

# **IMPLEMENTASI LVQ DENGAN OPTIMASI ALGORITMA GENETIK PADA KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES**

Komputasi Intelegensia Kelompok 4



# ANGGOTA GRUP 4



**CORNELIUS JUSTIN  
SATRYO HADI**

**2006529796**

**MUHAMMAD HANIF  
PRAMUDYA ZAMZAMI**

**2006487566**

**JAVIER BINTORO**

**2006572150**



**MUHAMMAD DAFFA**

**2006568626**

**DAFFA AL GHIFARY**

**2006463420**

**RAFI ALVANZAH**

**2006528736**

**TULUS SETIAWAN**

**2006568802**

**01**





# DAFTAR ISI

**01**

**PENDAHULUAN**

**02**

**RUMUSAN  
MASALAH**

**03**

**TUJUAN**

**+**

**+**

**04**

**METODE**

**05**

**DESKRIPSI  
DATA**

**06**

**IMPLEMENTASI**

**02**



**01**

+

# PENDAHULUAN

+

**03**



# + PENDAHULUAN

Diabetes adalah penyakit jangka panjang atau kronis dan ditandai dengan kadar gula (glukosa) darah yang tinggi atau di atas nilai normal. Glukosa yang menumpuk di dalam darah akibat tidak diserap sel tubuh dengan baik dapat menyebabkan berbagai gangguan pada organ tubuh. Jika diabetes tidak dikontrol dengan baik, berbagai komplikasi yang dapat membahayakan nyawa pasien dapat muncul.



# PERSENTASE PERBANDINGAN PENDERITA DIABETES DI INDONESIA PADA TAHUN 2021

Penderita Diabetes

[1] **10.6%**

179,72 Juta

Bukan Penderita  
Diabetes

**89.4%**



# PENDAHULUAN

Kecerdasan buatan (AI) memberikan kemampuan untuk belajar dan meningkatkan secara otomatis dari pengalaman tanpa diprogram secara eksplisit. Teknik-teknik dari AI memungkinkan kita untuk memperoleh hasil akhir prediksi. [8]



# PENDAHULUAN

Teknik klasifikasi dari AI dapat memperkirakan secara akurat kelas target untuk setiap kasus dalam data. Algoritma klasifikasi umumnya mensyaratkan label kelas yang didefinisikan berdasarkan nilai atribut data.



07





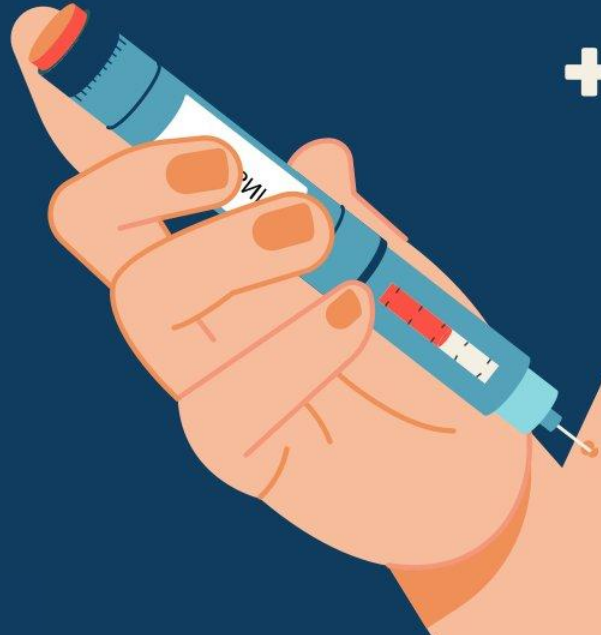
**02**

# **RUMUSAN MASALAH**

+



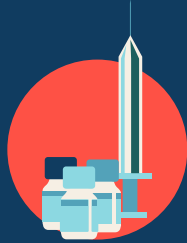
**08**



**5**



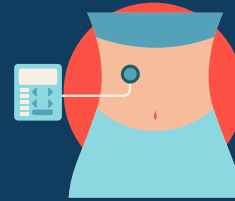
# RUMUSAN MASALAH



1.



Bagaimana implementasi Learning Vector Quantizer dengan optimasi Algoritma Genetik pada klasifikasi diagnosis penyakit diabetes?



2.

Bagaimana performa Learning Vector Quantizer dengan optimasi Algoritma Genetik terhadap masalah klasifikasi diagnosis penyakit diabetes berdasarkan akurasi, precision, recall, dan f1 score?



