

**PERANCANGAN SISTEM PENDETEKSI DIABETES
DENGAN MENGGUNAKAN METODE REGRESI**



Diusulkan Oleh :

Andrew C Handoko (672019250)

Gruda Sakti Krida P (672019252)

Arya Damar Pratama (672019227)

**FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN SATYA WACANA
SALATIGA
2021**

DAFTAR ISI

DAFTAR ISI

BAB 1. PENDAHULUAN

BAB 2. TINJAUAN PUSTAKA

BAB 3. METODE PENELITIAN

BAB 4. HASIL DAN PEMBAHASAN

DAFTAR PUSTAKA

Perancangan Sistem Pendeteksi Diabetes Dengan Menggunakan Metode Regresi

Andrew C Handoko, Gruda Sakti Krida P & Arya Damar P

Program Studi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
Jl. Dr. O. Notohamidjojo No.1 - 10, Blotongan, Kec. Sidorejo, Kota Salatiga,
Jawa Tengah 50715, Indonesia
672019250@student.uksw.edu

ABSTRAK

Diabetes merupakan penyakit metabolik yang jumlah penderitanya termasuk banyak di seluruh dunia, bahkan Indonesia sendiri. Dan penyakit ini memiliki jenis diabetes gestasional yang mana menyerang perempuan yang sedang hamil. Sehingga dapat menimbulkan banyak komplikasi yang akan terjadi pada saat kehamilan maupun setelah melahirkan. Maka dari itu deteksi dini atau skrining sangat perlu dilakukan, untuk mencegah dan dapat menurunkan risiko komplikasi bagi perempuan yang sedang hamil serta bagi anaknya juga. Dalam perancangan sistem deteksi ini, digunakan pendekatan *machine learning* dengan metode regresi logistik. Yang nantinya akan melihat pengaruh dari hubungan antar variabel gejala dan outputnya.

Kata Kunci : *Diabetes, Diabetes Gestasional, Machine Learning, Regresi, Regresi Logistik*

ABSTRACT

Diabetes is a metabolic disease whose number of sufferers includes many around the world, even Indonesia itself. And this disease has a type of gestational diabetes which attacks women who are pregnant. So that it can cause many complications that will occur during pregnancy and after delivery. Therefore, early detection or screening is very necessary, to prevent and reduce the risk of complications for women who are pregnant and for their children as well. In designing this detection system, a machine learning approach with logistic regression method is used. Which will later see the effect of the relationship between the symptom variables and their output.

Keyword : *Diabetes, Machine Learning, Regression, Logistic Regression*

PENDAHULUAN

Diabetes merupakan salah satu penyakit yang menjadi ancaman serius terhadap kesehatan global di antara populasi manusia yang semakin menua ini. Dan menurut WHO (Organisasi Kesehatan Dunia) di tahun 2014 menunjukkan bahwa prevalensi diabetes, terutama tipe 2 terus meningkat dalam tiga dekade terakhir, terutama pada Negara dengan penghasilan rendah dan menengah. Sedangkan untuk Indonesia sendiri di tahun 2015 menurut WHO bahwa Indonesia menempati peringkat ke tujuh dunia dimana prevalensi diabetes tertinggi dunia dimana estimasi penderita diabetes sebesar 10 juta orang dari total penderita diabetes di dunia sebesar 415 juta orang. Dimana pada tahun 2017 juga menurut IDF (International Diabetes Federation) Atlas menunjukkan bahwa diabetes di Indonesia cenderung meningkat. Indonesia menempati posisi ke enam di dunia dengan jumlah penderita diabetes di usia 20 – 79 tahun sekitar 10,3 juta orang. Dan pada diabetes gestasional, hampir 80% penderitanya berada di Negara dengan penghasilan rendah dan menengah. Di Negara Eropa diabetes gestasional ini berkembang sebesar 5,4%, di negara Afrika sebesar 14%, di negara-negara Asia sebesar 1% - 20%, sedangkan di Indonesia sendiri prevalensi diabetes gestasional sebesar 1,9% - 3,6%.

