

**KLASTERISASI VARIABEL INDIKATOR DIABETES MELLITUS
DENGAN PENDEKATAN ALGORITMA FUZZY C-MEANS**

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data**



Disusun Oleh:

AFDHAL RESHANDA ADIWIDYATMA

20083010024

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2024**

LEMBAR PENGESAHAN

**KLASTERISASI VARIABEL INDIKATOR DIABETES MELLITUS
DENGAN PENDEKATAN ALGORITMA FUZZY C-MEANS**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data
pada : **Senin, 15 Juli 2024**

**Program Studi S-1 Sains Data
Fakultas Ilmu Komputer**

**Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya**

Oleh :

AFDHAL RESHANDA ADIWIDYATMA

NPM. 20083010024

Disetujui oleh Tim Penguji Skripsi :

Penguji 1

Penguji 2

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU
NIP. 198012052005011002

Wahyu Syaifulah, S.Kom., M.Kom
NIP. 198608252021211003

Pembimbing 1

Pembimbing 2

Dr. Ir. F Gede Susrama Mas Divasa, ST., MT., IPU
NIP. 197006192021211009

Trinono, S.Si., M.Si.
NIP. 199509082022031003

Mengetahui,

**Fakultas Ilmu Komputer
Dekan,**

**Program Studi Sains Data
Fakultas Ilmu Komputer
Koordinator,**

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 196811261994032001

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.
NIP. 198012052005011002

Surabaya, Juli, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Afdhal Reshanda Adiwidyatma
NPM : 20083010024
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

Klasterisasi Variabel Indikator Diabetes *Mellitus* dengan Pendekatan Algoritma *Fuzzy C-Means*

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ software/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 4 Juli 2024

Hormat Saya



Afdhal Reshanda Adiwiyatma

NPM. 20083010024

ABSTRAK

KLASTERISASI VARIABEL INDIKATOR DIABETES *MELLITUS* DENGAN PENDEKATAN ALGORITMA *FUZZY C-MEANS*

Nama Mahasiswa / NPM : Afdhal Reshanda Adiwidyatma / 20083010024
Program Studi : Sains Data, FASILKOM, UPN Veteran Jatim
Dosen Pembimbing 1 : Dr.Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU
Dosen Pembimbing 2 : Trimono, S.Si., M.Si

Abstraks

Penelitian ini berfokus pada efektivitas algoritma *Clustering*, yaitu *Fuzzy C-Means* dengan menggunakan algoritma *K-means* sebagai metode pendukung, dalam menganalisis variabel indikator penyebab Diabetes *Mellitus*. Diabetes *mellitus* merupakan penyakit kronis yang ditandai oleh tingginya kadar gula (glukosa) dalam darah. Indonesia menempati urutan ke-5 dengan penderita Diabetes *Mellitus* tertinggi di dunia. Penelitian ini bertujuan untuk memahami pola variabel indikator penyebab diabetes *mellitus* serta menguji efektivitas algoritma klastering yang digunakan. Adapun metode analisis data meliputi pengumpulan data, pra-pemrosesan data, pembagian jumlah klaster, implementasi algoritma, penyesuaian model, pelatihan model, evaluasi model, dan analisis hasil. Hasil penelitian menunjukkan bahwa algoritma *Fuzzy C-Means* mendapatkan skor *Coefficient of Fuzzynes* sebesar 0,23 dengan validasi skor 0,40, sementara untuk metode pendukung yang dipakai algoritma *K-means* mendapatkan skor validasi 0,32. Hasil ini menunjukkan bahwa algoritma *Fuzzy C-Means* lebih unggul dalam mengklasterisasi variabel indikator penyebab diabetes *mellitus*. Adapun hasil variabel apa saja yang paling berpengaruh pada nilai *cluster* 0 dan 1. Dimana *cluster* 0 merupakan *cluster* yang menunjukkan variabel mana yang lebih beresiko mengalami diabetes, sementara *cluster* 1 merupakan *cluster* yang nilainya menunjukkan variabel apa saja yang jauh dari resiko penyebab diabetes *mellitus*.

