

**IMPLEMENTASI ALGORITMA REGRESI LOGISTIK
MENGUNAKAN *MACHINE LEARNING* UNTUK
MEMPREDIKSI DIABETES**

SKRIPSI

NAMA: ABI YUSUP

NIM: 20200050028

NAMA: INDRA MAAJID ADITAMA

NIM: 20200050032

NAMA: RISA NURSANIA

NIM: 20200050039



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK KOMPUTER DAN DESAIN
UNIVERSITAS NUSA PUTRA SUKABUMI**

2024

**IMPLEMENTASI ALGORITMA REGRESI LOGISTIK
MENGUNAKAN *MACHINE LEARNING* UNTUK
MEMPREDIKSI DIABETES**

SKRIPSI

*Diajukan Untuk Memenuhi Salah Satu Syarat Dalam Menempuh Gelar Sarjana
Komputer*

Nama: Abi Yusup

NIM: 20200050028

Nama: Indra Maajid Aditama

NIM: 20200050032

Nama: Risa Nursania

NIM: 20200050039



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK KOMPUTER DAN DESAIN
UNIVERSITAS NUSA PUTRA SUKABUMI**

2024

PERNYATAAN PENULIS

JUDUL: IMPLEMENTASI ALGORITMA REGRESI LOGISTIK
MENGUNAKAN *MACHINE LEARNING* UNTUK MEMPREDIKSI
DIABETES

Nama: Abi Yusup

NIM: 20200050028

Nama: Indra Maajid Aditama

NIM: 20200050032

Nama: Risa Nursania

NIM: 20200050039

“Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut”

Sukabumi,....Juli 2024

Mahasiswa

Mahasiswa

Mahasiswa



Abi Yusup

Penulis I

Indra Maajid Aditama

Penulis II



Risa Nursania

Penulis III

PENGESAHAN SKRIPSI

JUDUL : IMPLEMENTASI ALGORITMA REGRESI LOGISTIK
MENGUNAKAN PENDEKATAN *MACHINE LEARNING*
UNTUK MEMPREDIKSI DIABETES

NAMA	:	Abi Yusup	NIM	:	20200050028
NAMA	:	Indra Maajid Aditama	NIM	:	20200050032
NAMA	:	Risa Nursania	NIM	:	20200050039

Skripsi ini telah diujikan dan di pertahankan di depan Dewan Penguji pada Sidang Skripsi tanggal Juli 2024 menurut pandangan kami, Skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan gelar Sarjana Komputer (S.Kom).

