

**PREDIKSI PENYAKIT DIABETES
MENGUNAKAN ALGORITMA
REGRESI LOGISTIK**

SKRIPSI



**Oleh:
Sugianto
200210039**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2025**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Sugianto
NPM : 200210039
Fakultas : Teknik dan Komputer
Program Studi : Teknik Informatika

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

PREDIKSI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA REGRESI LOGISTIK

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, di dalam Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip di dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan serta proses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 8Februari 2025


Sugianto

200210039

PREDIKSI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA REGRESI LOGISTIK

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar Sarjana**

Oleh

Sugianto

200210039

**Telah disetujui oleh Pembimbing
pada tanggal seperti tertera di bawah ini**

Batam, 8 Februari 2025

A handwritten signature in black ink, appearing to read 'Andi Maslan', with a stylized flourish at the end.

Andi Maslan, S.T., M.SI., Ph.D.

Pembimbing

ABSTRAK

Machine Learning adalah salah satu cabang AI yang berfokus pada pengembangan algoritma dan teknik yang membuat komputer menghasilkan prediksi. Jumlah orang yang menderita diabetes meningkat setiap tahunnya. Adapun survei yang dilakukan oleh Federasi Diabetes Internasional (IDF), populasi diabetes global diperkirakan akan mencapai 537 juta pada tahun 2021, dan angka ini diperkirakan akan meningkat lebih lanjut untuk melebihi 780 juta pada 2045. Oleh sebab itu, peneliti ingin meneliti kecenderungan penderita penyakit diabetes melibatkan variabel apa saja. Tujuan utama dari penelitian ini adalah untuk mendiagnosis dan memprediksi apakah pasien memiliki diabetes atau tidak. Pendekatan ini menggunakan regresi logistik, alat statistik untuk memodelkan klasifikasi individual keberadaan atau tidak adanya diabetes. Untuk penelitiannya peneliti memodelkan *dataset* serta melatih dan mentraining data tersebut. Tingkat akurasi dalam mendeteksi penyakit diabetes menggunakan algoritma regresi logistik sebesar 0.76, presisi 0.66, *recall* 0.55 dan *f1-score* = 0.60. Menurut hasil prediksi risiko diabetes, 43% responden mempertimbangkan kondisi tersebut. Akibatnya, telah terbukti bahwa normalisasi meningkatkan akurasi prediksi risiko diabetes menggunakan metode regresi logistik. Hasil penelitiannya adalah *website* prediksi penyakit diabetes. Berdasarkan variabel yang disertakan, diharapkan bahwa prediksi yang dibuat oleh model ini akan berfungsi sebagai panduan bagi masyarakat umum dalam memahami kehidupan sehat dan pencegahan diabetes.

Kata kunci : Prediksi Diabetes; Regresi Logistik; Prediksi Risiko, Variabel Diabetes; Panduan Gaya Hidup Sehat

ABSTRACT

Machine Learning is one of the branches of AI that focuses on the development of algorithms and techniques that make computers produce predictions. The number of people suffering from diabetes is increasing every year. As for the survey conducted by the International Diabetes Federation (IDF), the global diabetes population is projected to reach 537 million by 2021, and this figure is expected to rise further to exceed 780 million by 2045. Therefore, the researchers wanted to study the tendency of people with diabetes to involve any variable. The main objective of this study is to diagnose and predict whether a patient has diabetes or not. This approach uses logistical regression, a statistical tool to model individual classifications of the existence or absence of diabetes. For research, researchers model data sets and train and train the data. Then the result of the prediction is a value of accuracy of 0.76, precision of 0.66, recall of 0.55 and f1-score = 0.60. According to the prediction of diabetes risk, 43% of respondents considered the condition. As a result, it has been proven that normalization improves the accuracy of predicting diabetes risk using logistical regression methods. The results of his research are a website for the prediction of diabetes. Based on the included variables, it is expected that the predictions made by this model will serve as a guide for the general public in understanding healthy living and diabetes prevention.

Keywords: Diabetes Prediction; Logistical Regression; Risk Prediction; Diabetes Variable; Healthy Lifestyle Guidance

KATA PENGANTAR

Puji syukur kepada Tuhan Yang maha Esa yang telah melimpahkan segala rahmat dan karunia-nya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyampaikan ucapan terimakasih kepada:

1. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI. selaku Rektor Universitas Putera Batam;
2. Bapak Welly Sugianto, S.T., M.M. Ph.D selaku Dekan Fakultas Teknik dan Komputer;
3. Bapak Andi Maslan, S.T., M.SI.,Ph.D selaku Ketua Program Studi Teknik Informatika dan juga pembimbing skripsi pada Program Studi Teknik Informatika Universitas Putera Batam;
4. Dosen dan Staf Universitas Putera Batam;
5. Orang tua penulis yang selalu memberikan dukungan dan doa agar penulis dapat menyelesaikan skripsi ini;
6. Pasangan saya Ernestina yang selalu mensupport saya dalam penulisan skripsi ini.
7. Teman seangkatan 2020 yang selalu memberikan dukungan, semangat dan bantuan selama proses pengerjaan skripsi ini.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurahkan hidayah serta taufik-nya, Amin.

