



**OPTIMALISASI *FEATURE SELECTION* UNTUK
MENDETEKSI PENYAKIT DIABETES MELLITUS
MENGUNAKAN METODE *DECISION TREE* (STUDI
KASUS: PUSKESMAS SEMENDAWAI BARAT)**

SKRIPSI

**Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Program Studi Informatika**

Oleh:

Aplea Pameka

2020.11.0038

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER DAN SAINS
UNIVERSITAS IGM**

2023

**OPTIMALISASI *FEATURE SELECTION* UNTUK
MENDETEKSI PENYAKIT DIABETES MELLITUS
MENGUNAKAN METODE *DECISION TREE* (STUDI
KASUS: PUSKESMAS SEMENDAWAI BARAT)**



SKRIPSI

**Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Program Studi Informatika**

Oleh:

Aplea Pameka

2020.11.0038

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER DAN SAINS
UNIVERSITAS IGM**

2023

LEMBAR PENGESAHAN SKRIPSI

LEMBAR PENGESAHAN SKRIPSI

Optimalisasi *Feature Selection* Untuk
Mendeteksi Penyakit Diabetes Mellitus
Menggunakan Metode *Decision Tree*
(Studi Kasus: Puskesmas Semendawai
Barat)

Oleh

Aplea Pameka

NPM : 2020.11.0038

Palembang , 31 Januari 2024

Pembimbing I



Rudi Heriansyah, S.T., M.Eng., Ph.D
NIK. 2022.01.0315

Pembimbing II



Lastri Widya Astuti, M. Kom
NIK. 2003.01.0063

Mengetahui,

Dekan Fakultas Ilmu Komputer dan Sains

FAKULTAS ILMU KOM & SAINS



Rudi Heriansyah, S.T., M.Eng. Ph.D
NIK:2022.01.0315

LEMBAR PERSETUJUAN DEWAN PENGUJI

LEMBAR PERSETUJUAN DEWAN PENGUJI

Pada hari Kamis tanggal 25 Januari 2024 telah dilaksanakan ujian sidang skripsi :

Nama : Aplea Pameka

NPM : 2020.11.0038

Judul : Optimalisasi *Feature Selection* Untuk Mendeteksi Penyakit
Diabetes Mellitus Menggunakan Metode *Decision Tree* (Studi
Kasus: Puskesmas Semendawai Barat)

Oleh Prodi Teknik Informatika Fakultas Ilmu Komputer dan Sains Universitas Indo
Global Mandiri Palembang

Palembang, 31 Januari 2024


Penguji 1,



Dr. Shinta Puspasari, S.Si., M.Kom

NIK: 2015.01.0132

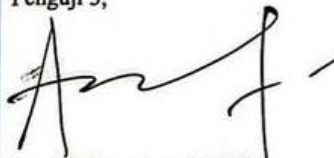
Penguji 2,



Muhammad Haviz Irfani, S.Si., M.T.I

NIK: 2021.03.0291

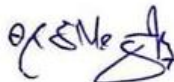
Penguji 3,



Lastri Widya Astuti, M.Kom

NIK: 2003.01.0063

Menyetujui,
Ka. Prodi Teknik Informatika



Zaid Romegar Mair, S.T., M.Cs

NIK: 2021.01.0307

LEMBAR REVISI SKRIPSI



SURAT KETERANGAN REVISI SKRIPSI
PROGRAM STUDI TEKNIK INFORMATIKA (S1)
FASILKOM DAN SAINS UNIVERSITAS INDO GLOBAL MANDIRI

Kami yang bertanda tangan dibawah ini, menerangkan bahwa :

Nama : Aplea Pameka
NPM : 2020.11.0038
Judul : Optimalisasi *Feature Selection* Untuk Mendeteksi Penyakit
Diabetes Mellitus Menggunakan Metode *Decision Tree* (Studi
Kasus: Puskesmas Semendawai Barat)

Mahasiswa yang namanya tercantum diatas, telah selesai merevisi penulisan SKRIPSI

Palembang, 31 Januari 2024

Penguji 1,

Dr. Shinta Puspasari, S.Si., M.Kom
NIK: 2015.01.0132

Penguji 2,

Muhammad Naviz Irfani, S.Si., M.T.I
NIK: 2021.03.0291

Penguji 3,

Lastri Widya Astuti, M.Kom
NIK: 2003.01.0063

Menyetujui,
Ka. Prodi Teknik Informatika

Zaid Romegar Mair, S.T., M.Cs
NIK: 2021.01.0307

**OPTIMALISASI *FEATURE SELECTION* UNTUK MENDETEKSI
PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE
DECISION TREE (STUDI KASUS: PUSKESMAS SEMENDAWAI BARAT)**