Sukabumi, Juli 2024

Pembimbing I

Falentino Sembiring, M.Kom
NIDN. 0408029102

Pembimbing II

Rieska Rahayu Ayuningsih, ST., M.Kom
NIDN. 0407058603

Ketua Penguji

Arny Lattu, S.Pd.Kom., M.Kom
NIDN. 0424089206

Ketua Program Studi Sistem Informasi



Adhittia Erfina, ST., M.Kom
NIDN. 0417049102

Plh. Dekan Fakultas Teknik Komputer dan Desain

Ir. Paikun, ST., M.T., IPM., ASEAN Eng
NIDN. 0402037401

ABSTRAK

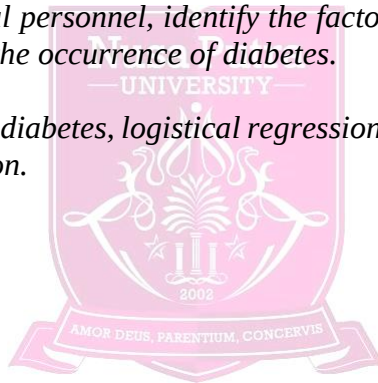
Penyakit diabetes perlu diprediksi dengan akurat karena penyakit diabetes merupakan penyakit sosial yang serius dan bisa terkena orang dalam jumlah besar, serta menyebabkan komplikasi dan melibatkan biaya yang tinggi serta dapat meningkatkan keadaan sakit melalui penyakit diabetes terutama pada anak-anak dan anak muda. Oleh karena itu, untuk mengurangi efek berbahaya diabetes pada orang dan masyarakat secara keseluruhan, pemahaman yang lebih lengkap tentang faktor risiko, pencegahan, dan pengobatan diabetes sangat penting. Tujuan dari penelitian ini adalah untuk mengintegrasikan algoritma regresi logistik untuk melakukan prediksi diabetes atau tidak pada data yang diperoleh dari Kaggle.com yang memiliki *dataset* sejumlah 768 data. Berdasarkan penelitian-penelitian sebelumnya, maka penulis menggunakan algoritma regresi logistik untuk membuat sistem cerdas yang dapat melakukan prediksi diabetes atau tidak. Dimana datanya sendiri terdiri dari 768 *record* dengan beberapa variabel atau atribut prediktor medis (*Pregnancies*/Kehamilan, *Glucose*/Glukosa, *Blood Pressure*/Tekanan Darah, *Skin Thickness*/Ketebalan Kulit, *Insulin*, *BMI*/Indeks Masa Tubuh, *Diabetes Pedigree Function*/Keturunan, *Age*/Umur dan *Outcome*/Hasil). Hasil dari penelitian ini menunjukkan bahwa menggunakan alat evaluasi seperti *cross-validation* dan *confusion matrix* dapat membuat hasil menjadi cukup baik dan akurat. Untuk *data training* hasilnya mendapat nilai ke akurataannya sebesar 77.1%, dan hasil dari *data testing* menghasilkan akurasi sebesar 74%. Sedangkan hasil akurasi dengan menggunakan *Python* mendapatkan hasil 77,54% untuk hasil akurasi pada *data training*, sedangkan hasil pada *data testing* sebesar 78,57%. Akhir dari tujuan penelitian ini adalah mengembangkan model regresi logistik untuk memprediksi data diabetes serta dapat memberikan kontribusi dan menjadi rekomendasi dalam membantu masyarakat bahkan tenaga medis untuk mengidentifikasi faktor-faktor penyebab munculnya penyakit diabetes sehingga masyarakat dapat mencegah terjadinya penyakit diabetes.

Kata Kunci : Prediksi, Diabetes, algoritma regresi logistik, *cross-validation*, *confusion matrix*, *Python*.

ABSTRACT

Diabetes needs to be predicted accurately because it is a serious social disease that can affect a large number of people, cause complications, involve high costs, and improve the condition of diabetes, especially in children and young people. Therefore, to reduce the harmful effects of diabetes on people and society as a whole, a better understanding of risk factors, prevention, and treatment of diabetes is essential. The aim of this study is to integrate logistical regression algorithms to predict diabetes or not in the data obtained from Kaggle.com, which has a data set of 768. Based on previous research, the authors use logistics regression algorithms to create intelligent systems that can predict or not diabetes. The data itself consists of 768 records with several medical predictor variables or attributes (pregnancies, glucose, blood pressure, skin thickness, insulin, BMI/body time index, diabetes pedigree function, age, and outcome). The results of this study show that using evaluation tools such as cross-validation and a confusion matrix can make the results fairly good and accurate. For training data, the results were 77.1%, and the results from testing data produced 74%. While accuracy results using Python obtained 77.54% for the results of training data, the results on testing data were 78.57%. The ultimate goal of this research is to develop a logistic regression model to predict diabetes data that can contribute and be a recommendation for helping the public, even medical personnel, identify the factors that cause diabetes so that the public can prevent the occurrence of diabetes.

Keywords: *predictions, diabetes, logistical regression algorithms, cross-validation, confusion matrix, Python.*



KATA PENGANTAR

Puji dan syukur kami panjatkan kehadiran Allah SW, Berkat Rahmat-Nya, Kami dapat menyelesaikan penyusunan Skripsi “**IMPLEMENTASI ALGORITMA REGRESI LOGISTIK MENGGUNAKAN MACHINE LEARNING UNTUK MEMPREDIKSI DIABETES**”. Skripsi ini diajukan untuk memenuhi persyaratan memperoleh gelar sarjana untuk mata kuliah skripsi di Fakultas Teknik Komputer Dan Desain Universitas Nusa Putra di Sukabumi.

Sehubungan dengan hal tersebut penulis menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Kurniawan, ST., M.Si., MM., selaku Rektor Universitas Nusa Putra Sukabumi.
2. Prof. Ir. Paikun, S.T., M.T., IPM., ASEAN Eng selaku Dekan Fakultas Teknik Komputer dan Desain Universitas Nusa Putra Sukabumi.
3. Bapak Adhitia Erfina, S.T., M.Kom selaku Ketua Program Studi Sistem Informasi Universitas Nusa Putra Sukabumi.
4. Bapak Falentino Sembiring, S.Kom., M.Kom selaku Dosen Pembimbing I yang telah meluangkan waktu dan tenaga dan juga pikirannya untuk memberikan bimbingan yang sangat amat membantu penulis dalam menyelesaikan Skripsi ini.
5. Ibu Rieska Rahayu Ayuningsih S.T., M.Kom selaku Dosen Pembimbing II yang telah meluangkan waktu dan memberikan instruksi dalam segi sistematika penulisan Skripsi ini.
6. Seluruh jajaran Dosen Program Studi Sistem Informasi Universitas Nusa Putra Sukabumi.
7. Kedua orang tua, keluarga, dan juga Orang terkasih kami yang selalu memberikan motivasi, doa dan juga semangat yang tiada henti kepada kami selama kami menyelesaikan Skripsi ini.
8. Rekan-rekan seperjuangan di jurusan Sistem Informasi Angkatan 2020
9. Himpuna Mahasiswa Sistem Informasi (HMSI) Universitas Nusa Putra

