2 Usulan Solusi	3
1.2.3 Pertanyaaan Penelitian	3
1.3 Tujuan dan Manfaat Penelitian.....	3
1.3.1 Tujuan	3
1.3.2 Manfaat	3
1.4 Batasan Masalah.....	3
1.5 Sistematika Penulisan.....	3
BAB II	5
STUDI PUSTAKA	5

2.1	Definisi Stroke.....	5
2.2	Klasifikasi.....	6
2.3	Logistik Regression	6
2.4	Regresi Logistik Biner.....	7
2.5	Distribusi Bernoulli	8
2.6	Penelitian Terkait	9
2.7	Confusion Matrix	11
BAB III.....		14
METODE USULAN		14
3.1	Desain Rancangan Sistem	14
3.2	Dataset	15
3.3	Metode Logistic Regression	16
3.4	Model Regresi Logistik	17
3.5	Evaluasi dan Validasi	19
3.6	Perhitungan Manual Menggunakan Excel.....	20
3.7	Skenario Pengujian.....	26
BAB IV		27
HASIL DAN PEMBAHASAN		27
4.1	Pembahasan.....	27
4.2	Tahap Pembangunan Sistem	27
4.3	Hasil dan Pembahasan.....	27
4.4	Import Library dan Dataset	27
4.5	Proses Drop Kolom.....	28
4.6	Mengisi Nilai Nand (Missing Value).....	28
4.7	Proses Encoding Data	29
4.8	Memisahkan Data Fitur dan Data Target Class	30

4.9 Split Dataset Untuk Training dan Testing.....	31
4.10 Implementasi Logistik Regresi dan Konfusi Matrix.....	32
4.11 Hasil Pengujian Tingkat Akurasi (Confusion Matrix).....	33
BAB V	34
PENUTUP	34
5.1 Kesimpulan.....	34
5.1 Saran.....	34
REFERENSI	35

DAFTAR GAMBAR

Gambar 2.1 Kurva Regresi Logistik	8
Gambar 3.1 Ilustrasi Rancangan Sistem	14
Gambar 3.2 Model Regresi Logistik.....	17
Gambar 3.3 Solver Excel	24
Gambar 3.4 Solver Excel	25
Gambar 4.1 Hasil Output Import Dataset	28
Gambar 4.2 Hasil Output Drop Kolom.....	28
Gambar 4.3 Output Isi Nilai Nand	29
Gambar 4.4 Output Encoding	30
Gambar 4.5 Output Data Fitur	31
Gambar 4.6 Hasil Split Data	31
Gambar 4.7 Hasil Uji Konfusi Matrix.....	33

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	10
Tabel 2.2 Confusion Matrix	11
Tabel 3.1 Fitur Attribute dataset Stroke Prediction Dataset	15
Tabel 3.2 Input Data.....	20
Tabel 3.3 Lanjutan Input Data	21
Tabel 3.4 Variabel Penduga (β)	21
Tabel 3.5 Rumus Logit Excel	22
Tabel 3.6 Exponensial Excel.....	22
Tabel 3.7 Probabilitas Excel	22
Tabel 3.8 Likelihood	23
Tabel 3.9 Log Likelihood.....	23
Tabel 3.10 Nilai Sum Log Likelihood	23
Tabel 3.11 Variabel (β) Penduga	25
Tabel 3.12 Confusion Matrix	26