Penyakit diabetes yang mana sering disebut sebagai penyakit gula ini merupakan penyakit berbahaya yang mana dapat menyebabkan kematian akibat dari komplikasi yang timbul dari penyakit ini. Diabetes sendiri merupakan penyakit metabolik yang mana tubuh penderita tidak mampu mencukupi kebutuhan insulin secara efektif sehingga gula di dalam darah menjadi berlebih. Dan diabetes pada ibu hamil tiap tahunnya meningkat, dimana diabetes ini disebut diabetes gestasional. Dimana kemudahan dalam skrinning atau deteksi dini terkait diabetes gestasional sangatlah penting untuk mengurangi risiko atau komplikasi yang mungkin terjadi pada ibu maupun bayinya nanti. Sehingga tindakan pencegahan maupun penanganan yang harus dilakukan dapat segera diambil agar dapat melindungi perempuan atau ibu hamil serta mengurangi angka kejadian diabetes gestasional. Dan dapat menurunkan morbiditas untuk ibu dan anak. Dimana diabetes gestasional ini memberikan dampak buruk kepada ibu maupun janin, dimana meningkatkan risiko ibu mengalami hipertensi pada kehamilan

hingga pre-eklamsia. Dan sekitar 50-70% ibu hamil penderita diabetes gestasional juga berisiko menderita diabetes tipe 2 dalam kurun waktu 5-10 tahun setelah melahirkan dan anak-anak yang lahir dari ibu penderita diabetes berisiko mengalami diabetes tipe 2 pada usia dewasa. Sehingga diabetes gestasional ini perlu mendapat perhatian khusus dan serius dalam pengobatan demi melindungi perempuan maupun ibu hamil serta anaknya.

Machine Learning merupakan proses dimana sistem bisa meningkatkan performanya dari pengalaman atau bisa dikatakan sebuah mesin yang dapat belajar dari data-data yang diberikan. Dalam *machine learning* sendiri terdapat jenis-jenisnya, dan yang akan digunakan kali ini adalah *supervised learning*. Dimana dalam *supervised learning* metode yang akan digunakan adalah regresi. Regresi merupakan salah satu metode statistika yang digunakan untuk melihat pola hubungan antara dua variabel atau lebih, yaitu variabel *dependent* dan variabel *independent*. Dan dalam perancangan sistem deteksi ini akan menggunakan regresi logistik yang merupakan bentuk khusus dari analisa regresi yang mendeskripsikan hubungan antar variabel *dependent* yang mempunyai dua atau lebih kategori dengan satu atau lebih variabel *independent* berskala kontinu. Sehingga dipilih metode regresi logistik ini untuk perancangan sistem deteksi diabetes yang nantinya akan dilihat berdasarkan pengaruh beberapa variabel untuk mengklasifikasikan apakah seseorang diabetes atau tidak. Dan pada penelitian ini data yang kami gunakan yaitu “*Diabetes Dataset*” yang dimana *dataset* tersebut didapat dari *National Institute of Diabetes and Digestive and Kidney Diseases* yang dipublikasikan di *website Kaggle*, yang berisi data dari perempuan dengan umur minimal 21 tahun dari “Pima Indian Heritage”.

TUJUAN

Tujuan Perancangan Sistem Pendeteksi Diabetes ini diharapkan dapat mendeteksi penyakit diabetes gestasional pada perempuan terutama ibu hamil, sehingga tindakan penanganan maupun pencegahan dapat dilakukan sebaik mungkin. Dan dapat menurunkan risiko ibu hamil mengalami komplikasi selama proses kehamilan, melahirkan dan setelah melahirkan. Agar perempuan nantinya juga terhindar dari risiko diabetes tipe 2 setelah 5-10 tahun melahirkan dan

menurunkan risiko diabetes tipe 2 kepada anaknya setelah dewasa. Dengan adanya sistem ini diharapkan dapat mengurangi perempuan atau ibu hamil penderita diabetes dengan skrining atau deteksi dini yang dapat dilakukan dengan mudah melalui sistem deteksi diabetes ini.

TINJAUAN PUSTAKA

A. Penelitian Terdahulu

Penelitian dengan judul Deteksi Diabetik Retinopati menggunakan Regresi Logistik membahas tentang rancangan sistem pendeteksi untuk penyakit diabetes terutama diabetes retinopati. Dalam perancangannya digunakan pendekatan *machine learning* dengan metode regresi logistik. Dan dalam pelatihan data pada model sistem deteksi yang dirancang ada empat macam kondisi yaitu dengan parameter bawaan, standarisasi atribut, pemilihan atribut dan pengaturan parameter. Dan hasil yang didapatkan menunjukkan model mempunyai akurasi yang cukup bagus dengan persentase sebesar 80,17% (Raras & Adhi Yoga, 2020).