10. Serta semua pihak terkait yang tidak bisa kami sebutkan satu-persatu yang sudah ikut membantu peneliti baik secara langsung maupun tidak dalam menyelesaikan laporan penelitian ini.

Dengan ini semua semoga atas izin Allah SWT semua pihak yang telah membantu proses penyelesaian skripsi ini kebaikannya Allah balas dengan pahala yang berlipat ganda.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu kritik dan saran yang membangun sangat kami butuhkan dari berbagai pihak. Aamiin Yaa Rabbal 'Alamiin.

Sukabumi, 10 Juli 2024



Penulis

PERSETUJUAN SKRIPSI
HALAMAN PERSYARATAN PERSETUJUAN PUBLIKASI
SKRIPSI UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik UNIVERSITAS NUSA PUTRA, kami yang bertanda tangan dibawah ini:

NAMA	:	Abi Yusup	NIM	:	20200050028
NAMA	:	Indra Maajid Aditama	NIM	:	20200050032
NAMA	:	Risa Nursania	NIM	:	20200050039

Program Studi: Sistem Informasi

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Nusa Putra **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti Free Right*)** atas karya ilmiah kami yang berjudul: **"IMPLEMENTASI ALGORITMA REGRESI LOGISTIK MENGGUNAKAN MACHINE LEARNING UNTUK MEMPREDIKSI DIABETES"**. Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Nonesklusif ini Universitas Nusa Putra berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pengkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Dibuat di : Sukabumi

Pada tanggal : Juli 2024

Yang Menyatakan:

Mahasiswa



Abi Yusup

Mahasiswa

" -

Indra Maajid Aditama

Mahasiswa



Risa Nursania

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
PERNYATAAN PENULIS	iii
PENGESAHAN SKRIPSI.....	iv
PERSETUJUAN SKRIPSI.....	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	15
1.1 Latar Belakang	15
1.2 Rumusan Masalah.....	17
1.3 Batasan Masalah	18
1.4 Tujuan Penelitian	18
1.5 Manfaat Penelitian	18
1.6 Sistematika Penulisan	19
BAB II TINJAUAN PUSTAKA.....	20
2.1 Penelitian Terdahulu.....	20
2.2 Landasan Teori	27
2.2.1 Diabetes Milletus (DM).....	27
2.2.2 Regresi Logistik.....	29
2.2.3 <i>Machine Learning</i>	29
2.2.4 <i>Data Mining</i>	30
2.2.5 Analisis Prediksi	30
2.2.6 Bahasa Pemograman <i>Pyhton</i>	31
2.3 Kerangka Berpikir	32
BAB III METODE PENELITIAN	33
3.1 Tahapan Penelitian.....	33
3.2 Pengambilan Data.....	34

3.3 Populasi dan Sampel.....	36
3.3.1 Definisi Populasi dan Sampel.....	36
3.3.2 Populasi dalam <i>Dataset</i>	36
3.3.3 Contoh Sampel dari <i>Dataset</i>	37
3.4 Metode Analisis Data	38
3.5 <i>Orange Data Mining</i>	39
3.6 <i>Preprocessing Data</i>	40
3.6.1 <i>Cleaning Data</i>	40
3.6.2 <i>Discretization Data</i>	40
3.6.3 <i>Normalization Data</i>	41
BAB IV HASIL DAN PEMBAHASAN.....	44
4.1 Hasil.....	44
4.1.1 Pengolahan Data.....	44
4.1.2 Algoritma Regresi Logistik.....	46
4.1.3 Hasil Data	47
4.2 Pembahasan	58
BAB V PENUTUP.....	60
5.1 Kesimpulan.....	60
5.2 Saran	60
DAFTAR PUSTAKA.....	62
CURRICULUM VITAE.....	65