**03**

+

**TUJUAN**

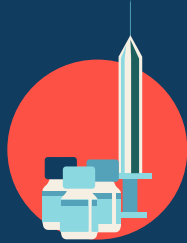
+

**10**





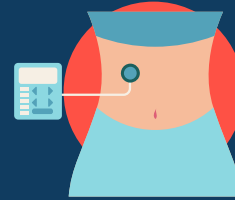
# TUJUAN



1.



Mengimplementasikan Learning Vector Quantizer dengan optimasi Algoritma Genetik pada klasifikasi diagnosis penyakit diabetes.



2.

Menganalisis performa Learning Vector Quantizer dengan optimasi Algoritma Genetik terhadap masalah klasifikasi diagnosis penyakit diabetes berdasarkan akurasi, precision, recall, dan f1 score.



04

METODE



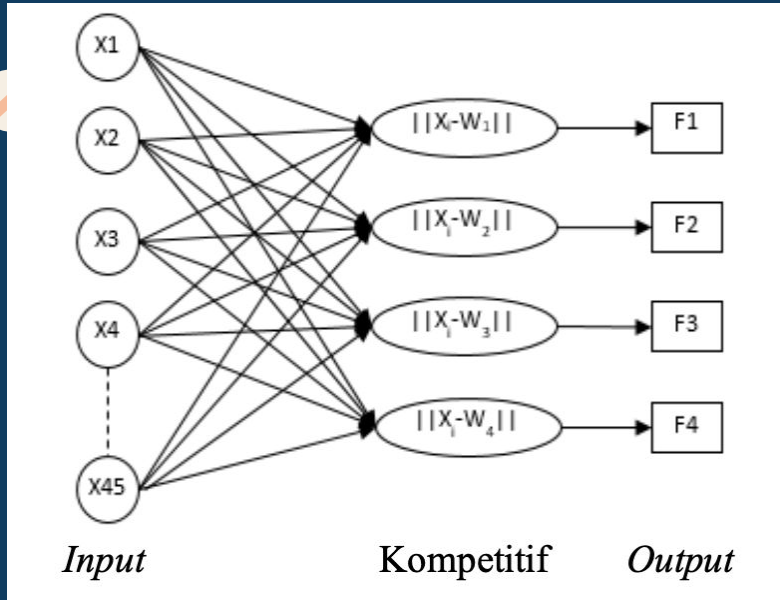
12

# + LVQ

## (LEARNING VECTOR QUANTIZER)+

<b>DEFINISI</b>	adalah keluarga algoritma untuk teknik klasifikasi AI, yang bertujuan untuk mempelajari prototipe ( <i>codebook vector</i> ) yang mewakili wilayah kelas.
<b>SIFAT</b>	bersifat heuristik dan menunjukkan kepekaan terhadap inisialisasi, masalah konvergensi lambat, dan ketidakstabilan.
<b>TUJUAN</b>	untuk menentukan vektor bobot (prototipe), sehingga sampel data pelatihan dipetakan ke label kelas yang sesuai.

## ARSITEKTUR LVQ



Lapisan input, menerima informasi (input) terkait permasalahan.

Lapisan kompetitif, memproses jarak kedekatan input dengan bobot yang berpadanan untuk setiap output. Jarak terdekat akan menentukan kelas suatu input.

Lapisan output, lapisan yang menampilkan output dari kelas yang sudah ditentukan di kelas kompetitif.

# + LANGKAH PROSES LVQ +

1. Penghitungan jarak terhadap masing-masing kelas:  
Euclidean Distance

$$D_{i,j} = \|X_i - W_j\|$$

$X_i$  = Data atau input ke- $i$   
 $W_j$  = Bobot kelas ke- $j$

2. Menentukan kelas dari data ke- $i$  berdasarkan jarak minimum terhadap seluruh bobot:

$$C_i = \begin{cases} 0 & , \min(D_{i,0}, D_{i,1}) = D_{i,0} \\ 1 & , \min(D_{i,0}, D_{i,1}) = D_{i,1} \end{cases}$$

3. Perbarui Bobot





# LANGKAH PROSES LVQ



4. Ulangi langkah (1-3) hingga tercapai epoch maksimum:
5. Hitung Akurasi

$$\text{Akurasi} = \frac{\text{jumlah data dengan kelas sesuai target}}{\text{jumlah seluruh data}} \times 100$$





# GA (GENETIC ALGORITHM)



*Genetic Algorithm (GA)* adalah bagian dari *Evolutionary Algorithm* yaitu suatu algoritma yang mencontoh proses evolusi alami di mana konsep utamanya adalah individu-individu yang paling unggul akan bertahan hidup, sedangkan individu-individu yang lemah akan punah [2].