Pada penelitian dengan judul Analisis Faktor Resiko Penyebab Diabetes Mellitus dengan Regresi Logistik Biner ini membahas mengenai diabetes mellitus yang menjadi salah satu masalah perawatan kesehatan utama di seluruh dunia. Yang mana penyakit berbahaya ini disebabkan oleh berbagai faktor, dan beberapa diantaranya yaitu usia, merokok, serum sodium dan lainnya. Dalam menganalisis faktor-faktor tersebut digunakan regresi logistik untuk melihat pengaruh dari variabel independen usia, merokok, serum sodium dan platelet. Dan didapatkan hasil bahwa semua variabel independen saling berpengaruh di dalam model. Tetapi dalam analisis oleh model ini terjadi misklasifikasi bahwa orang yang menderita diabetes sebagai orang yang tidak mengalami diabetes sebesar 60%. Dan tentunya klasifikasi dengan metode ini agak beresiko dalam mengelompokkan hasil analisa berdasar empat variabel independen tadi jika bertujuan sebagai skrining atau deteksi cepat (I Gusti Bagus & Kartika, 2021).

Lalu pada penelitian berjudul Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression membahas bagaimana mendeteksi penyakit jantung yang merupakan salah satu penyakit berbahaya

secara cepat menggunakan data dari pasien. Dalam pendeteksian penyakit jantung telah terdapat berbagai metode, tetapi pada penelitian ini digunakan metode regresi logistik. Setelah dilakukan pengimplementasian dengan metode ini didapatkan keunggulan yang berbeda terhadap metode lainnya berdasarkan hasil evaluasi dengan *confussion matrix*. Pada *data training* didapatkan akurasi sebesar 88.54% dan pada *data testing* sebesar 87.50% dibanding metode lainnya. Yang mana pada model ini, semakin banyak data akan semakin bagus akurasi dari model karena sistem deteksi yang dirancang memiliki banyak data sebagai acuan pengklasifikasian (Jefri, Henry & Kenichi, 2021).

Berdasarkan penelitian-penelitian terdahulu mengenai rancangan sistem deteksi cepat untuk penyakit dengan metode regresi logistik, maka akan dilakukan penelitian yang membahas tentang rancangan sistem deteksi cepat untuk penyakit diabetes berjenis diabetes gestasional. Model rancangan yang akan dibuat menggunakan metode regresi logistik untuk melihat hubungan dari 8 variabel independen yang dapat menyebabkan diabetes gestasional.

B. Diabetes

Menurut American Diabetes Association (ADA) tahun 2010, Diabetes melitus merupakan suatu kelompok penyakit metabolik dengan karakteristik hiperglikemia yang terjadi karena kelainan sekresi insulin, kerja insulin, atau kedua-duanya.

Klasifikasi etiologi Diabetes melitus menurut American Diabetes Association, 2010 adalah sebagai berikut :

1. Diabetes Melitus tipe 1 adalah penyakit gangguan metabolik yang ditandai oleh kenaikan kadar gula darah akibat destruksi (kerusakan) sel beta pancreas karena suatu sebab tertentu yang menyebabkan produksi insulin tidak ada sama sekali sehingga penderita sangat memerlukan tambahan insulin dari luar.
2. Diabetes Melitus tipe 2 adalah penyakit gangguan metabolik yang ditandai oleh kenaikan kadar gula darah akibat penurunan sekresi insulin oleh sel beta pankreas dan atau fungsi insulin (resistensi insulin)

3. Diabetes Melitus tipe lain adalah penyakit gangguan metabolik yang ditandai oleh kenaikan kadar gula darah akibat defek genetik fungsi sel beta, defek genetik kerja insulin, penyakit eksokrin pankreas, endokrinologi, karena obat atau zat kimia, infeksi, sebab imunologi yang jarang, sindrom genetik lain yang berkaitan dengan DM
4. Diabetes Melitus tipe Gestasional adalah penyakit gangguan metabolik yang ditandai oleh kenaikan kadar gula darah yang terjadi pada wanita hamil, biasanya terjadi pada usia 24 minggu masa kehamilan, dan setelah melahirkan gula darah kembali normal.