DAFTAR TABEL

Tabel 2. 1 Tabel <i>Research GAP</i>	22
Tabel 3. 2 Variabel <i>Dataset</i>	32
Tabel 3. 3 Contoh Hasil Sampel.....	35



DAFTAR GAMBAR

Gambar 1.1. Prevalensi Diabetes Mellitus Indonesia (2018 dan 2023)	14
Gambar 2.2 Kerangka Berpikir	29
Gambar 3.3 Model Tahapan Penelitian	30
Gambar 3.4 <i>Flowchart</i> Pengunduhan <i>Dataset</i> Dari <i>Kaggle.com</i>	32
Gambar 3.5 Contoh Hasil Data Sampel Pada <i>Python</i>	35
Gambar 3.6 Data Setelah Proses Diskretisasi Pada <i>Orange</i>	38
Gambar 3.7 Data Setelah Proses Diskretisasi Pada <i>Python</i>	38
Gambar 3.8 Data Sebelum Proses Normalisasi Pada <i>Orange</i>	39
Gambar 3.9 Data Sesudah Proses Normalisasi Pada <i>Orange</i>	39
Gambar 3.10 Contoh Hasil Data Normalisasi <i>Min-Max Scaling</i>	39
Gambar 3.11 Contoh Hasil Data Normalisasi <i>Decimal Scaling</i>	40
Gambar 4.12 <i>Workflow</i> Penelitian Pada <i>Orange</i>	42
Gambar 4.13 <i>Source Code</i> Untuk <i>Impor Library</i>	42
Gambar 4.14 <i>Widget</i> Algoritma Regresi Logistik	43
Gambar 4.15 <i>Source Code</i> Model Regresi Logistik.....	44
Gambar 4.16 Hasil Data <i>Test And Score Data Training</i>	44
Gambar 4.17 Hasil data test and score data testing	45
Gambar 4.18 <i>Source Code Test And Score Data Training</i>	46
Gambar 4.19 Hasil <i>Test And Score Data Training</i> Pada <i>Python</i>	46
Gambar 4.20 <i>Source Code Test And Score Data Testing</i>	46
Gambar 4.21 Hasil <i>Test And Score Data Testing</i> Pada <i>Python</i>	46
Gambar 4.22 <i>Confusion Matrix Data Training</i>	47
Gambar 4.23 <i>Confusion Matrix Data Testing</i>	48
Gambar 4.24 <i>Source Code Confusion Matrix Display Data Training</i>	49
Gambar 4.25 <i>Confusion Matrix Data Training</i> Pada <i>Python</i>	49
Gambar 4.26 <i>Source Code Confusion Matrix Display Data Testing</i>	51

Gambar 4.27 <i>Confusion Matrix Data Testing Pada Python</i>	51
Gambar 4.28 Hasil Prediksi <i>Data Training</i>	53
Gambar 4.29 Hasil Prediksi <i>Data Testing</i>	54