Kata kunci: *Diabetes mellitus, Fuzzy C-Means, k-means, Clustering.*

ABSTRACT

CLASSTERIZATION OF DIABETES MELLITUS INDICATOR VARIABLES USING THE FUZZY C-MEANS ALGORITHM APPROACH

Student Name / NPM : Afdhal Reshanda Adiwidyatma / 20083010024
Study Program : Sains Data, FASILKOM,UPN Veteran Jatim
Advisor 1 : Dr.Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU
Advisor 2 : Trimono, S.Si., M.Si

Abstract

This research focuses on the effectiveness of the Clustering algorithm, namely Fuzzy C-Means using the K-means algorithm as a supporting method, in analyzing indicator variables that cause Diabetes Mellitus. Diabetes mellitus is a chronic disease characterized by high levels of sugar (glucose) in the blood. Indonesia ranks 5th with the highest number of Diabetes Mellitus sufferers in the world. This research aims to understand the pattern of indicator variables that cause diabetes mellitus and test the effectiveness of the clustering algorithm used. The data analysis methods include data collection, data pre-processing, dividing the number of clusters, algorithm implementation, model adjustment, model training, model evaluation, and results analysis. The research results show that the Fuzzy C-Means algorithm gets a Coefficient of Fuzzyness score of 0.23 with a validation score of 0.40, while the supporting method used by the K-means algorithm gets a validation score of 0.32. These results indicate that the Fuzzy C-Means algorithm is superior in clustering indicator variables that cause diabetes mellitus. The results of which variables have the most influence on the values of clusters 0 and 1. Where cluster 0 is a cluster that shows which variables are more at risk of developing diabetes, while cluster 1 is a cluster whose values show which variables are far from the risk of causing diabetes mellitus.

Keywords: *Diabetes mellitus, Fuzzy C-Means, k-means, Clustering.*

KATA PENGANTAR

Dengan puji syukur kehadirat Tuhan Yang Maha Esa, atas segala limpahan rahmat dan hidayah-Nya kepada kita semua, sehingga Penulis dapat menyelesaikan laporan skripsi dengan judul “Klasterisasi Variabel Indikator Diabetes *Mellitus* dengan Pendekatan Algoritma *Fuzzy C-Means*”. Penulis menyadari bahwa di dalam pembuatan laporan ini berkat bantuan dan tuntunan Tuhan Yang Maha Esa dan tidak lepas dari bantuan berbagai pihak, pihak-pihak tersebut antara lain:

1. Kedua orang tua dan keluarga serta teman-teman Sains Data yang senantiasa memberikan dorongan dan doa.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T.selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur
4. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Bapak Dr. Ir. I Gede Susrama Mas Diyasa, S.T., M.T., IPU dan Bapak Trimono, S.Si., M.Si selaku Dosen Pembimbing 1 dan 2.
6. Bapak dan Ibu dosen Program Studi Sains Data UPN “Veteran” Jawa Timur yang sudah berkenan untuk memberikan waktu untuk berkontribusi dalam penelitian ini.
7. NPM 20083010029 sebagai pendamping yang selalu memberikan bantuan terhadap jalanya pembuatan skripsi ini.

Penulis menyadari bahwa dalam proses penulisan laporan ini masih dari jauh dari kesempurnaan baik materi maupun cara penulisannya. Namun demikian, Penulis telah berupaya dengan segala kemampuan dan pengetahuan yang dimiliki sehingga dapat selesai dengan baik.