ABSTRAK

Diabetes atau juga dikenal sebagai diabetes mellitus, adalah penyakit kronis yang ditandai oleh tingginya kadar gula (glukosa) dalam darah. Penyebabnya berkaitan dengan masalah dalam produksi atau penggunaan insulin, hormon yang mengatur kadar gula darah. Penyakit diabetes mellitus tipe 2 merupakan masalah kesehatan yang semakin meningkat prevalensinya di seluruh dunia. Data riwayat medis dan karakteristik pasien, termasuk faktor-faktor seperti usia, jenis kelamin, riwayat penyakit keluarga, dan hasil tes laboratorium, dikumpulkan dari Puskesmas Semendawai Barat. Pada Penelitian ini menggunakan *Feature selection* dengan metode *Information Gain* dalam proses pemilihan subset fitur yang relevan dan didapatkan 5 fitur dengan nilai tertinggi. Selanjutnya, algoritma klasifikasi *Decision Tree* diterapkan pada subset fitur terpilih dan mendapatkan tingkat akurasi sebesar 96% dimana hasil ini konsisten dengan hasil tingkat akurasi yang hanya menggunakan algoritma *decision tree* tanpa melibatkan *feature selection* yaitu sebesar 96%.

Keywords: Diabetes Mellitus, *Feature Selection*, *Information Gain*, *Decision Tree*, Puskesmas Semendawai Barat, Deteksi

**OPTIMIZATION OF FEATURE SELECTION TO DETECT DIABETES
MELLITUS USING DECISION TREE METHOD (CASE STUDY:
SEMENDAWAI BARAT HEALTH CENTER)**

ABSTRACT

Diabetes or also known as diabetes mellitus, is a chronic disease characterized by high levels of sugar (glucose) in the blood. The cause is related to problems in the production or use of insulin, a hormone that regulates blood sugar levels. Type 2 diabetes mellitus is a health problem that is increasing in prevalence worldwide. Data on medical history and patient characteristics, including factors such as age, gender, family history of disease, and laboratory test results, were collected from the Semendawai Barat Health Center. This research uses Feature selection with the Information Gain method in the process of selecting a subset of relevant features and obtained 5 features with the highest value. Furthermore, the Decision Tree classification algorithm is applied to the selected feature subset and gets an accuracy rate of 96% where this result consistent with the result of the accuracy rate that only uses the decision tree algorithm without involving feature selection which is 96%.

Keywords: *Diabetes Mellitus, Feature Selection, Information Gain, Decision Tree, Semendawai Barat Health Center, Detection*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT dan Nabi Muhammad SAW berkat rahmat dan karunianya akhirnya penulisan skripsi ini dapat diselesaikan dengan baik dan tepat pada waktunya. Shalawat dan salam selalu dilimpahkan kepada junjungan kita Nabi besar Nabi Muhammad SAW semoga kita semua mendapatkan safaatnya di akhir nanti.

Proposal skripsi dengan judul “**Optimalisasi *Feature Selection* Untuk Mendeteksi Penyakit Diabetes Mellitus Menggunakan Metode *Decision Tree* (Studi Kasus: Puskesmas Semendawai Barat)**” disusun guna memenuhi persyaratan untuk menyelesaikan Pendidikan Strata I Program Studi Informatika Universitas Indo Global Mandiri Palembang.

Tak lupa penulis ucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan segala kemudahan, bimbingan, pengarahan, dorongan, bantuan baik moril maupun materil selama penyusunan skripsi ini. Selanjutnya, ucapan terima kasih penulis tujukan kepada yang terhormat:

1. Orangtua, saudara, dan keluarga saya yang telah memberikan doa, restu serta dukungan yang sangat besar selama menjalani kehidupan hingga perkuliahan di Universitas Indo Global Mandiri.
2. Bapak Dr. H. Marzuki Alie, S.E. M.M, selaku Rektor Universitas Indo Global Mandiri.
3. Bapak Rudi Heriansyah, S.T., M.Eng. Ph.D, selaku Dekan Fakultas Ilmu Komputer dan Pembimbing 1 yang telah membimbing saya selama pembuatan skripsi ini.
4. Bapak Zaid Romegar Mair, S.T., M.Cs, selaku Ketua Program Studi Teknik Informatika dan selaku Dosen Pembimbing Akademik saya.
5. Ibu Lastri Widya Astuti, M.Kom, selaku Pembimbing II yang telah membimbing saya selama pembuatan skripsi ini.