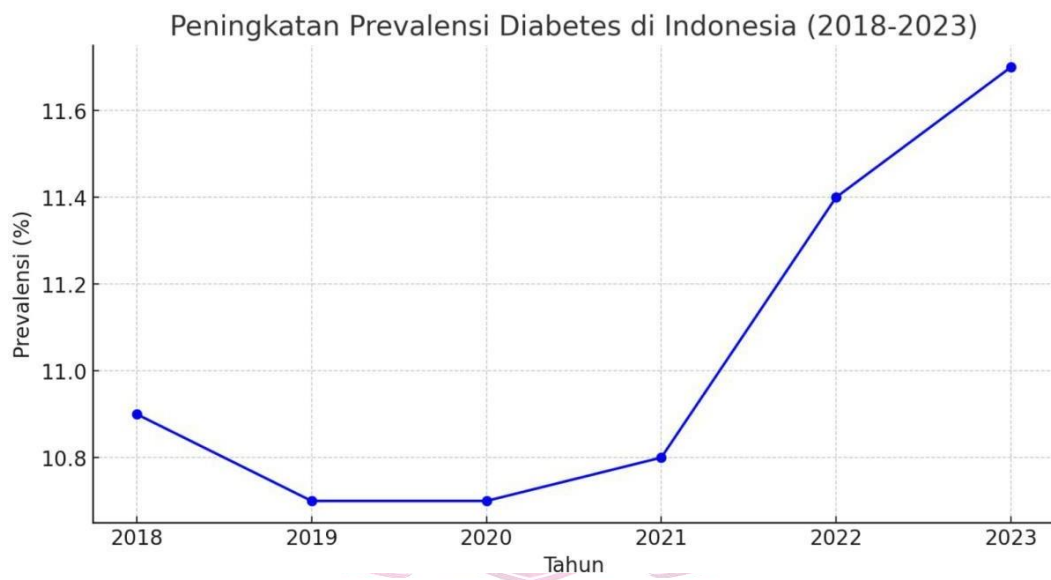


BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes merupakan penyakit kronis yang dapat merenggut nyawa penderita penyakit tersebut. Penyakit ini disebabkan oleh tingginya kadar gula darah pada tubuh manusia. Glukosa merupakan salah satu sumber energi utama bagi sel tubuh manusia. Akan tetapi pada penderita diabetes, glukosa tidak dapat diserap dan tidak dapat digunakan oleh tubuh[1].



Gambar 1.1. Prevalensi *Diabetes Mellitus* Indonesia (2018 Sampai 2023)

Laporan Survei Kesehatan Indonesia (SKI) 2023 yang dikeluarkan Kementerian Kesehatan mengungkap, terjadi peningkatan prevalensi penyakit *diabetes mellitus* (DM) pada penduduk umur di atas 15 tahun berdasarkan hasil pengukuran kadar gula darah[2].

Data dari beberapa tahun terakhir menunjukkan bahwa prevalensi diabetes di Indonesia terus meningkat. Pada tahun 2018, prevalensi diabetes di Indonesia mencapai 10,9%. Angka ini mengalami sedikit penurunan pada tahun 2019 menjadi

10,7%, namun angka tersebut tetap sama pada tahun 2020. Pada tahun 2021, prevalensi kembali meningkat menjadi 10,8%, dan pada tahun 2022 mengalami lonjakan signifikan menjadi 11,4%. Pada tahun 2023, prevalensi diabetes mencapai 11,7%, yang merupakan angka tertinggi dalam lima tahun terakhir[2].

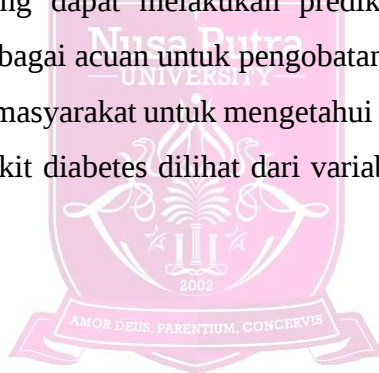
Penyakit diabetes perlu diprediksi dengan akurat karena penyakit diabetes merupakan penyakit sosial yang serius dan bisa terkena orang dalam jumlah besar, serta menyebabkan komplikasi dan melibatkan biaya yang tinggi serta dapat meningkatkan keadaan sakit melalui penyakit diabetes terutama pada anak-anak dan anak muda[3]. Diabetes mellitus dapat mengalami perbaikan jika pasien mendeteksinya dan menjalani pengobatan sejak awal sebelum keparahannya bertambah. Oleh sebab itu dari beban yang ditimbulkan cukup banyak maka penyakit diabetes penting untuk diteliti.

Data Mining deteksi dini penyakit diabetes perlu dilakukan sebagai upaya dalam menurunkan tingkat kematian yang diakibatkan oleh faktor penyakit tersebut[4]. Diabetes dapat menyerang siapa saja tanpa mengenal usia baik lansia, orang dewasa, maupun anak-anak yang ditandai dengan meningkatnya kadar gula (glukosa) darah dalam tubuh manusia[5]. Diabetes dapat disebabkan oleh banyak faktor seperti tekanan darah tinggi, kadar gula berlebih, berat badan, riwayat keturunan diabetes, usia, jumlah kehamilan seseorang, ketebalan lipatan kulit, jumlah kadar insulin dalam tubuh, kurangnya aktivitas fisik dan pola hidup, serta diet tidak sehat[1]. Faktor-faktor tersebut merupakan variabel yang digunakan dalam penelitian ini untuk membuat sistem cerdas yang dapat memprediksi penyakit diabetes. Metode-metode yang digunakan dalam melakukan prediksi yaitu metode *Data Mining*. *Data Mining* merupakan serangkaian tindakan untuk menemukan hubungan dari pola dan kecenderungan dari data yang disimpan, sehingga memungkinkan penerapan algoritma klasifikasi yang efektif untuk mengklasifikasikan tingkat keparahan diabetes[6].

