

RANCANG BANGUN APLIKASI PENDETEKSI PENYAKIT DIABETES SECARA DINI MENGGUNAKAN ALGORITMA BINARY LOGISTIC REGRESSION

¹*Rezky Maulana*

²*Dr. Anggraeni Ridwan, S.Kom., MMSI*

.

¹*rezkymaulana0905@gmail.com*

²*neni_rid@staff.gunadarma.ac.id*

*Jurusan Teknik Informatika, Fakultas Teknologi Industri,
Universitas Gunadarma*

.

ABSTRAK

Diabetes adalah masalah kesehatan masyarakat yang penting, menjadi salah satu dari empat penyakit tidak menular prioritas yang menjadi target tindak lanjut oleh para pemimpin dunia. WHO memperkirakan bahwa, secara global, 422 juta orang dewasa berusia di atas 18 tahun hidup dengan diabetes pada tahun 2014. Jumlah terbesar orang dengan diabetes diperkirakan berasal dari Asia Tenggara dan Pasifik Barat, terhitung sekitar setengah kasus diabetes di dunia. Proses untuk mendeteksi penyakit diabetes umumnya masih menggunakan cara manual yaitu dengan melakukan konsultasi secara langsung ke dokter spesialis penyakit diabetes atau dokter penyakit dalam dan melakukan beberapa pemeriksaan laboratorium yang kemudian harus dikonsultasikan lagi dengan dokter spesialis penyakit diabetes. Selama proses pemeriksaan yang dilakukan oleh laboratorium memerlukan waktu yang cukup lama untuk menyimpulkan hasil diagnosa secara jelas. Aplikasi Pendeteksi Penyakit Diabetes Menggunakan Algoritma *Binary Logistic Regression* ini bertujuan untuk membantu pihak laboratorium untuk melakukan diagnosa secara efisien dan cepat untuk memberikan hasil kepada pasien. Aplikasi Pendeteksi Penyakit Diabetes Secara Dini Menggunakan Algoritma *Binary Logistic Regression* ini dibuat dengan menggunakan aplikasi Jupyter Notebook dengan menerapkan bahasa pemrograman Python dan diimplementasikan kedalam GUI.

Kata Kunci: Penyakit Diabetes, Klasifikasi, Algoritma Binary Logistic Regression, Aplikasi, Deteksi.

**DESIGN AND BUILD APPLICATION OF DIABETES DETECTION
EARLY BY USING BINARY LOGISTIC
REGRESSION ALGORITHM**

ABSTRACT

Diabetes is an important public health problem, being one of the four priority non-communicable diseases that are being targeted by world leaders. WHO estimates that, globally, 422 million adults aged over 18 years lived with diabetes in 2014. The largest number of people with diabetes is estimated to come from Southeast Asia and the Western Pacific, accounting for about half the cases of diabetes in the world. The process for detecting diabetes generally still uses the manual method by consulting directly with a diabetes specialist or internist and conducting several laboratory examinations which then must be consulted again with a diabetes specialist. During the examination process carried out by the laboratory requires a long time to conclude the diagnosis clearly. Diabetes Detection Application Using Binary Logistic Regression Algorithm aims to help the laboratory to diagnose efficiently and quickly to provide results to patients. Diabetes Detection Application Using Binary Logistic Regression Algorithm was made using the Jupyter Notebook application by implementing the Python programming language and implemented into the GUI.

Keyword : *Diabetes illness, Classification, Binary Logistic Regression Algorithm, Application, Detection.*

PENDAHULUAN

Diabetes adalah penyakit kronis serius yang terjadi karena pankreas tidak menghasilkan cukup insulin (hormon yang mengatur gula darah atau glukosa), atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dihasilkannya. Diabetes adalah masalah kesehatan masyarakat yang penting, menjadi salah satu dari empat penyakit tidak menular prioritas yang menjadi target tindak lanjut oleh para pemimpin dunia. Jumlah kasus dan prevalensi diabetes terus meningkat selama beberapa dekade terakhir. Diabetes menyebabkan 1,5 juta kematian pada tahun 2012. Gula darah yang lebih tinggi dari batas maksimum mengakibatkan tambahan 2,2 juta kematian, dengan meningkatkan risiko penyakit kardiovaskular dan lainnya.