B.1. Diabetes Gestasional

Diabetes gestasioanl atau disebut sebagai *Gestasional Diabetes Mellitus* (GDM) adalah diabetes yang terjadi pada saat kehamilan ditandai dengan kadar gula darah yang naik biasanya pada usia 24 minggu masa kehamilan. Dan dalam pemeriksaan untuk mengetahui apakah ibu hamil mengalami GDM dilakukan dengan mengukur tingkat toleransi gula darah plasma setelah 2 jam berpuasa. GDM ini dapat terjadi pada perempuan manapun, yang risikonya akan berbeda-beda tergantung dengan gaya hidup di tiap daerah. Serta perempuan dengan usia lebih dari 35 tahun memiliki risiko lebih tinggi daripada yang hamil di usia lebih muda. Perempuan dengan usia 35 tahun lebih cenderung memiliki kadar gula darah tinggi karena kadar insulin yang diproduksi tubuh makin berkurang. Lalu ada faktor berat badan, yang mana jika Indeks Massa Tubuh masuk ke dalam obesitas, maka insulin sulit bekerja secara maksimal karena kadar lemak darah yang tinggi terutama kolesterol. Serta beberapa faktor lainnya seperti riwayat keturunan keluarga apakah mempunyai riwayat diabetes, faktor dari tekanan darah dan lainnya.

Dan komplikasi yang dapat terjadi pada ibu hamil penderita GDM ada beberapa seperti gangguan penglihatan, pre-eklamsia atau sindrom dengan tanda hipertensi, ukuran janin yang besar, keguguran, persalinan yang lama, premature serta persalinan *sectio caesarea* (SC). Serta komplikasi yang akan terjadi setelah bersalin pada ibu hamil yaitu adanya risiko infeksi kandung kemih, memperberat komplikasi diabetes yang sudah ada sebelumnya

(gangguan pada organ lain seperti jantung, ginjal, saraf dan lainnya), lalu risiko menderita diabetes mellitus tipe 2 dalam jangka waktu 5-10 tahun setelah melahirkan. Komplikasi pada bayi juga dapat terjadi seperti bayi kuning (Ikterus Neonatorum), sindrom gangguan pernafasan bayi, hiperglikemia akut, peningkatan risiko obesitas dan diabetes pada saat anak-anak dan remaja, risiko mengalami diabetes mellitus tipe 2 pada saat usia dewasa dan berat bayi yang baru lahir bisa lebih dari 4000 gram.

C. Machine Learning

Machine Learning yang berbasis pada computer merupakan sebuah metode yang dilakukan computer tanpa perlu diatur dahulu oleh manusia dan seiring waktu dengan banyaknya data yang diolah akan membuatnya menjadi semakin pintar atau belajar dari pengalaman. Dan menurut Arthur pada tahun 1959 *machine learning* adalah kemampuan komputer dalam melakukan pembelajaran tanpa programmer harus menuliskan secara eksplisit kode programnya. Di dalam *machine learning* sendiri, proses pengolahan data akan terbagi menjadi dua yaitu *data training & data testing*. Dimana *data training* digunakan untuk melatih metode ataupun algoritma yang digunakan, sedangkan *data testing* digunakan untuk mengetahui kinerja dari metode maupun algoritma yang telah dilatih menggunakan *data training*, agar akurasi dan kinerja ketika program menemui data baru dapat diketahui.

Machine Learning jika dilihat dari teknik pembelajarannya maka akan terbagi dua yaitu *supervised learning* dan *unsupervised learning*. *Supervised learning* adalah teknik mengolah sekumpulan data (dataset) yang sudah memiliki label data untuk dipelajari sehingga saat mesin melakukan proses klasifikasi maupun prediksi, data yang telah berlabel dapat diidentifikasi. Sedangkan *unsupervised learning* merupakan teknik yang mengizinkan mesin dalam mengambil keputusan sendiri atas informasi yang diperlukan dan data biasanya tidak dilabeli.

D. Regresi

Regresi termasuk dalam teknik *supervised learning*, dimana metode ini akan membandingkan dua atau lebih variabel untuk mengetahui bagaimana

pengaruhnya satu sama lain, sehingga akan berdampak pada keputusan nantinya. Variabel yang dibandingkan adalah variabel *independent* dan variabel *dependent*. Dalam metode regresi sendiri terbagi menjadi beberapa sub-metode lagi, salah satunya adalah regresi logistik.