6. Bapak/Ibu Dosen Program Studi Informatika yang telah memberikan ilmu dan pembelajaran selama masa perkuliahan.
7. Hubertus Maria Wira, yang telah meluangkan waktu menemani saya mengerjakan skripsi ini.
8. Sahabat dan rekan seperjuangan yang telah menjadi pendengar dan membantu penulis dalam setiap keluh dan kesah.
9. Almamater.

Tiada harapan lain selain semoga Allah SWT membalas segala niat baik kepada semua pihak yang telah membantu. Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan karena kurangnya pengetahuan dan pengalaman penulis. Kritik dan saran yang membangun sangat diharapkan sebagai bahan acuan untuk menyempurnakan laporan ini.

Palembang, 19 November 2023

Penulis,

Aplea Pameka

2020.11.0038

DAFTAR ISI

HALAMAN JUDUL DALAM	ii
LEMBAR PENGESAHAN SKRIPSI	iii
LEMBAR PERSETUJUAN DEWAN PENGUJI	iv
LEMBAR REVISI SKRIPSI	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR RUMUS	xvi
DAFTAR LAMPIRAN	xvii
MOTTO DAN PERSEMBAHAN	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat	4
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEORI	6
2.1 Data Mining	6
2.2 <i>Decision Tree</i>	7
2.3 <i>Feature Selection</i>	10
2.4 Deteksi	11
2.5 Diabetes Mellitus	12
2.6 Normalisasi Data	14
2.7 <i>Cross Validation</i>	14
2.8 <i>Confusion Matrix</i>	15
2.9 Flowchart	17

2.10	Python	19
2.10	Penelitian Terdahulu.....	23
BAB III METODOLOGI PENELITIAN.....		27
3.1	Tahapan Penelitian	27
3.2	Studi Literatur.....	28
3.3	Pengumpulan Data	28
3.4	Kebutuhan Perangkat Keras dan Perangkat Lunak	30
3.4.1	Kebutuhan Perangkat Keras	30
3.4.2	Kebutuhan Perangkat Lunak	30
3.5	Perancangan.....	30
3.5.1	Persiapan Data.....	30
3.5.2	<i>Feature Selection</i>	31
3.5.3	Implementasi Algoritma <i>Decision Tree</i>	32
3.5.4	<i>K-Fold Cross Validation</i>	33
3.6	Perhitungan <i>Feature selection</i> dan Model <i>Decision Tree</i>	35
3.6.1	Cara Menghitung dan Mendapatkan Fitur Terbaik Menggunakan <i>Feature Selection</i>	37
3.6.2	Perhitungan Manual Algoritma <i>Decision Tree</i>	39
3.7	Evaluasi	45
3.8	Ringkasan	46
BAB IV IMPLEMENTASI DAN EVALUASI.....		47
4.1	Implementasi	47
4.1.1	Memuat <i>Dataset</i>	47
4.1.2	<i>Data Cleansing</i>	52
4.1.3	<i>K-Fold Cross Validation (5 Folds)</i>	54
4.1.4	Implementasi <i>Feature Selection</i>	57
4.1.5	Implementasi Algoritma <i>Decision Tree</i>	61
4.2	Evaluasi	62
4.2.1	Hasil Pengujian	67
4.2.2	Hasil Pengujian Tanpa menggunakan <i>Feature Selection</i>	68
BAB V KESIMPULAN DAN SARAN.....		70
DAFTAR PUSTAKA		71

