

ABSTRAK

Nama : Shintia Hadigatri Henaulu
Program Studi : Teknik Informatika
Judul : Implementasi *Machine Learning* untuk Prediksi Serangan Stroke
Dosen Pembimbing : Dino Hariatma Putra, M.Kom

Pada saat ini banyak macam penyakit yang dapat melumpuhkan manusia seperti *bell's palsy*, *multiple sclerosis*, tumor otak, dan stroke. Stroke adalah suatu kondisi yang terjadi ketika pasokan darah ke suatu jaringan bagian otak tiba-tiba terganggu, karena sebagian sel-sel otak mengalami kematian akibat gangguan aliran darah karena sumbatan atau pecahnya pembuluh darah otak. Ada beberapa faktor yang dapat memprediksi serangan stroke seperti umur, berat badan, hipertensi, kadar gula darah dan lain-lain. Faktor tersebut diprediksi dalam penelitian ini menggunakan *Machine Learning* dengan metode *Random Forest Clasifier*, dataset diperoleh dari *Kaggle* dengan jumlah data 449 dan 12 attribut diantaranya jenis kelamin, umur, hipertensi, riwayat jantung, status pernikahan, pekerjaan, tempat tinggal, kadar gula darah, index massa tubuh, status merokok, dan stroke. Data dianalisis dengan melihat korelasi antar attribut, pelabelan data dilakukan dengan *Machine Learning* agar dapat dimodelkan menggunakan Algoritma *Random Forest*. Hasil akhir dari penelitian dengan Implementasikan *Machine Learning* untuk prediksi serangan stroke, memiliki hasil akurasi yang didapat sebesar 74%.

Kata Kunci : *Machine Learning*, *Random Forest Clasifier*, *Stroke*

ABSTRACT

At this time many kinds of diseases that can paralyze humans such as Bell's palsy, multiple sclerosis, brain tumors, and stroke. Stroke is a condition that occurs when the blood supply to a tissue part of the brain is suddenly interrupted, because some brain cells die due to disruption of blood flow due to blockage or rupture of brain blood vessels. There are several factors that can predict a stroke such as age, weight, hypertension, blood sugar levels and others. These factors were predicted in this study using Machine Learning with the Random Forest Classifier method, the dataset was obtained from Kaggle with a total of 449 data and 12 attributes including Gender, Age, Hypertension, Heart History, Marital Status, Occupation, Place of Residence, Blood Sugar Levels, Body Mass Index, smoking status, and stroke. Data were analyzed by looking at the correlation between attributes, data labeling was done with Machine Learning so that it could be modeled using the Random Forest Algorithm. The final result of the research by Implementing Machine Learning for prediction of stroke, has an accuracy of 74%.

Keywords: Machine Learning, Random Forest Classifier, Stroke

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	v
ABSTRAK	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB 1. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah	3
1.5. <i>State of The Art</i>	4
1.6. Sistematika Penulisan	7
BAB 2. TINJAUAN PUSTAKA	9
2.1. Penyakit Stroke	9
2.1.1. Definisi Stroke	9
2.1.2. Penyebab Stroke	9
2.2. Prediksi	9
2.3. <i>Machine Learning</i>	10
2.3.1. <i>Supervised Learning</i>	10
2.3.2. <i>Unsupervised Learning</i>	12
2.4. <i>Random Forest</i>	12

2.5. <i>Confusion matrix</i>	14
BAB 3. METODE	16
3.1. Analisis Masalah	16
3.2. Sumber data	16
3.3. Analisis Data	17
3.4. <i>Preprocessing Data</i>	17
3.4.1. <i>Data selection</i>	18
3.4.2. <i>Cleaning Data</i>	19
3.6. Analisis Metode <i>Random Forest</i>	20
BAB 4. PEMBAHASAN	39
4.1. Lingkungan Pengembangan	39
4.2. Implementasi Data	39
4.2.1. Penggunaan <i>Jupyter Notebook</i>	39
4.2.2. Pengambilan Data	40
4.3. Mengisi Baris Kosong	40
4.4. Proses <i>Encoder</i>	41
4.5. Data Uji	41
4.6. Hasil	42
BAB 5. KESIMPULAN DAN SARAN	44
5.1. Kesimpulan	44
5.2. Saran	44
DAFTAR REFERENSI	45

DAFTAR GAMBAR

Gambar 2.1. Skema Klasifikasi	11
Gambar 2.2. <i>Decision Tree</i>	13
Gambar 3.1. Dataset	16
Gambar 3.2. Korelasi Data	17
Gambar 3.3. Info Dataset	18
Gambar 3.4. <i>Cleaning Data</i>	19
Gambar 3.5. Pembentukan <i>root tree</i>	33
Gambar 3.6. Pembentukan <i>tree</i>	33
Gambar 3.7. Hasil <i>Tree</i>	37
Gambar 3.8. Alur <i>Random Forest</i>	38
Gambar 4.1. <i>Windows PowerShell</i>	39
Gambar 4.2. Pengambilan dataset	40
Gambar 4.3. <i>Syntax fillna</i>	40
Gambar 4.4. <i>Syntax Encoder</i>	41
Gambar 4.5. <i>Confusion Matrix</i>	43

DAFTAR TABEL

Tabel 2.1. <i>Confusion Matrix</i>	14
Tabel 3.1. <i>Label Encoding</i>	20
Tabel 3.2. <i>Sample Dataset</i>	21
Tabel 3.3. Perhitungan Gini Stroke	32
Tabel 3.4. Perhitungan Gini kadar gula	36
Tabel 4.1 Data Uji	42