Di seluruh dunia, jumlah penderita diabetes telah meningkat secara substansial antara tahun 1980 dan 2014, meningkat dari 108 juta menjadi 422 juta atau sekitar empat kali lipat. Oleh karena itu penyakit diabetes ini meningkat dari secara signifikan, maka diperlukan pencegahan, mengobati dan memantau

secara berkala sepanjang hidup. Sehingga penyakit diabetes dapat dicegah tingkat keparahannya jika terpantau sejak dini.

Proses untuk mendeteksi penyakit diabetes umumnya masih menggunakan cara manual yaitu dengan melakukan konsultasi secara langsung ke dokter spesialis penyakit diabetes atau dokter penyakit dalam dan melakukan beberapa pemeriksaan laboratorium yang kemudian harus dikonsultasikan lagi dengan dokter spesialis penyakit diabetes. Maka, dengan tersedianya teknologi komputer yang semakin berkembang dalam berbagai bidang termasuk dalam mendiagnosa penyakit, seharusnya teknologi komputer bisa di implementasikan untuk memberikan hasil diagnosa awal secara cepat dan efisien.

Penelitian dalam mendiagnosa penyakit dengan menggunakan teknologi komputer semakin meningkat. Salah satu peran Teknologi Informasi dalam pendeteksian dini penyakit diabetes sudah dilakukan dengan teknologi data mining. Data mining dengan algoritma cerdas dapat digunakan untuk mengatasi masalah diagnosis dengan dataset medis yang melibatkan beberapa inputan data.

Data mining dalam dunia medis memiliki potensi besar untuk mengetahui pola tersembunyi pada dataset medis. Pola tersebut dapat dimanfaatkan untuk mengungkapkan ada atau tidak ada suatu penyakit. Implementasi yang akurat dan efisien dibutuhkan teknik data mining yang tepat. Data mengenai faktor-faktor diagnosis penyakit diabetes dapat diakses secara luas dan bebas pada situs Kaggle atau *UCI Machine Learning*. Sehingga peluang riset untuk mendeteksi dini penyakit diabetes sangat luas.

Maka, dengan tersedianya teknologi komputer yang semakin berkembang dalam berbagai bidang termasuk dalam mendiagnosa penyakit, seharusnya teknologi komputer bisa diimplementasikan untuk memberikan hasil diagnosa awal secara cepat dan efisien dengan menerapkan proses *Data Mining*.

Dalam penelitian ini menggunakan algoritma *Logistic Regression Binary* dengan data Diabetes yang akan dijadikan sebagai data training dan data testing, kemudian akan melakukan prediksi, sehingga akan mendapatkan informasi mengenai diagnosa penyakit diabetes. Hasil prediksi awal yang didapatkan

dapat digunakan oleh para petugas medis sebagai alat bantu dalam penentuan penyakit diabetes atau bisa disebut juga sebagai kesimpulan awal dalam mendiagnosa penyakit diabetes sebelum di tindak lanjuti. Dengan mekanisme tersebut, proses dalam mendiagnosa penyakit diabetes bisa menjadi lebih cepat dan efisien dalam memberikan hasil kepada pasien.

Dalam penelitian ini menggunakan algoritma Logistik Regresi Biner dengan data Diabetes yang akan dijadikan sebagai data training dan data testing, kemudian akan melakukan prediksi, sehingga akan mendapatkan informasi mengenai diagnosa penyakit diabetes. Hasil prediksi awal yang didapatkan dapat digunakan oleh para petugas medis sebagai alat bantu dalam penentuan penyakit diabetes atau bisa disebut juga sebagai kesimpulan awal dalam mendiagnosa penyakit diabetes sebelum di tindak lanjuti. Dengan mekanisme tersebut, proses dalam mendiagnosa penyakit diabetes bisa menjadi lebih cepat dan efisien dalam memberikan hasil kepada pasien.

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini menggunakan *System Development Cycle* (SDCL) dan model yang digunakan adalah model *Waterfall*. Berikut tahap tahap yang menggambarkan metode SDLC dengan model *Waterfall*.(sommerville, 2011)

1. *Requirements analysis and definition*, mengumpulkan apa yang dibutuhkan secara lengkap untuk kemudian dianalisis guna mendefinisikan kebutuhan yang harus dipenuhi oleh program yang akan dibangun. Fase ini harus dikerjakan dengan lengkap untuk menghasilkan desain yang lengkap.
2. *System and software design*, setelah apa yang dibutuhkan telah selesai dikumpulkan dan sudah lengkap maka desain kemudian di kerjakan.
3. *Implementation and unit testing*, desain program diterjemahkan dalam kode-kode dengan menggunakan bahasa

pemerograman yang sudah di tentukan. Program yang dibangun langsung diuji secara unit, apakah sudah bekerja dengan baik.