Pada masalah klasifikasi diagnosis penyakit diabetes ini, GA akan digunakan untuk mengoptimasi vektor bobot awal pada LVQ yang diharapkan dapat menghasilkan nilai fitness tertinggi.

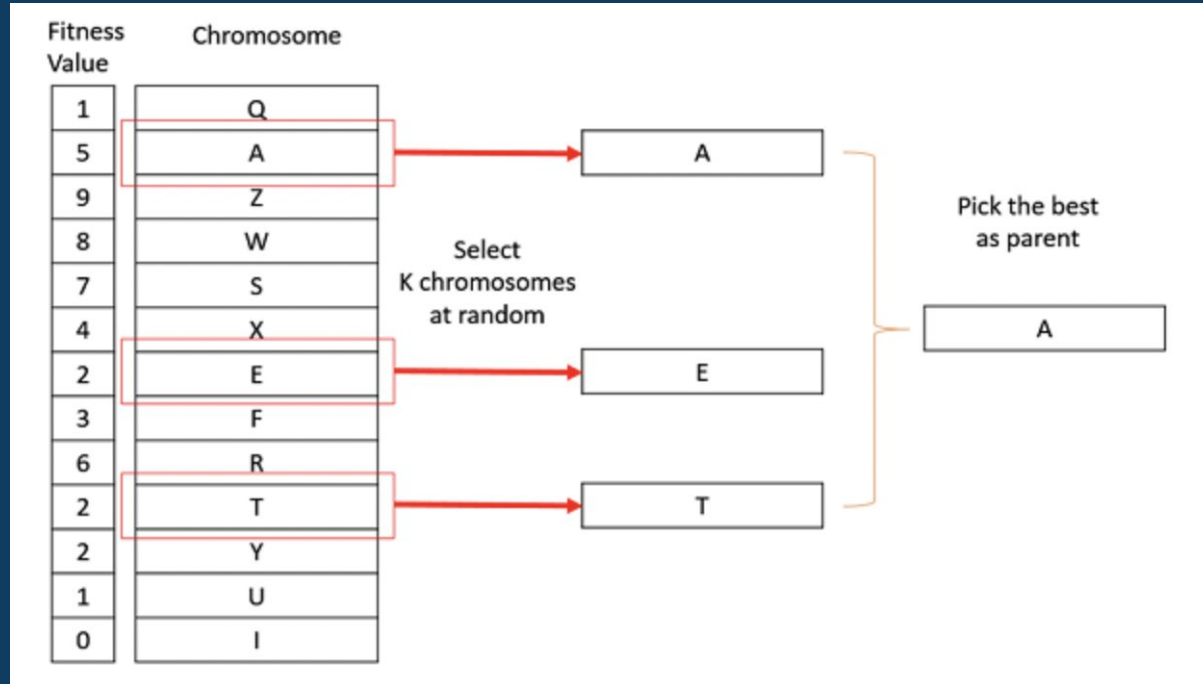


# + LANGKAH PROSES GA +

1. Membentuk populasi awal
2. Menghitung fitness dari setiap kromosom:  
*Fitness = Akurasi LVQ*
3. Memilih kromosom sebagai parent:  
*Tournament Rank Selection*
4. Melakukan crossover untuk membentuk kromosom baru:  
*Two-point crossover*
5. Melakukan mutasi pada kromosom untuk memberi variasi:  
*Random mutation*

# + PENJELASAN MATEMATIS

Seleksi Parent → Tournament Parent Selection



Ket.

Pilih K-kromosom dari populasi secara acak dan pilih yang terbaik dari ini untuk menjadi *Parent*. Proses yang sama diulangi untuk memilih *parent* berikutnya.