Surabaya, 4 Juli 2024

Afdhal Reshanda
Adiwidyatma

DAFTAR ISI

LEMBAR PENGESAHAN.....	<i>ii</i>
SURAT PERNYATAAN.....	<i>iii</i>
ABSTRAK	<i>iv</i>
<i>ABSTRACT</i>	<i>v</i>
KATA PENGANTAR.....	<i>vi</i>
DAFTAR ISI.....	<i>vii</i>
DAFTAR GAMBAR	<i>ix</i>
DAFTAR TABEL	<i>x</i>
DAFTAR LAMPIRAN	<i>xi</i>
BAB I PENDAHULUAN	<i>1</i>
1.1 Latar Belakang.....	1
1.1.1 Rumusan Masalah	6
1.2. Batasan Masalah.....	6
1.3. Tujuan Penelitian.....	7
1.4. Manfaat Penelitian.....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 Dasar Teori	9
2.1.1 Diabetes Mellitus.....	9
2.1.2 <i>Machine learning</i>	11
2.1.3 <i>Data mining</i>	12
2.1.4 <i>Clustering</i>	14
2.1.5 <i>K-means Clustering</i>	16
2.1.6 <i>Fuzzy C-Means Clustering</i>	18
2.1.7 Metode Elbow	25
2.1.8 <i>Silhouette Score</i>	26
2.2 Penelitian Terdahulu.....	27
BAB III METODOLOGI PENELITIAN.....	32
3.1. Sumber Data	32
3.2. Diagram Alir Penelitian.....	32
3.3. Jadwal Penelitian	35
BAB IV HASIL DAN PEMBAHASAN.....	36

4.1.	Hasil Penelitian.....	36
4.1.1	Pengumpulan Data.....	36
4.1.2	Praproses Penelitian.....	37
4.1.3	<i>Fuzzy C-Means</i>	39
4.1.4	<i>K-means</i>	47
4.1.5	<i>Deployment Aplikasi</i>	49
4.2.	Pembahasan	50
BAB V.....		49
PENUTUP.....		49
5.1	Kesimpulan.....	49
5.2	Saran.....	50
DAFTAR PUSTAKA.....		53
BIODATA PENULIS.....		64

DAFTAR GAMBAR

Gambar 2.1 Contoh Proses <i>Clustering</i>	16
Gambar 2.2 Alur Pengerjaan <i>Fuzzy C-Means</i>	19
Gambar 3.1 Flowchart Penelitian	33
Gambar 4.1 Tampilan <i>Input Library</i>	37
Gambar 4.2 Tampilan Dataset Setelah Drop Variabel	37
Gambar 4.3 Hasil Cek <i>Missing Value</i>	38
Gambar 4.4 Hasil Cek Duplikasi Data	39
Gambar 4.5 Hasil Cek <i>Unique Value</i>	39
Gambar 4.6 <i>Scatter Plot</i> Hasil Analisa <i>Cluster</i>	41
Gambar 4.7 Hasil Distribusi Data pada <i>Cluster</i>	42
Gambar 4.8 Hasil Analisa Rata-rata Variabel Gula Darah Acak	43
Gambar 4.9 Hasil Analisa Rata-rata Variabel Gula Darah 2 Jam PP.....	44
Gambar 4.10 Hasil Analisa Rata-rata Variabel Gula Darah Puasa	44
Gambar 4.11 Hasil Analisa Rata-rata Variabel Hba1c.....	45
Gambar 4.12 Histogram Variabel Paling Berpengaruh.....	46
Gambar 4.13 Hasil Validasi <i>Score</i> Terhadap Model	46
Gambar 4.14 Grafik <i>Elbow Method</i>	47
Gambar 4.15 Inisialisasi Nilai <i>Cluster</i> pada Model	48
Gambar 4.16 Hasil Sebaran Data pada <i>Cluster</i>	48
Gambar 4.17 Hasil Validasi <i>Score</i> Terhadap Model.....	48
Gambar 4.18 Halaman Awal Aplikasi.....	49
Gambar 4.19 Hasil <i>Cluster</i> pada Aplikasi	50

DAFTAR TABEL

Tabel 2.1. Tabel studi literatur yang relevan dalam Skripsi.....	27
Tabel 3.1. Tabel Struktur Data	32
Tabel 3.2. Jadwal Kegiatan	35
Tabel 4.1. Dataset Penelitian	36
Tabel 4.2 <i>Value Input</i>	40
Tabel 4.3. Hasil Analisa <i>Cluster</i>	40

DAFTAR LAMPIRAN

Lampiran 1	57
Lampiran 2	57
Lampiran 3	57
Lampiran 4	62