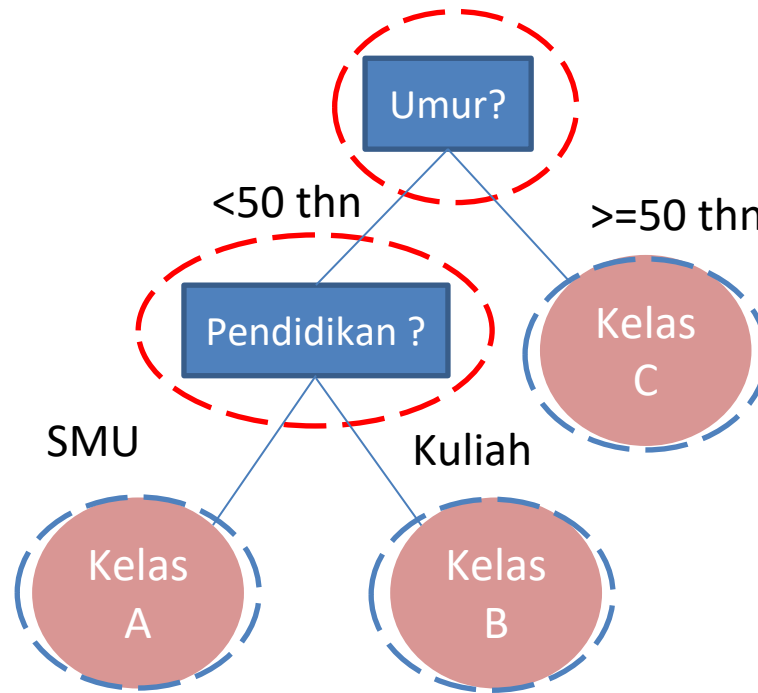


Learning



Decision Tree Learning

CONTOH DECISION TREE



Entropy

$$Entropy(S) = \sum_i^c -p_i \log p_i,$$

Di mana :

c : Jumlah nilai yang ada pada atribut target

p_i : Menyatakan jumlah sampe untuk i

Information gain

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \left(\frac{|S_v|}{|S|} \right) Entropy(S_v)$$

Di mana :

A : atribut

S : menyatakan suatu nilai yang mungkin untuk atribut A

Values(A) : Himpunan nilai-nilai yang mungkin untuk atribut A

$|S_v|$: Jumlah sampel untuk nilai v

$|S|$: Jumlah seluruh sampel data

Entropy(S_v) : entropy untuk sampel-sampel yang memiliki nilai v

Pelamar	IPK	Psikologi	Wawancara	Diterima
P1	Bagus	Tinggi	Baik	Ya
P2	Bagus	Sedang	Baik	Ya
P3	Bagus	Sedang	Buruk	Ya
P4	Bagus	Rendah	Buruk	Tidak
P5	Cukup	Tinggi	Baik	Ya
P6	Cukup	Sedang	Baik	Ya
P7	Cukup	Sedang	Buruk	Ya
P8	Cukup	Rendah	Buruk	Tidak
P9	Kurang	Tinggi	Baik	Ya
P10	Kurang	Sedang	Buruk	Tidak
P11	Kurang	Rendah	Baik	Ya

$$S = |8+, 3-|, S = 11$$

$$S_{\text{Bagus}} = |3+, 1-|, S = 4$$

$$S_{\text{Cukup}} = |3+, 1-|, S = 4$$

$$S_{\text{Kurang}} = |2+, 1-|, S = 3$$

$$\begin{aligned} Entropy(S) &= -\left(\frac{8}{11}\right) \log_2 \left(\frac{8}{11}\right) - \left(\frac{3}{11}\right) \\ &= 0,8454 \end{aligned}$$

$$\begin{aligned} Entropy(S_{Bagus}) &= -\left(\frac{3}{4}\right) \log_2 \left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) \log_2 \left(\frac{1}{4}\right) \\ &= 0,8113 \end{aligned}$$

$$\begin{aligned} Entropy(S_{Cukup}) &= -\left(\frac{3}{4}\right) \log_2 \left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) \log_2 \left(\frac{1}{4}\right) \\ &= 0,8113 \end{aligned}$$

$$\begin{aligned} Entropy(S_{Kurang}) &= -\left(\frac{3}{4}\right) \log_2 \left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) \log_2 \left(\frac{1}{4}\right) \\ &= 0,8113 \end{aligned}$$

ID₃

Function ID3 (Kumpulan Sampel, AtributTarget, KumpulanAtribut)

1. Pertama kita buat simpul Root

2. IF semua sampel adalah kelas i , maka Return pohon satu simpul Root dengan label = i

3. if Kumpulan atribut kosong, Return pohon satu simpul root dengan label = nilai atribut target yang paling umum (yang paling sering muncul)