# PENJELASAN MATEMATIS



Crossover  $\longrightarrow$  Two-Point Crossover

$$O1 = [P1(1), \dots, P1(p), P2(p+1), \dots, P2(q), P1(q+1), \dots, P1(n)]$$

$$O2 = [P2(1), \dots, P2(p), P1(p+1), \dots, P1(q), P2(q+1), \dots, P2(n)]$$

$$0 < p < q < n$$

$O1$  = Offspring 1

$n$  = Jumlah gen pada kromosom

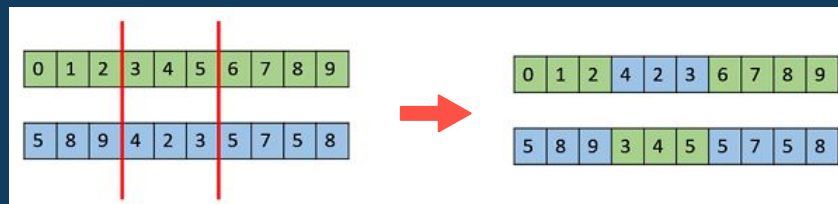
$O2$  = Offspring 2

$p$  = Nilai acak antara 0 dan  $n - 1$

$P1$  = Parent 1

$q$  = Nilai acak antara  $p$  dan  $n$

$P2$  = Parent 2





# PENJELASAN MATEMATIS



Mutation → Random Mutation

$$G_n = G_n + r(\max_n - \min_n)$$

$G_n$  = Gen ke- $n$

$r$  = Nilai acak  $[-0.1, 0.1]$

$\max_n$  = Nilai maksimal pada gen ke- $n$

$\min_n$  = Nilai minimal pada gen ke- $n$



# + PENJELASAN MATEMATIS

Seleksi Fitness → Elitism

Kromosom	Fitness
Kromosom 1	75
Kromosom 2	50
Kromosom 3	92
Kromosom 4	60
Kromosom 5	40
Kromosom 6	33



Kromosom	Fitness
Kromosom 3	92
Kromosom 1	75
Kromosom 4	60
Kromosom 2	50

Ket.

Memilih K-kromosom terbaik untuk menjadi populasi di generasi berikutnya.

# PENJELASAN METRICS

## PRECISION

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$\text{Precision} = (TP) / (TP + FP)$$

## RECALL

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$\text{Recall} = (TP) / (TP + FN)$$

## F1 SCORE

Mengukur keseimbangan antara Precision – Recall. Nilai terbaik F1-Score adalah 1.0 dan nilai terburuknya adalah 0

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

## AKURASI

$$\text{Akurasi} = \frac{\text{jumlah data dengan kelas sesuai target}}{\text{jumlah seluruh data}} \times 100$$



# PENJELASAN PARAMETER

## POPULASI

Sekumpulan individu yang menyimpan solusi. Setiap individu dalam populasi mewakili sebuah solusi dan setiap individu terdiri dari sekumpulan nilai atau variabel yang disebut gen.

## GENERASI

Generasi pada *genetic algorithm* mengacu pada satu iterasi pada algoritma. Setiap generasi menghasilkan solusi berdasarkan generasi sebelumnya.

## LEARNING RATE ( $\alpha$ )

*Learning Rate* menentukan seberapa besar bobot pada neuron di ubah pada setiap iterasi.

## CROSSOVER RATE (CR)

Tingkat kemungkinan terjadinya penggabungan dua individu sehingga terbentuk individu baru yang memiliki sifat-sifat dari kedua individu tersebut.

# PENJELASAN PARAMETER

## PROBABILITY Crossover (PC)

*Probability Crossover* adalah parameter yang menentukan probabilitas dipilihnya dua solusi (*parent solution*) untuk menghasilkan solusi baru (*child solution*).

## PROBABILITY MUTATION (PM)

*Probability Mutation* adalah parameter yang menentukan probabilitas solusi (*chromosome*) akan di modifikasi.

## DECREASING LEARNING RATE

Laju penurunan learning rate pada setiap iterasi.

## MINIMUM LEARNING RATE

Learning rate minimum. Jika tercapai learning rate minimum iterasi akan dihentikan.