C.1. Regresi Logistik

Menurut Hosmer & Lemeshow, regresi logistik adalah metode analisis statistika yang digunakan untuk mendeskripsikan hubungan antara variabel respon dengan variabel prediktor. Dimana variabel respon (*dependent*) berbentuk data dikotomi seperti 1 atau 0, ya atau tidak, hidup atau mati, sakit atau sehat dan lainnya. Dan variabel prediktor (*independent*) berupa data kategorik maupun numerik. Sehingga, jika variabel responnya bernilai 1 maka menyatakan bahwa variabel respon sesuai dengan kriteria dari hasil pembelajaran menggunakan variabel prediktor, dan berlaku sebaliknya, dimana jika variabel bernilai 0 maka tidak sesuai dengan kriterianya.

METODE

A. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data sekunder yang didapatkan melalui repository dataset yang bersumber dari website kaggle.com. Kaggle merupakan tempat perkumpulan data scientist dimana tersedia banyak dataset yang dapat diambil dan tersedia banyak perlombaan yang berkaitan dengan Machine Learning. Dataset yang digunakan berjudul “Diabetes Dataset” yang dimana dataset tersebut didapat dari National Institute of Diabetes and Digestive and Kidney Diseases yang dipublikasikan di website Kaggle.

Dataset yang kami gunakan memuat data dari semua wanita dengan umur minimal dua puluh satu tahun yang berasal dari “Pima Indian Heritage”. Dengan sembilan atribut yang digunakan serta tujuh ratus enam puluh delapan *instance*. Dimana ada delapan atribut yang akan digunakan sebagai variabel independen, yaitu :

1. Pregnancies : berapa kali mengalami kehamilan

2. Glucose : konsentrasi *plasma glucose* selama dua jam yang didapat dari tes oral terhadap toleransi glukosa
3. Blood Pressure : tekanan darah diastolic (mmHg)
4. Skin Thickness : ketebalan dari lipatan kulit pada bagian trisep (mm)
5. Insulin : Serum insulin yang diberikan selama dua jam (mu U/ml)
6. BMI : Body Mass Index atau Indeks Massa Tubuh (bobot dalam kg/(tinggi dalam m)²)
7. Diabetes Pedigree Function : riwayat keturunan apakah penderita diabetes atau tidak
8. Age : umur (tahun)

Dan satu atribut sisanya sebagai variabel dependen yaitu Outcome (target/output) dengan nilai numerik 0 atau 1. Dimana jika bernilai 1 maka hasilnya positif menderita diabetes.

B. Tahapan Penelitian

Dalam melakukan penelitian ini, yang perlu dipersiapkan terlebih dahulu adalah menentukan latar belakang serta tujuan apa yang diinginkan nantinya. Dan batasan serta ruang lingkup dalam penelitian juga harus didefinisikan. Tinjauan pustaka juga harus ditulis, karena dapat membantu meningkatkan pemahaman terhadap komponen-komponen yang terlibat dalam penelitian, yaitu mengenai diabetes, machine learning, regresi serta regresi logistik.

Tahapan selanjutnya adalah langkah penelitian yaitu menyiapkan data yang akan digunakan, dimana penelitian ini menggunakan data dari website *kaggle*. Lalu data yang didapatkan harus dilakukan pembagian variabel menjadi variabel independen dan variabel dependen. Lalu data juga harus dibagi menjadi dua, yaitu yang akan digunakan sebagai *data training* dan satu lagi untuk *data testing*.

Lalu langkah yang ketiga adalah mengimplementasikan algoritma atau metode regresi logistik untuk mendeteksi apakah seseorang sesuai dengan gejala yang ada mengalami diabetes atau tidak. Pada langkah ini *data training* akan dilatih sehingga menghasilkan model yang sesuai dengan data. Lalu dilakukan test model yang ada kepada data yang menjadi *data testing*.

Dan langkah terakhir adalah melakukan analisa atau evaluasi terhadap hasil implementasi regresi logistik terhadap dataset, agar mendapat hasil yang sesuai dan baik. Dalam mengevaluasi digunakan *confusion matrix*. Lalu akan didapatkan akurasi dari sistem kita. Setelah itu, baru kesimpulan diambil dan memberikan saran dari hasil penelitian.

HASIL DAN PEMBAHASAN

Pada bagian ini diuraikan hasil dari model *machine learning* dengan metode regresi logistik yang telah dibuat. Model dibuat dengan menggunakan bahasa Python melalui *Google Collab*.