LAMPIRAN.....	75
---------------	----

DAFTAR GAMBAR

Gambar 2.1 Konsep <i>Decision Tree</i>	7
Gambar 2.2 Proses <i>Decision Tree</i>	8
Gambar 2.3 Ilustrasi dari <i>K-Fold Validation</i>	15
Gambar 3.1 Diagram Alir Tahapan Penelitian	27
Gambar 3.2 Flowchart <i>Data Preparation</i>	31
Gambar 3.3 Flowchart <i>Data Cleaning</i>	31
Gambar 3.4 Flowchart <i>feature selection</i>	32
Gambar 3.5 Flowchart implementasi Algoritma <i>Decicion Tree</i>	33
Gambar 3.6 Ilustrasi Implementasi <i>K-fold Cross Validation 5 fol</i>	34
Gambar 3.7 <i>Decision Tree Root</i>	42
Gambar 3.8 <i>Decision tree Node Level 1 (Node 1.1)</i>	45
Gambar 4.1 Kode program Input Data	47
Gambar 4.2 Proses penyimpanan dataframe dalam file csv	48
Gambar 4.3 Kode Program Normalisasi Data	48
Gambar 4.4 Hasil Proses Normalisasi Pada Data.....	49
Gambar 4.5 Kode Program Menampilkan Data	49
Gambar 4.6 Tampilan Dataset yang telah dimuat	50
Gambar 4.7 Kode Program Menampilkan Jumlah Dataset	50
Gambar 4.8 Tampilan jumlah dataset yang telah dimuat	50
Gambar 4.9 Kode Program Menampilkan 5 baris Akhir dan 5 Baris Awal.....	51
Gambar 4.10 Tampilan 5 baris pertama dalam dataset	51
Gambar 4.11 Tampilan 5 baris akhir dataset.....	51
Gambar 4.12 Kode program Pemeriksaan Data Duplikat	52
Gambar 4.13 Tampilan kolom apabila tidak memiliki data yang null	53
Gambar 4.14 Kode Program Data Duplikat	53
Gambar 4.15 Tampilan jumlah data duplikat	54
Gambar 4.16 Kode Program <i>K-Fold Cross Validation</i>	55
Gambar 4.17 Dataset training pada folds 1	56

Gambar 4.18 Dataset training pada Fold 2	56
Gambar 4.19 Dataset Training pada fold 5.....	57
Gambar 4.20 Kode Program Implementasi Feature Selection	58
Gambar 4.21 Nilai dari Atribut dengan perhitungan feature selection.....	59
Gambar 4.22 Hasil Feature Selection yang dilakukan pada dataset penderita penyakit diabetes mellitus	60
Gambar 4.23 Kode Program Implementasi Algoritma Decision Tree	61
Gambar 4.24 Kode Program Pemodelan Algoritma Decision Tree	63
Gambar 4.25 Kode Program Tree Algoritma Decision Tree	67
Gambar 4.26 Hasil Tree terbaik pada model decision tree.....	68

DAFTAR TABEL

Tabel 2.1 Simbol <i>Flowchart</i>	17
Tabel 2.2 Penelitian Terdahulu.....	24
Tabel 3.1 Atribut yang digunakan dalam dataset.....	29
Tabel 3.2 Data Sample Penderita Penyakit Diabetes Mellitus	36
Tabel 3.3 Data Sample Setelah Perhitungan <i>Feature Selection</i>	39
Tabel 3.4 Hasil perhitungan <i>Node</i> 1	41
Tabel 3.5 Hasil perhitungan <i>Node</i> Level 1 (<i>Node</i> 1.1).....	44
Tabel 3.6 Tabel Perbandingan hasil <i>confusion matrix</i> model <i>decision tree</i> pada masing masing <i>fold</i>	45
Tabel 4.1 <i>Confusion Matrix</i> pada model <i>decision tree</i> yang telah dibuat.....	63
Tabel 4.2 Tabel Perbandingan hasil <i>confusion matrix</i> model <i>decision tree</i> pada masing masing <i>fold</i>	67
Tabel 4.3 Tabel Perbandingan hasil <i>confusion matrix</i> model <i>decision tree</i> pada masing masing <i>fold</i> tanpa <i>feature selection</i>	68

DAFTAR RUMUS

Rumus 2.1 Gain <i>Decision Tree</i>	9
Rumus 2.2 Entropy <i>Decision Tree</i>	10
Rumus 2.3 Entropy <i>Feature Selection</i>	11
Rumus 2.4 Rumus <i>Information Gain</i>	11
Rumus 2.5 Menghitung Nilai Akurasi	16
Rumus 2.6 Menghitung Nilai Presisi	16
Rumus 2.7 Menghitung Nilai Recall.....	16

DAFTAR LAMPIRAN

Lampiran 1 Daftar Riwayat Hidup.....	75
Lampiran 2 Kartu Bimbingan.....	76
Lampiran 3 Surat Keterangan Tidak Plagiat	77
Lampiran 4 Surat Pengantar Kampus.....	78
Lampiran 5 Surat Keterangan Siap Sidang Proposal Skripsi.....	79
Lampiran 6 Surat Keterangan Revisi Proposal Skripsi	80
Lampiran 7 Surat Keterangan Siap Sidang Skripsi.....	81
Lampiran 8 Surat Persetujuan Ujian Skripsi	82
Lampiran 9 Bebas Pustaka	83
Lampiran 10 Surat Balasan Tempat Pengambilan Data Penelitian.....	84
Lampiran 11 Tabel Dataset Penyakit Diabetes Mellitus	80