# PENJELASAN PARAMETER

## EPOCH

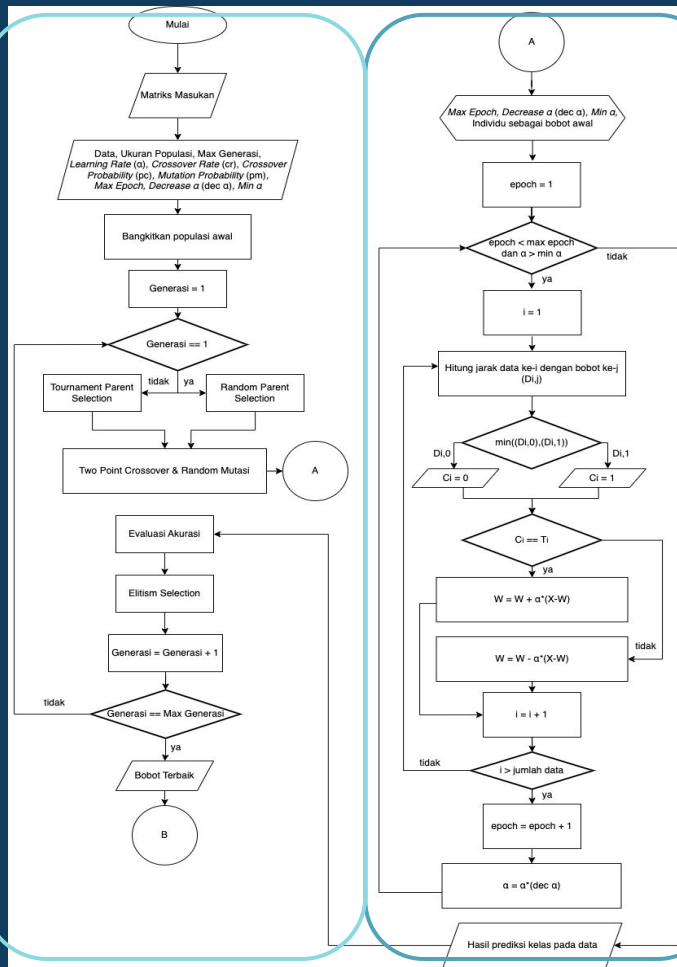
Ketika seluruh dataset sudah melalui proses training pada Neural Network sampai dikembalikan ke awal untuk sekali putaran.

## BOBOT (W)

Bobot dari setiap kriteria yang akan dijadikan perhitungan dalam penentuan label kelas data.

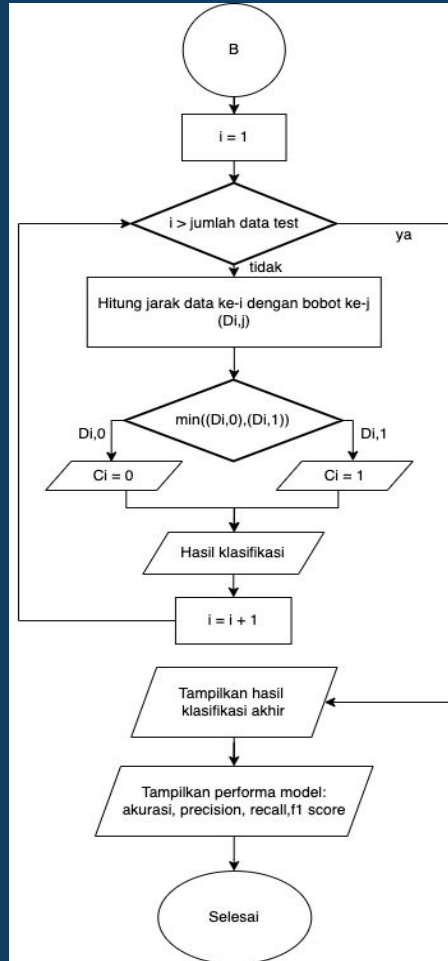
# FLOWCHART TRAINING LVQ-GA PROCESS

PROSES GA



PROSES LVQ

# FLOWCHART TESTING LVQ-GA PROCESS



+

05

+

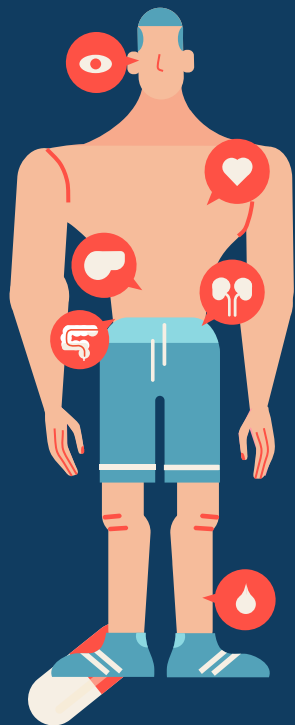
# DESKRIPSI DATA

+





# DATASET DIABETES



Jumlah Sample: 768

Format dataset: CSV

Fitur pada dataset:

1. Pregnancies
2. Glucose
3. BloodPressure
4. SkinThickness
5. Insulin
6. BMI
7. DiabetesPedigreeFunction
8. Age
9. Outcome