A. Model Regresi Logistik

Pertama adalah dengan mengimpor semua library yang kita perlukan. Lalu dengan dataset yang sudah kita miliki tadi, akan kita masukkan ke program. Dan pada gambar berikut dapat dilihat sebagian datanya.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

Dengan jumlah data sebanyak 768 dan memiliki 9 variabel. Lalu variabel akan dibagi menjadi variabel independen sebanyak 8 buah dan 1 buah lagi sebagai variabel dependen, dimana variabel dependen sebagai outcome atau target prediksi. Setelah didapat kedua variabel, maka dari jumlah total data akan dibagi menjadi dua bagian data, yaitu data training dan data testing. Data training akan diambil sebanyak 80% dari total data dan data testing sebanyak 20%. Dengan jumlah data 614 untuk data training dan 154 untuk data testing.

```
(
  Pregnancies  Glucose  BloodPressure  ...  BMI  DiabetesPedigreeFunction  Age
603          7      150           78    ...  35.2          0.692    54
118           4       97           60    ...  28.2          0.443    22
247           0      165           90    ...  52.3          0.427    23
157           1      109           56    ...  25.2          0.833    23
468           8      120           0     ...  30.0          0.183    38
..          ...      ...           ...    ...  ...          ...    ...
763          10     181           76    ...  32.9          0.171    63
192           7     159           66    ...  30.4          0.383    36
629           4       94           65    ...  24.7          0.148    21
559          11       85           74    ...  30.1          0.300    35
684           5      136           82    ...   0.0          0.640    69

[614 rows x 8 columns], 603    1
118    0
247    0
157    0
468    1
..
763    0
192    1
629    0
559    0
684    0
Name: Outcome, Length: 614, dtype: int64)

(
  Pregnancies  Glucose  BloodPressure  ...  BMI  DiabetesPedigreeFunction  Age
661           1      199           76    ...  42.9          1.394    22
113           2      107           74    ...  33.6          0.484    23
122           4       76           62    ...  34.0          0.391    25
14           5      166           72    ...  25.8          0.587    51
529           0      111           65    ...  24.6          0.660    31
..          ...      ...           ...    ...  ...          ...    ...
476           2      185           80    ...  33.7          0.711    29
482           4       85           58    ...  27.8          0.306    28
230           4      142           86    ...  44.0          0.645    22
527           3      116           74    ...  26.3          0.107    24
380           1      107           72    ...  30.8          0.821    24

[154 rows x 8 columns], 661    1
122    0
113    0
14     1
529    0
..
476    1
482    0
230    1
527    0
380    0
Name: Outcome, Length: 154, dtype: int64)
```

Lalu memodelkan dengan regresi logistik. Dimana data yang dibagi sebagai data training tadi akan di training. Lalu beri contoh ke model dengan parameter default pada regresi logistik. Setelah itu model yang sudah ada dimasukkan dan disesuaikan dengan data dari data training. Maka didapatkan persentase skor model trainingnya.

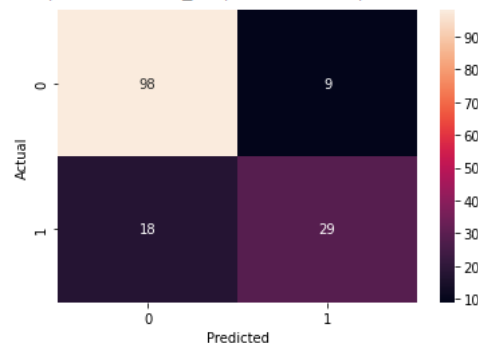
```
model train score: 0.764
```

Setelah itu model yang sudah ada digunakan untuk melakukan prediksi pada data testing. Sehingga menghasilkan output berupa 0 untuk tidak terindikasi diabetes dan 1 terindikasi diabetes. Prediksi dilakukan sesuai dari klasifikasi berdasar metode regresi logistik tadi.

```
array([1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1,
       1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1,
       1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1,
       0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0])
```

B. Evaluasi Model Regresi Logistik

Dalam mengavaluasi model yang sudah dibuat, maka digunakan *confussion matrix*. Dalam *confussion matrix* ini berisi nilai dari prediksi yang benar dan salah. Dan dalam model yang akan dievaluasi berbentuk objek *array* dalam matriks 2x2, karena model berbentuk biner. Dan didapatkan hasil dengan visualisai heatmap sebagai berikut.