Batam, 8 Februari 2025

Sugianto
200210039

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
SURAT PERNYATAAN ORISINALITAS.....	iii
HALAMAN PENGESAHAN	iv
ABSTRAK.....	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	4
1.3 Batasan Masalah.....	4
1.4 Rumusan Masalah	4
1.5 Tujuan Penelitian.....	5
1.6 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	7
2.1 Teori Dasar	7
2.1.1 <i>Machine Learning</i> (ML).....	7
2.1.2 <i>Method Machine Learning</i>	9
2.2 Teori Khusus.....	17
2.2.1 Metode Penelitian	17
2.2.2 UML (<i>Unified Modelling Language</i>)	20
1. Use Case Diagram	22
2. Activity Diagram	25
3. Sequence Diagram.....	26
4. Class Diagram	29
2.2.3 <i>Python</i>	31
2.2.4 Metode Pengujian	32
2.3 <i>Software</i> Pendukung	32
2.4 Penelitian Terdahulu.....	33

2.5	Kerangka Pemikiran	35
BAB III METODE PENELITIAN.....		37
3.1	Desain Penelitian	37
3.2	Teknik Pengumpulan Data.....	39
3.2.1	Observasi	39
3.2.2	Studi Pustaka	39
3.3	Variabel Penelitian.....	40
3.4	Teknik Pengambilan Sampel	41
3.5	Teknik Analisis Data.....	43
3.5.1	Proses Analisis Data	43
3.5.2	Pemilihan Atribut (Selection Features).....	53
3.5.3	Data Training dan Data Testing	55
3.5.4	Menentukan Nilai Logistik <i>Regression</i>	55
3.6	Perancangan Sistem.....	57
3.6.1	Perancangan UML.....	57
1.	Diagram Usecase	57
2.	Diagram Aktivitas.....	59
3.	Diagram Sequence.....	67
4.	Diagram Class	71
3.6.2	Perancangan <i>Interface</i>	73
3.7	Lokasi dan Jadwal Penelitian	76
BAB IV HASIL DAN PEMBAHASAN		78
4.1	Hasil Penelitian.....	78
4.1.1	Hasil Pengumpulan Dataset.....	78
4.1.2	Hasil Preprocessing	78
4.1.3	Hasil Prediksi Algoritma Regresi Logistik.....	79
4.2	Hasil Evaluasi Model Logistik Regression	85
4.2.1	Hasil <i>Training Dataset</i>	85
4.2.2	Hasil <i>Testing Dataset</i>	85
4.3	Hasil Implementasi Sistem Prediksi.....	88
4.3.1	Hasil Rancangan Aplikasi.....	88
4.3.2	Hasil Pengujian.....	92

4.4	Pembahasan	94
4.4.1	Memprediksi Penyakit Diabetes Menggunakan Algoritma Regresi Logistik.....	94
4.4.2	Mengimplementasikan Desain Website Prediksi Penyakit Diabetes Menggunakan Algoritma Regresi Logistik.....	97
BAB V	KESIMPULAN DAN SARAN.....	105
5.1	Simpulan.....	105
5.2	Saran.....	105
DAFTAR PUSTAKA.....		107
LAMPIRAN.....		110
	Lampiran 1. Surat Izin Penelitian	110
	Lampiran 2: Bukti Dokumentasi	111
	Lampiran 3: Surat Keterangan LOA	113
	Lampiran 4: Riwayat Hidup	114
	Lampiran 5: Hasil Turnitin Skripsi.....	115
	Lampiran 6: Hasil Turnitin Jurnal	116