Pada penelitian ini digunakan model regresi logistik biner untuk melihat faktor apa saja yang mempengaruhi seseorang sehingga menderita penyakit Diabetes. Algoritma ini memodelkan hubungan antara *variable dependen* (tingkat keparahan

diabetes) dan *variable independen* (faktor-faktor yang berhubungan dengan diabetes) menggunakan fungsi logistik. Regresi logistik telah terbukti berhasil dalam berbagai aplikasi klasifikasi, termasuk dalam bidang kesehatan. *Dataset* yang dijadikan sebagai objek penelitian yaitu dataset yang berasal dari *database* kesehatan diabetes. *Dataset* yang dapat diakses melalui <https://www.kaggle.com/uciml/pima-indians-diabetes-database>. Dimana datanya sendiri terdiri dari 768 record dengan beberapa variabel atau atribut prediktor medis (*Pregnancies*/Kehamilan, *Glucose*/Glukosa, *Blood Pressure*/Tekanan Darah, *Skin Thickness*/Ketebalan Kulit, *Insulin*, *BMI*/Indeks Masa Tubuh, *Diabetes Pedigree Function*/Keturunan, *Age*/Umur dan *Outcome*/Hasil).

Tujuan dari penelitian ini adalah untuk mengintegrasikan algoritma regresi logistik untuk melakukan prediksi diabetes atau tidak. Berdasarkan penelitian-penelitian sebelumnya, maka penulis menggunakan algoritma regresi logistik untuk membuat sistem cerdas yang dapat melakukan prediksi diabetes atau tidak, sehingga dapat digunakan sebagai acuan untuk pengobatan penderita diabetes bagi dokter di rumah sakit dan di masyarakat untuk mengetahui cara menjaga pola hidup dan cara menghindari penyakit diabetes dilihat dari variabel yang mempengaruhi terjadinya penyakit.



1.2 Rumusan Masalah

Berikut Rumusan Masalah yang peneliti tetapkan pada penelitian ini:

1. Bagaimana kinerja algoritma regresi logistik dalam memprediksi kemungkinan terjadinya diabetes berdasarkan faktor-faktor risiko yang ada?
2. Faktor-faktor apa saja yang signifikan dalam prediksi diabetes?
3. Bagaimana hasil prediksi penyakit diabetes dengan menggunakan pendekatan *machine learning*?

1.3 Batasan Masalah

Berdasarkan rumusan masalah yang telah ditulis, maka di identifikasikan bahwa batasan masalah akan membahas sebagai berikut:

1. Penelitian ini difokuskan untuk memprediksi penyakit diabetes dengan menggunakan *machine learning*.
2. Penelitian ini menggunakan *dataset* diabetes berupa angka dengan jumlah 768 populasi.
3. Penelitian ini menggunakan pendekatan *machine learning* dengan menggunakan model regresi logistik.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini ialah sebagai berikut:

1. Untuk mengetahui bagaimana kinerja algoritma regresi logistik dalam memprediksi kemungkinan terjadinya diabetes berdasarkan faktor-faktor risiko yang diberikan?
2. Mengidentifikasi dan menganalisis faktor-faktor risiko yang signifikan terhadap penyakit diabetes menggunakan metode regresi logistik.
3. Membangun model prediksi yang dapat mengestimasi kemungkinan seseorang menderita diabetes berdasarkan faktor-faktor risiko yang diidentifikasi.
4. Memberikan rekomendasi berdasarkan hasil analisis untuk meningkatkan pencegahan dan pengelolaan penyakit diabetes.

1.5 Manfaat Penelitian

Berikut beberapa manfaat yang bisa diambil dari hasil penelitian ini:

- 1) Bagi Penulis

Penulis berharap dari penelitian ini dapat menambah pemahaman serta wawasan ilmu yang telah diperoleh semasa perkuliahan, khususnya pada ilmu analisis data.

2) Bagi Instansi

Penulis berharap dari penelitian ini dapat menambah pengembangan ilmu penelitian bagi universitas nusa putra serta dijadikan sebagai bahan bacaan di bidang Sistem Informasi yang dapat membantu dalam proses belajar.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi teori-teori terkait pengertian dan penjelasan yang diambil dari penelitian terdahulu yang berhubungan dengan *machine learning* menggunakan algoritma regresi logistik.

BAB III METODOLOGI PENELITIAN

Bab ini berisi metode penelitian yang digunakan, pengumpulan data, analisis data, teknik, langkah-langkah penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil analisis data dari penelitian yang dilakukan dalam bentuk data ilmiah. Diperoleh dari hasil *machine learning* dengan menggunakan algoritma regresi logistik.