4. *Integration and system testing*, penyatuan unit-unit program untuk kemudian di uji secara keseluruhan (*system testing*).
5. *Operating and maintenance*, mengoperasikan program dilingkungannya dan melakukan pemeliharaan, seperti penyesuaian atau perubahan untuk adaptasi dengan situasi yang sebenarnya.

PEMBAHASAN

Tujuan yang ingin dicapai dalam penelitian ini adalah membangun aplikasi klasifikasi penyakit diabetes dengan menggunakan algoritma Logistik Regresi Biner yang dapat menunjukkan nilai akurasi prediksi penyakit diabetes dan mengimplementasikan-nya kedalam bentuk GUI dan Dapat digunakan oleh pihak Rumah Sakit sebagai alat bantu untuk mendiagnosa pasien dalam hal diagnose penyakit diabetes secara dini.

A. DATASET

Penelitian ini menggunakan dataset mengenai penyakit diabetes yang diperoleh dari situs *Kaggle* atau *Machine Learning UCI* (Ronit, 2018) sebanyak 768 *record*. Terdiri dari 9 atribut yang akan diberikan skala pengukuran berupa jumlah kehamilan yang pernah di alami (khusus wanita), kadar glukosa dalam darah, Tekanan darah diastolic (mm Hg), Ketebalan kulit triceps, Insulin Berat Badan, Keturunan diabetes dari keluarga, umur, dan hasil prediksi atau outcome.

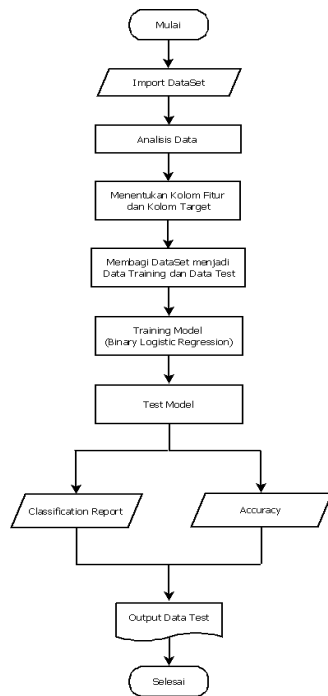
Tabel 1 Deskripsi *dataset*

Label	Makna Label	Singkatan	Nilai
Pregnancies	Jumlah kehamilan yang pernah di alami (khusus wanita)	Pregucies	Integer
Glucose	Kadar Glukosa	Gc	Integer
Diastolic blood pressure (mm Hg)	Tekanan Darah Diastolik (mm Hg)	Dbp	Integer
Triceps skin fold thickness (mm)	Ketebalan lipatan kulit trisep (mm)	Sfold	Integer
Insulin	Insulin serum dalam 2 jam (mu U / ml)	ISul	Integer
BMI	Indeks massa tubuh (berat dalam kg / (tinggi dalam m) ²)	Bmi	Float
Diabetes pedigree function	Keturunan Diabetes	DPedigree	Float
Age (years)	Umur	Age	integer
Outcome	Output penyakit Diabetes	Result	

B. FLOWCHART

Flowchart merupakan penggambaran secara grafik dari langkah-langkah dan urutan prosedur suatu program. Biasanya mempengaruhi penyelesaian masalah yang khususnya perlu dipelajari dan dievaluasi lebih lanjut. Tujuan membuat *flowchart* adalah menggambarkan suatu tahapan penyelesaian masalah secara sederhana, terurai, rapi dan jelas menggunakan simbol - simbol standar (Indrajani, 2011) .

Tahap pertama yang dilakukan adalah mengumpulkan dataset yang terdiri dari data *training* dan data *test*. Setelah mengumpulkan dan analisis dataset, kemudian dilakukan pembagian atribut. Atribut terbagi menjadi dua, yaitu terdapat fitur dan target.



Gambar 1 Flowchart Program

C. ALGORITMA *BINARY*

LOGISTIC REGRESSION

Binary Logistic Regression adalah sebuah pendekatan untuk membuat model prediksi seperti halnya regresi linear atau yang biasa disebut dengan istilah Ordinary Least Squares (OLS) regression. Perbedaannya adalah pada *Binary Logistic Regression*, peneliti memprediksi variabel terikat yang berskala dikotomi. Skala dikotomi yang dimaksud adalah

skala data nominal dengan dua kategori, misalnya: Ya dan Tidak, Baik dan Buruk, Tinggi dan Rendah, atau Cepat dan Lambat. Model regresi logistik biner digunakan jika variabel responnya menghasilkan dua kategori bernilai 0 dan 1 (Anonim,2015).