30





# PENJELASAN DATASET



1. Pregnancies : jumlah kehamilan
2. Glucose : konsentrasi glukosa plasma
3. Blood Pressure : tekanan darah diastolik (mm/Hg)
4. Skin Thickness : ketebalan lipatan kulit trisep (mm)
5. Insulin : insulin (U/mL)
6. BMI : indeks berat badan (kg/m<sup>2</sup>)
7. Diabetes Pedigree Function : nilai kecenderungan diabetes berdasarkan riwayat keluarga
8. Age : usia (tahun)







# SAMPLE DATASET



	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
369	1	133	102	28	140	32.8	0.234	45	1
316	3	99	80	11	64	19.3	0.284	30	0
396	3	96	56	34	115	24.7	0.944	39	0
542	10	90	85	32	0	34.9	0.825	56	1
480	3	158	70	30	328	35.5	0.344	35	1

## TARGET CLASS

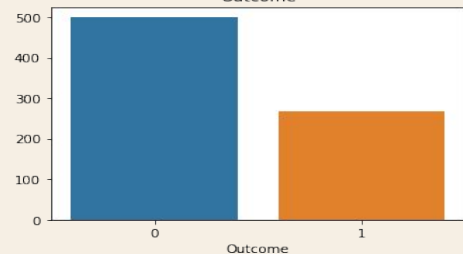
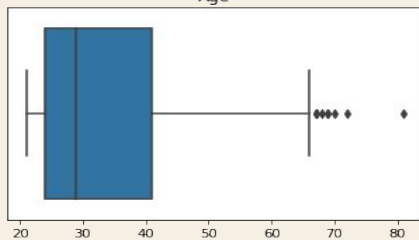
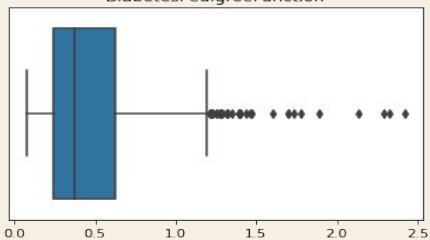
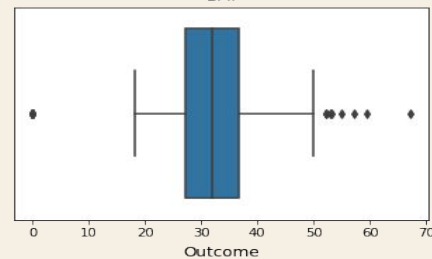
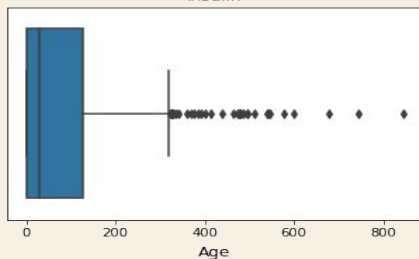
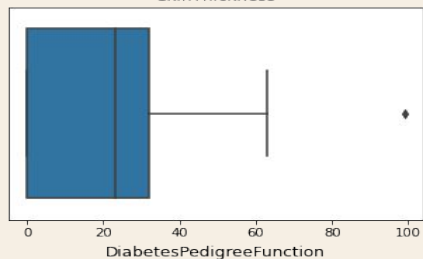
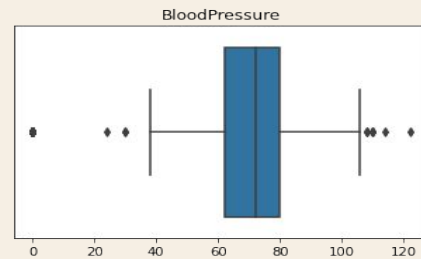
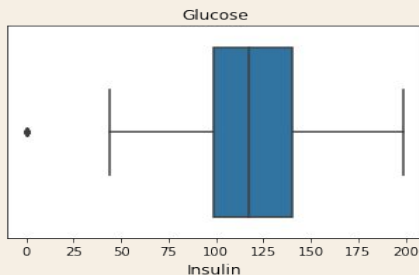
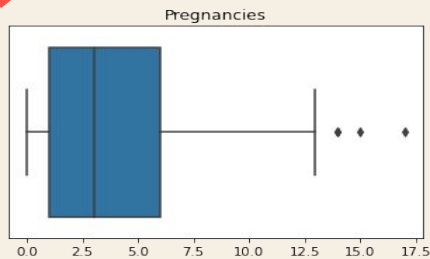
Dataset memiliki *target class* yang menyatakan apakah individu tersebut diabetes (1) atau tidak diabetes (0).