BAB V PENUTUPAN

Bab ini berisi Kesimpulan dan saran yang berkaitan dengan hasil dari analisis prediksi penyakit diabetes.



BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan pada penelitian yang telah diteliti menunjukkan hasil dari data yang telah diprediksi pada *dataset* penyakit diabetes dengan menggunakan *Orange Data Mining* dengan metode Logistik Regresi dapat disimpulkan bahwa pada *Orange* berhasil diaplikasikan. Hasil dari penelitian ini menunjukkan bahwa menggunakan alat evaluasi seperti *cross-validation* dan *confusion matrix* dapat membuat hasil menjadi cukup baik dan akurat. Untuk *data training* hasilnya mendapat nilai ke akuratanya sebesar 77.1%, dan hasil dari *data testing* menghasilkan akurasi sebesar 74%. Sedangkan hasil akurasi dengan menggunakan *Python* mendapatkan hasil 77,54% untuk hasil akurasi pada *data training*, sedangkan hasil pada *data testing* sebesar 78,57%.

Pada penelitian ini dapat menunjukkan bahwa hasil akurasi dari analisis data prediksi menggunakan bahasa pemrograman *Python* lebih akurat daripada menggunakan *Orange Data Mining*. Selain itu juga peneliti dapat membangun model prediksi yang dapat mengestimasi kemungkinan seseorang positif diabetes atau negatif diabetes. Akhir dari tujuan penelitian ini adalah mengembangkan model regresi logistik untuk memprediksi data diabetes serta dapat memberikan kontribusi dan menjadi rekomendasi dalam membantu masyarakat bahkan tenaga medis untuk mengidentifikasi faktor-faktor penyebab munculnya penyakit diabetes sehingga masyarakat dapat mencegah terjadinya penyakit diabetes.

5.2 Saran

Untuk mendapatkan hasil yang lebih baik lagi dalam penelitian selanjutnya, agar dapat menggunakan metode yang lain dengan algoritma yang digunakan pada penelitian ini atau menggunakan algoritma lainnya untuk perbandingan hasil akurasi mana yang lebih akurat. Peneliti mengusulkan pengembangan penelitian

lebih lanjut, agar membantu para peneliti selanjutnya untuk mengembangkan dan menganalisis data prediksi dengan menggunakan metode lainnya seperti algoritma *Support Vector Machine* (SVM), algoritma *Naive Bayes*, algoritma C 4.5, algoritma *K-Means* untuk meningkatkan pengalaman pengguna dalam memprediksi data diabetes.



DAFTAR PUSTAKA

- [1] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, “Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM),” *Informatics Digit. Expert*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.
- [2] Santika Erlina F., “Prevalensi Diabetes Indonesia Naik Jadi 11,7% pada 2023.” Accessed: Jul. 05, 2024. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2024/07/05/prevalensi-diabetes-indonesia-naik-jadi-117-pada-2023>
- [3] A. Fauzi and A. H. Yunial, “Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K – Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 3, p. 470, 2022, doi: 10.26418/jp.v8i3.56656.
- [4] N. Fitria, M. Andela, Y. Z. Syaputri, and H. Nasif, “Analisis Efektivitas Biaya Penggunaan Metformin-Glimepiride Terhadap Penurunan Kadar Gula Darah Puasa Pada Pasien Diabetes Mellitus Tipe 2 di RS Universitas Andalas,” *J. Sains Farm. Klin.*, vol. 9, no. sup, p. 202, 2023, doi: 10.25077/jsfk.9.sup.202-207.2022.
- [5] N. Ulya, A. Z. E. Sibuea, S. S. Purba, A. I. Maharani, and C. K. Herbawani, “Analisis Faktor Risiko Diabetes Pada Remaja Di Indonesia,” *J. Kesehat. Tambusai*, vol. 4, no. 3, pp. 2332–2341, 2023, doi: 10.31004/jkt.v4i3.16210.
- [6] H. Apriyani and K. Kurniati, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus,” *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issue3.year2020.page133-143.
- [7] Q. R. Cahyani *et al.*, “Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction using Logistic Regression Algorithm Article Info ABSTRAK,” *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 2, pp. 2828–9099, 2022, doi: 10.55123/jomlai.v1i2.598.
- [8] R. Tyasnurita and A. Y. M. Pamungkas, “Deteksi Diabetik Retinopati menggunakan Regresi Logistik,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 130–135, 2020, doi: 10.33096/ilkom.v12i2.578.130-135.
- [9] D. Diana Dewi, N. Qisthi, S. S. S. Lestari, and Z. H. S. Putri, “Perbandingan Metode Neural Network Dan Support Vector Machine Dalam Klasifikasi Diagnosa Penyakit Diabetes,” *Cerdika J. Ilm. Indones.*, vol. 3, no. 09, pp. 828–839, 2023, doi: 10.59141/cerdika.v3i09.662.
- [10] Dewi Nasien *et al.*, “Perbandingan Implementasi Machine Learning

- Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes,” *JEKIN - J. Tek. Inform.*, vol. 4, no. 1, pp. 10–17, 2024, doi: 10.58794/jekin.v4i1.640.
- [11] A. M. Argina, “Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [12] M. R. Hunafa and A. Hermawan, “Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Imbalance Class Dataset Penyakit Diabetes,” *Media Online*, vol. 4, no. 3, pp. 1551–1561, 2023, doi: 10.30865/klik.v4i3.1486.
- [13] F. Handayanna, “Diabetes Mellitus Dengan Metode Support Vector Machine Berbasis Particle Swarm,” *J. Tek. Inform. STMIK Antar Bangsa*, vol. II, no. 1, pp. 30–37, 2016.
- [14] D. Hardianto, “Telaah Komprehensif Diabetes Melitus: Klasifikasi, Gejala, Diagnosis, Pencegahan, Dan Pengobatan,” *J. Bioteknol. Biosains Indones.*, vol. 7, no. 2, pp. 304–317, 2021, doi: 10.29122/jbbi.v7i2.4209.
- [15] E. Oktavia, A. Id Hadiana, and F. Rahmat Umbara, “Penerapan Metode Regresi Logistik Dalam Prediksi Risiko Diabetes Melitus Gestasional,” *J. Informatics Commun. Technol.*, vol. 5, no. 2, pp. 177–185, 2023.
- [16] I. R. Hikmah and R. N. Yasa, “Perbandingan Hasil Prediksi Diagnosis pada Indian Liver Patient Dataset (ILPD) dengan Teknik Supervised Learning Menggunakan Software Orange,” *J. Telemat.*, vol. 16, no. 2, pp. 69–76, 2021.
- [17] S. O. Pramudya, F. W. Nabili, R. Himawan, and A. P. Sari, “Sistem Pendeteksi Diabetes Menggunakan Algoritma Tsukamoto Pada Bahasa Pemrograman Python,” vol. 3, pp. 53–57, 2023.
- [18] U. M. LEARNING, “Pima Indians Diabetes Database.” [Online]. Available: <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database/code>
- [19] H. Najjichah, A. Sukur, and H. Subagyo, “Pengaruh Text Preprocessing dan Kombinasinya,” *J. Teknol. Inf.*, vol. 15, no. 1, pp. 1–11, 2019, [Online]. Available: [https://download.garuda.kemdikbud.go.id/article.php?article=2887463&val=25402&title=PENGARUH TEXT PREPROCESSING DAN KOMBINASINYA PADA PERINGKAS DOKUMEN OTOMATIS TEKS BERBAHASA INDONESIA](https://download.garuda.kemdikbud.go.id/article.php?article=2887463&val=25402&title=PENGARUH%20TEXT%20PREPROCESSING%20DAN%20KOMBINASINYA%20PADA%20PERINGKAS%20DOKUMEN%20OTOMATIS%20TEKS%20BERBAHASA%20INDONESIA)
- [20] V. C. Wangikar and R. R. Deshmukh, “Data Cleaning: Current Approaches and Issues,” *IEEE Data Eng. Bull.*, vol. 23, no. 4, pp. 3–13, 2015.
- [21] S. Ramírez-Gallego *et al.*, “Data discretization: Taxonomy and big data challenge,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 6, no. 1, pp. 5–21, 2016, doi: 10.1002/widm.1173.

- [22] I. Permana, “The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation,” *Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022, [Online]. Available: <https://media.neliti.com/media/publications/485639-pengaruh-normalisasi-data-terhadap-perfo-e19e3a00.pdf>
- [23] L. 2023, “No Titleการบริหารจัดการการบรรการหม ุณภาพใน
โรงพยาบาลสังกัดกระทรวงสาธารณสุข,”
ค



CURRICULUM VITAE

วารสารวช



CURRICULUM VITAE

าการมหาวิทยาลัย



CURRICULUM VITAE

อัสหรี น



CURRICULUM VITAE

ເອເຊຍ



CURRICULUM VITAE

, vol. 4, no. 1, pp. 88–100, 2023.