DAFTAR GAMBAR

Gambar 2. 1 Proses dari ML.....	9
Gambar 2. 2 Algoritma <i>K-NN</i>	10
Gambar 2. 3 Algoritma <i>Decision Tree</i>	11
Gambar 2. 4 Algoritma <i>Support Vector Machine</i>	12
Gambar 2. 5 Algoritma <i>Random Forest</i>	13
Gambar 2. 6 Algoritma Regresi.....	14
Gambar 2. 7 Algoritma <i>K-means Clustering</i>	15
Gambar 2. 8 Algoritma <i>Hierarchical Clustering</i>	16
Gambar 2. 9 Algoritma DBSCAN.....	17
Gambar 2. 10 Tahapan Algoritma Regresi Logistik	19
Gambar 2. 11 <i>Visual Studio Code</i>	32
Gambar 2. 12 Kerangka Berpikir	36
Gambar 3. 1 Desain Penelitian	37
Gambar 3. 2 Proses Analisis Data	44
Gambar 3. 3 Rentang Nilai Tiap Variabel	46
Gambar 3. 4 Analisa <i>Missing Value</i>	48
Gambar 3. 6 Analisa Nilai rata-rata Kolom.....	48
Gambar 3. 8 Penghitungan <i>Z-score</i>	50
Gambar 3. 9 Perintah Pengecekan <i>Outlier</i>	50
Gambar 3. 10 Gambar <i>Outlier</i>	51
Gambar 3. 11 Analisis Fitur.....	52
Gambar 3. 12 Korelasi antar Fitur.....	53
Gambar 3. 13 <i>Dataset Training dan Data Testing</i>	55
Gambar 3. 14 Diagram <i>Usecase</i>	57
Gambar 3. 15 Diagram Aktivitas <i>Home</i>	60
Gambar 3. 16 Diagram Aktivitas <i>Data Info</i>	61
Gambar 3. 17 Diagram Aktivitas <i>Prediction</i>	62
Gambar 3. 18 Diagram Aktivitas <i>Visualization</i>	63
Gambar 3. 19 Diagram Aktivitas <i>Contact Us</i>	64
Gambar 3. 20 Diagram Aktivitas <i>Chatbot</i>	66
Gambar 3. 21 Diagram Urutan Menu <i>Home</i>	67
Gambar 3. 22 Diagram Urutan Menu <i>Data Info</i>	68
Gambar 3. 23 Diagram Urutan Menu <i>Prediction</i>	68
Gambar 3. 24 Diagram Urutan Menu <i>Visualization</i>	69
Gambar 3. 25 Diagram Urutan Menu <i>Contact Us</i>	70
Gambar 3. 26 Diagram Urutan Menu FAQ	70
Gambar 3. 27 Diagram Urutan Menu FAQ.....	71
Gambar 3. 28 Diagram <i>Class</i>	72
Gambar 3. 29 Halaman <i>Home</i>	73
Gambar 3. 30 Halaman <i>Data Info</i>	73
Gambar 3. 31 Halaman <i>Prediction</i>	74

Gambar 3. 32 Halaman <i>Visualization</i>	74
Gambar 3. 33 Halaman <i>Contact Us</i>	75
Gambar 3. 34 Halaman FAQ.....	75
Gambar 3. 35 Halaman <i>Chatbot</i>	76
Gambar 3. 36 Lokasi Penelitian	77
Gambar 3. 37 Jadwal Penelitian	77
Gambar 4. 1 Dataset yang Digunakan.....	78
Gambar 4. 2 Input framework yang Akan Digunakan.....	79
Gambar 4. 3 <i>Upload Dataset</i>	80
Gambar 4. 4 Pembacaan file <i>dataset</i>	80
Gambar 4. 5 Menampilkan Lima Baris Pertama Data.....	81
Gambar 4. 6 Memisahkan Fitur.....	81
Gambar 4. 7 Bagi Data menjadi <i>training</i> dan <i>testing</i>	81
Gambar 4. 8 Model Regresi Logistik	82
Gambar 4. 9 Melatih Model	82
Gambar 4. 10 Hasil Jika Berhasil Melatih Data	82
Gambar 4. 11 Cara Mendapatkan Nilai Probabilitas	83
Gambar 4. 12 Nilai Probabilitas	83
Gambar 4. 13 Analisis Korelasi.....	84
Gambar 4. 14 Hubungan Korelasi	84
Gambar 4. 15 Proses Testing.....	87
Gambar 4. 16 Halaman Utama	88
Gambar 4. 17 Menu <i>Data Info</i>	89
Gambar 4. 18 Menu <i>Prediction</i>	89
Gambar 4. 19 Menu <i>Visualization</i>	90
Gambar 4. 20 Menu <i>Contact Us</i>	90
Gambar 4. 21 Menu FAQ.....	91
Gambar 4. 22 Menu <i>Chatbot</i>	91
Gambar 4. 23 Halaman Utama	97
Gambar 4. 24 Menu <i>Prediction</i>	99
Gambar 4. 25 Menu <i>Visualization</i>	99
Gambar 4. 26 Menu <i>Contact Us</i>	100
Gambar 4. 27 Menu FAQ.....	100
Gambar 4. 28 Menu <i>Chatbot</i>	101

DAFTAR TABEL

Tabel 2. 1 Diagram Kasus Pengguna.....	23
Tabel 2. 2 <i>Activity Diagram</i> (Diagram Aktivitas)	25
Tabel 2. 3 <i>Sequence Diagram</i> (Diagram Utama).....	27
Tabel 2. 4 <i>Class Diagram</i> (Diagram Kelas)	30
Tabel 3. 1 Definisi Operational Variable	41
Tabel 3. 2 <i>Missing Value</i> pada Tabel	48
Tabel 3. 3 Nilai 0 Diganti dengan NaN	49
Tabel 4. 1 Fitur Variabel.....	79
Tabel 4. 2 <i>Confussion Matrix</i> dengan Normalisasi.....	85
Tabel 4. 3 <i>Confussion Matrix</i> Tanpa Normalisasi.....	86
Tabel 4. 4 Pengujian <i>Alpha</i>	92
Tabel 4. 5 <i>Confussion Matrix</i> dengan Normalisasi.....	96
Tabel 4. 6 <i>Confussion Matrix</i> Tanpa Normalisasi.....	96
Tabel 4. 7 Pengujian <i>Alpha</i>	102