32





# DISTRIBUSI DATA



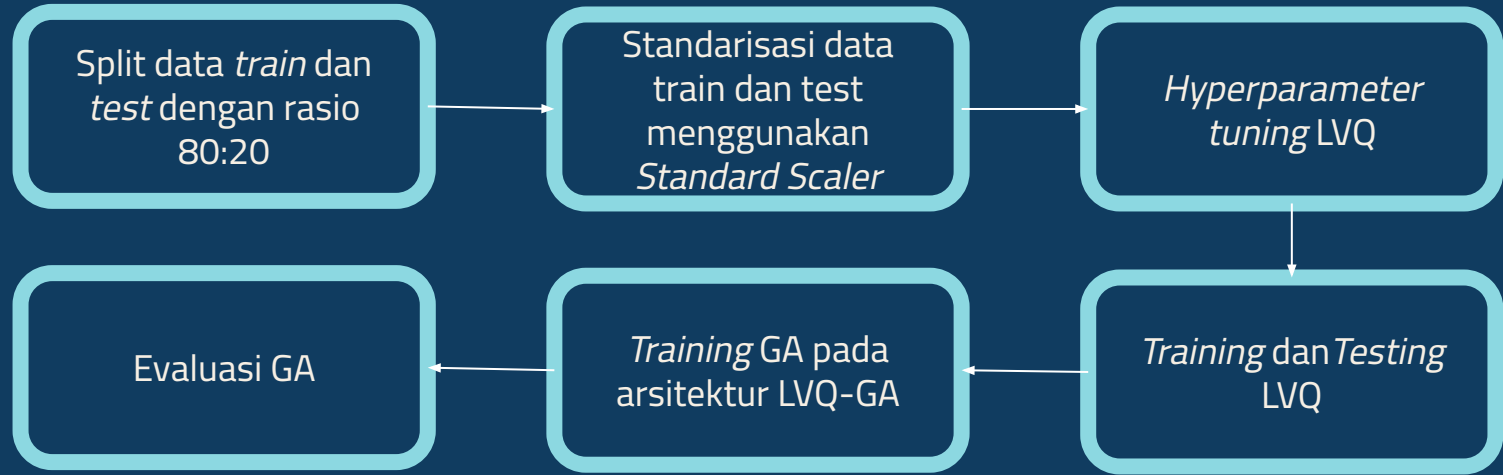
# 06

## IMPLEMENTASI



34

# ALUR IMPLEMENTASI



# HYPERPARAMETER TUNING LVQ

Tuning dilakukan menggunakan Randomized Search dengan 5-fold cross validation dan iterasi sebanyak 200.

List kandidat Hyperparameter:

Hyperparameter	Nilai yang akan di-tuning
Learning rate ( $\alpha$ )	Bilangan real dalam range (0.001, 1)
Decrease alpha (dec $\alpha$ )	Bilangan real dengan range (0.1, 0.9)
Minimum alpha (min $\alpha$ )	Bilangan real dengan range ( $10^{-100}$ , $10^{-4}$ )
Epochs	Bilangan bulat dengan range (50, 200)

# HYPERPARAMETER TUNING LVQ

Hasil hyperparameter *tuning* dengan 5 akurasi terbaik pada data *cross validation*:

Rank	Alpha	dec alpha	min alpha	Epochs	Mean Akurasi
1	0.074589	0.1	0.0001	200	75.73%
2	0.05423	0.491641	0.000056	197	75.72%
3	0.067427	0.223951	0.000094	193	75.57%
4	0.06424	0.483451	0.000093	200	75.56%
5	0.136618	0.132077	0.000034	197	75.41%



# HASIL TESTING LVQ



Dengan menggunakan *hyperparameter* teroptimal pada saat proses *tuning*, diperoleh metrik evaluasi pada data test sebagai berikut:

Akurasi	F1-Score	Precision	Recall
74%	62%	65%	60%





# HASIL TESTING LVQ-GA

Menggunakan tiga kombinasi *hyperparameter* GA yang berbeda, diperoleh metrik evaluasi pada data test sebagai berikut:

Ukuran populasi	Max generasi	cr	pc	pm	Akurasi	F1-Score	Precision	Recall
50	100	0.5	0.9	0.5	75%	51%	79%	38%
30	20	0.9	0.9	0.1	76%	69%	64%	75%
100	100	0.9	0.9	0.5	79%	70%	70%	70%







# ANALISIS

- Dapat dilihat berdasarkan proses tuning bahwa didapat parameter terbaik untuk model LVQ dengan performa model:



Alpha	dec alpha	min alpha	Epochs
0.07458	0.1	0.0001	200

Akurasi	F1-Score	Precision	Recall
74%	62%	65%	60%





# ANALISIS

- Dapat dilihat berdasarkan proses tuning bahwa didapat parameter terbaik untuk model LVQ-GA dengan performa model:



Ukuran populasi	Max generasi	cr	pc	pm	Alpha	dec alpha	min alpha	Epochs
100	100	0.9	0.9	0.5	0.07458	0.1	0.0001	200



Akurasi	F1-Score	Precision	Recall
79%	70%	70%	70%





# ANALISIS



Perbandingan performa antara model LVQ dan LVQ-GA

LVQ			
Akurasi	F1-Score	Precision	Recall
74%	62%	65%	60%

LVQ-GA			
Akurasi	F1-Score	Precision	Recall
79%	70%	70%	70%

Ket.

Performa LVQ-GA mungkin saja dapat dikembangkan agar memperoleh metric yang lebih baik dengan mencoba metode *crossover*, *mutation*, *selection*, serta *hyperparameter* yang tidak digunakan pada penelitian ini.





# KESIMPULAN



- Didapat konstruksi akhir model LVQ-GA yang dapat diimplementasikan untuk klasifikasi diagnosis penyakit diabetes.
- Penentuan Bobot Awal dalam proses training model LVQ sangatlah penting, sehingga dengan adanya GA dapat mengoptimasi bobot awal tersebut. Hal ini dibuktikan dengan Model LVQ-GA yang memiliki performa yang lebih baik dibanding model LVQ dari tabel akurasi, f1-score, precision, recall.





# KESIMPULAN



- Didapat akurasi model maksimal 79%, yang berarti memprediksi baik penderita diabetes dan sehat dengan cukup baik
- Didapat F1-Score model maksimal 70%, yang berarti bahwa model klasifikasi kita punya precision dan recall yang cukup baik
- Didapat Precision model maksimal 70%, yang berarti dapat memprediksi penderita diabetes dari keseluruhan penderita diabetes dalam data dengan cukup baik
- Didapat Recall model maksimal 70%, yang berarti dapat memprediksi penderita diabetes dari keseluruhan data dengan cukup baik



Dengan begitu, didapat hasil akhir konstruksi model yang menjawab tujuan awal dari penelitian.





# REFERENSI

- [1] Pahlevi, R. (2021, November 22). *Jumlah Penderita Diabetes Indonesia Terbesar Kelima di Dunia*. Katadata.  
<https://databoks.katadata.co.id/datapublish/2021/11/22/jumlah-penderita-diabetes-indonesia-terbesar-kelima-di-dunia>
- [2] *Genetic Algorithm*. (2018, December 8). School of Computer Science.  
<https://socs.binus.ac.id/2018/12/08/genetic-algorithm/>
- [3] Engelbrecht, A. (2006). *Fundamentals of Computational Swarm Intelligence*. John Wiley & Sons.
- [4] Ariyawan, M. D., Wibawa, I. G., & Ayu, L. A. (2020). *Diagnosis of Heart Disease Using Generalized Learning Vector Quantization (GLVQ) and Genetic Algorithms*.
- [5] Arniantya, R., Setiawan, B. D., & Adikara, P. P. (2018). *Optimasi Vektor Bobot Pada Learning Vector Quantization Menggunakan Algoritme Genetika Untuk Identifikasi Jenis Attention Deficit Hyperactivity Disorder Pada Anak*.





# REFERENSI

[6] Sulistyawati, D. H., & Murtadho, A. (2020). Performance Accuration Method of Machine Learning for Diabetes Prediction. *Jurnal Mantik*. <https://iocscience.org/ejournal/index.php/mantik/article/view/725/482>



[7] Nova, D., & Estévez, P. A. (2013). A review of learning vector quantization classifiers. *Neural Computing and Applications*, 25(3–4), 511–524. <https://doi.org/10.1007/s00521-013-1535-3>

[8] Dwi Harini Sulistyawati, Ali Murtadho (2020). Performance Accuration Method of Machine Learning for Diabetes Prediction.  
<https://iocscience.org/ejournal/index.php/mantik/article/view/725/482>



An illustration on a dark blue background showing a hand with a blue sleeve pointing a finger into a blue glucometer held by another hand. A red drop of blood is falling from the finger into the device's sample port. The glucometer's screen displays the number '104'. The scene is decorated with white plus signs and red-and-white capsules.

**TERIMA KASIH**