Regresi logistik biner bisa diartikan sebagai suatu model pendekatan matematika yang digunakan untuk menggambarkan hubungan antara beberapa variabel penjelas dengan suatu variabel dikotomi. Variabel dikotomi mempunyai dua kemungkinan yang biasanya dinyatakan dengan 0 (gagal) dan 1 (sukses) (Kleinbaum,2000).

$$Y = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i$$

Penjelasan :

Y = Dependent Variables

β_1 = Coefficient of Variables x_1

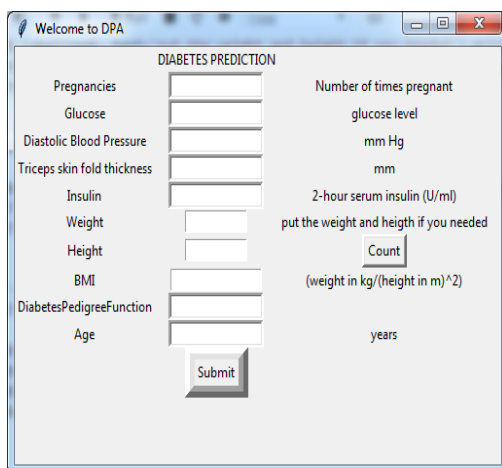
x_1 = Independent Variables

β = Constant

D. IMPLEMENTASI APLIKASI

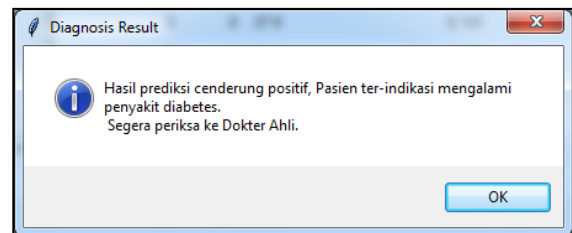
Implementasi ditampilkan hasil dari pembuatan aplikasi berbasis desktop yang telah dirancang dan dianalisa. Total data yang digunakan adalah 614 data latih dan 154 data uji hasil pra-pemrosesan. Uji coba aplikasi bertujuan untuk mengetahui apakah aplikasi sudah dapat berjalan dengan baik atau masih terdapat kesalahan. Tahap pengujian aplikasi dilakukan dengan menggunakan perangkat komputer atau *notebook*.

Tampilan *menu* utama



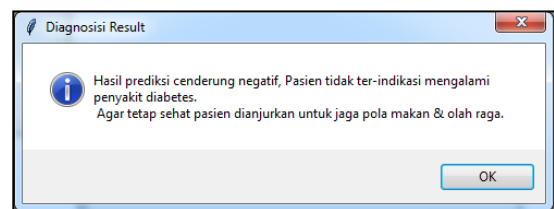
Gambar 3 Menu Utama

Tampilan selanjutnya yaitu tampilan *output* yang menyediakan 1 *button*, yaitu Ok juga tombol x untuk Keluar aplikasi, ditunjukkan pada Gambar 4 yang merupakan tampilan output hasil dari aplikasi prediksi diabetes jika melakukan input data yang terdiagnosa diabetes.



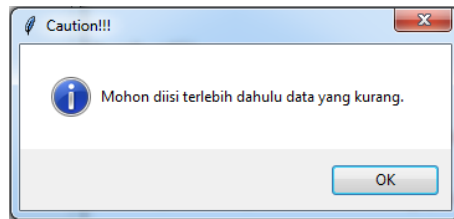
Gambar 4 hasil klasifikasi terdiagnosa

Tampilan selanjutnya Gambar 5 yang merupakan tampilan output hasil dari aplikasi prediksi diabetes jika melakukan input data yang tidak terdiagnosa diabetes.



Gambar 5 hasil klasifikasi tidak terdiagnosa

Tampilan selanjutnya Gambar 6 yang merupakan tampilan form notifikasi jika ada data yang belum diisi



Gambar 6 notifikasi data belum ter-input

Gambar 6 diatas merupakan tampilan output hasil dari aplikasi prediksi diabetes jika ada data input yang kosong.