Didapat output prediksi yang benar yaitu TP(True Positives) = 98 dan TN(True Negatives) = 29, serta prediksi yang tidak akurat dengan nilai FP(False Positives) = 18 dan FN(False Negatives) = 9. Dan didapat skor akurasi 0.825, dimana persentase akurasinya 82% dan cukup baik.

```
Accuracy score : 0.8246753246753247
Accuracy in Percentage : 82 %
```

KESIMPULAN

Penggunaan metode regresi logistic ini mendapatkan hasil yang bagus. Dengan akurasi dari model yang cukup baik diangka 82%, maka prediksi dengan metode regresi logistik ini sudah cukup sebagai awalan dalam pengembangan rancangan sistem deteksi ini kedepannya.

DAFTAR PUSTAKA

Safitri, A., Sudarmin & Nusrang, M. 2019. Model Regresi Logistik Biner pada Tingkat Pengangguran Terbukadi Provinsi Sulawesi Barat Tahun 2017. *Journal of Statistics and Its Application on Teaching and Research*. 1 (2):1-6.

Syukriah, A. 2018. Perbandingan Metode Regresi Logistik Dengan Analisis Diskrininan Untuk Mengetahui Faktor-Faktor Yang Berhubungan Dengan Kejadian Berat Badan Bayi Lahir Rendah (BBLR) Di Rumah Sakit Umum Pusat Haji Adam Malik Medan Tahun 2018. *Skripsi*. Fakultas Kesehatan Masyarakat Universitas Sumatra Utara, Medan.

Pangaribuan, J.J., Tanjaya, H. & Kenichi. 2021. Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression. *Journal Information System Development*. 6 (2).

Adikara, P. P., Wihandika, R. C., Utaminingrum, F., Sari, Y. A., Fauzi, M. A., Syauqy, D., & Maulana, R. 2017. Regresi Linier berbasis clustering untuk deteksi dan estimasi halangan pada smart wheelchair. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*. 3(1):11-16.

Tampil, Y.A., Komalig, A. & Langi, Y. 2017. Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *Jdc*. 6(2).

Suyanti & Sukestiyano, Y.L. 2014. Deteksi *Outlier* Menggunakan Diagnosa Regresi Berbasis Estimator Parameter *Robust*. *UNNES Journal of Mathematics*. 3(2).

Tyasnurita, R. & Pamungkas, A.Y.M. 2020. Deteksi Diabetik Retinopati Menggunakan Regresi Logistik. *ILKOM Jurnal Ilmiah*. 12(2):130-135.

Diksa, I.G.B.N & Fithriasari, K. 2021. Analisis Faktor Resiko Penyebab Diabetes Mellitus Dengan Regresi Logistik Biner. *Imferensi*. 4(1).

Biro Komunikasi dan Pelayanan Masyarakat, Kementerian Kesehatan RI 2018. *CEGAH, CEGAH, dan CEGAH: Suara Dunia Perangi Diabetes*. URL: <https://www.kemkes.go.id/article/view/18121200001/prevent-prevent-and-prevent-the-voice-of-the-world-fight-diabetes.html>. Diakses tanggal 10 November 2021.

Luwiharto, J. dan P. D. Ginanti 2010. *Diabetes Melitus*. URL : <https://prodiaohi.co.id/diabetes-melitus>. Diakses tanggal 11 November 2021.

Telaumbanua, F. D., Hulu, P., Nadeak, T. Z., Lumbantong, R. R., & Dharma, A. (2020). Penggunaan Machine Learning Di Bidang Kesehatan. *JURNAL TEKNOLOGI DAN ILMU KOMPUTER PRIMA (JUTIKOMP)*, 2(2), 391-399.

2021. Pengertian Machine Learning Menurut Para Ahli. URL : <https://www.contohapps.com/2021/05/pengertian-machine-learning-menurut.html>. Diakses tanggal 14 November 2021

Mufdillah, Ningsih, S.R., Subarto, C.B. & Fajarini Nurbita. 2019. *Mengenal dan Upaya Mengatasi Diabetes Melitus dalam Kehamilan*. Edisi ke-1, Nuha Medika. Jl.Nyi Wiji Andisoro, Pelemsari, Prenggan Kotagede Yogyakarta.

Fakultas Kedokteran, Kesehatan Masyarakat dan Keperawatan UGM. Diabetess Pada Ibu Hamil, Berbahayakah?. URL : <https://kanalpengetahuan.fk.ugm.ac.id/diabetes-pada-ibu-hamil-berbahayakah/>. Diakses tanggal 16 November 2021.