KESIMPULAN

Berdasarkan keseluruhan uraian penulisan yang penulis telah buat, dapat disimpulkan bahwa pembuatan Pendeteksi Penyakit Diabetes Menggunakan Algoritma *Logistik Regression binary* berbasis GUI telah berhasil dibuat dan informasi mengenai panyakit diabetes berhasil didapatkan menggunakan metode klasifikasi dengan menggunakan algoritma *logistic regression binary*. Informasi yang ada pada aplikasi berupa adanya indikasi penyakit

diabetes atau tidak pada seorang pasien. Dengan dilakukannya pembelajaran atau latih pada data-data yang tersedia menggunakan algoritma *logistic regression binary*, maka didapatkan informasi mengenai akurasi = 81%, error / loss 18%, precision = 84%, recall = 83%, f1-Score = 84%. Maka kesimpulan yang didapat yaitu aplikasi Pendeteksi Penyakit Diabetes Menggunakan Algoritma *Logistik Regression binary* berbasis GUI yang telah dibuat ini dapat melakukan deteksi penyakit diabetes dan mendiagnosa secara dini.

SARAN

Saran untuk pengembangan aplikasi ini yaitu menambah dataset atau data-data yang memiliki keterhubungan dengan dataset yang telah dipelajari dan diproses sebelumnya, karena semakin banyak data yang ditambahkan akan membuat model yang dilatih dapat meningkatkan tingkat akurasi dan mengurangi tingkat *loss*.

Dan menerapkan model klasifikasi yang lainnya seperti *C4.5*, *nearest neighbor*, dan *neural network* untuk membandingkan mana yang lebih baik dalam melakukan klasifikasi dengan menggunakan data yang lebih banyak dan parameter yang lebih dari data sebelumnya. Dan untuk tampilan *User Interface* dibuat lebih menarik.

DAFTAR PUSTAKA

- [1]. Anonim, 2010.
“InfoDiabetes”. <http://www.depkes.go.id/resources/download/pusdatin/infodatin/infodatin-diabetes.pdf>
- [2]. Suyanto, 2017. “Data Mining”. Bandung: Informatika.
- [3]. Anonim, 2015. “Logistik Regression”. <https://www.statistikian.com/2015/02/regresi-logistik.html>.
- [4]. Indrajani, 2011. “Flowchart”. <https://studylibid.com/doc/783897/bab-ii-landasan-teori>.
- [5]. Kleinbaum, David G., & M. Klein. 2010. *Logistic Regression : A Self-Learning Text (3rd edition)*, New York ; John Willey and Sons Inc.
- [6]. Taufiq Luthfi, Emha dan Kusri. 2009. *Algoritma Data Mining*. Yogyakarta: Penerbit ANDI.
- [7]. Anonim . 2009. “Diabetes”. <https://www.suara.com/health/2016/04/07/082912/who-422-juta-orang-dewasa-menderita-diabetes>.
- [8]. Anonim, 2005. “Regresi Logistik”. <https://swanstatistics.com/course/regresi-logistik/>
- [9]. Hosmer, D. W. and S. Lemeshow. 2000. *Applied Logistic Regression Second Edition*. New York: John Willey & Son inc.
- [10]. O'brien, 2000. “SDLC”. <https://timur.ilearning.me/2015/12/27/jelaskan-pengertian-dari-sdlc/>. Tanggal Akses : 29 Juli 2019
- [11]. Barry pratama, 2000. “Aplikasi Basis Data” <https://ilmukomputer.org/category/aplikasi-server/>. Tanggal Akses : 29 Juli 2019
- [12]. Fayyad, 1996. “Tahapan Data Mining” <https://www.kajianpustaka.com>

- /2017/09/ data-mining.html.
- [13]. Jubilee Enterprise, 2016. Trik Cepat Menguasai Python. Bandung : PT Elex Media Komputer.
- [14]. Hossin,M&Sulaiman,M.N,2015.” Optimized Classifier”.
<https://scholar.google.com.my/citations?user=5HyDXWUAAAAJ&hl=en>
 .
- [15]. Jiawei Han and Micheline Kamber, 2006. Data Mining: Concept and Techniques 2nd edition. Illnoiss: Morgan Kaufmann.
- [16]. Budi Raharjo, 2015. Mudah Belajar Python Untuk Aplikasi Deskrip. Jakarta: Informatika.
- [17]. Jorgensen,Paul C,1995. Software Testing: A Craftsman's Approach, Fourth Edition. Unites States of America : Auerbach Publications.
- [18]. Anonim, 2015. “Balsamiq Mockup”.<https://balsamiq.com/learn/>