Else

- $A \leftarrow$ Atribut yang merupakan the best classifier (dengan information gain terbesar)
- Atribut keputusan untuk Root $\leftarrow A$
- For v_1 (setiap nilai pada A)
 - Tambahkan suatu cabang di bawah Root sesuai dengan nilai v_1
 - Buat suatu variabel, misalnya Sampel_{v_1} , sebagai himpunan bagian (subnet) dari KumpulanSampel yang bernilai v_1 pada atribut A
 - If Sampel = kosong
 - Then di bawah ini ini tambahkan suatu simpul daun (leaf node, simpul yang tidak punya anak di bawahnya) dengan label = nilai atribut target yang paling umum (yang paling sering muncul)
 - Else di bawah cabang ini tambahkan subtree dengan memanggil fungsi ID3 (Sampel_{v_1} , Atribut Target, Atribut- $|A|$)

End

End

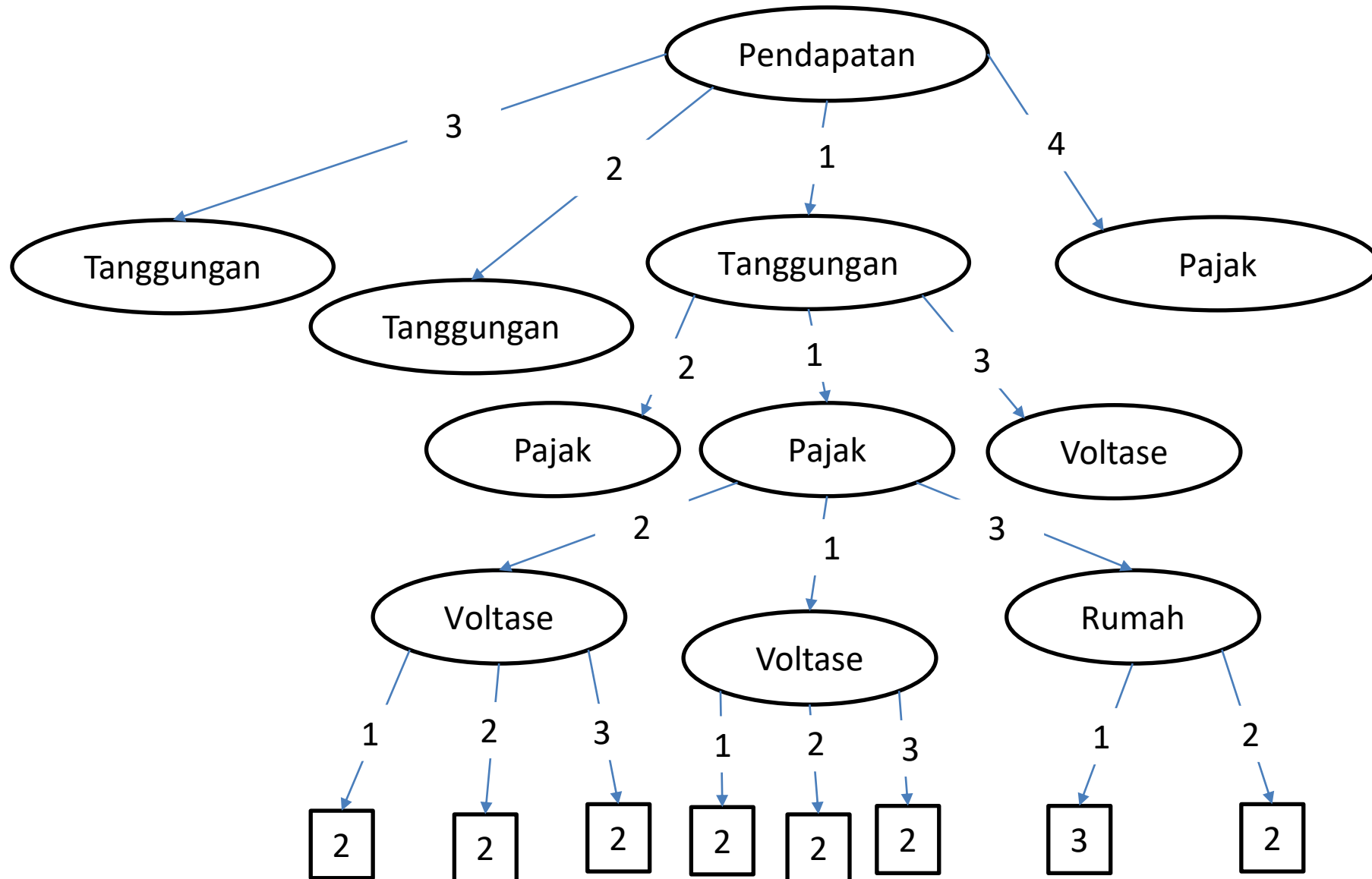
End

4. Return Root

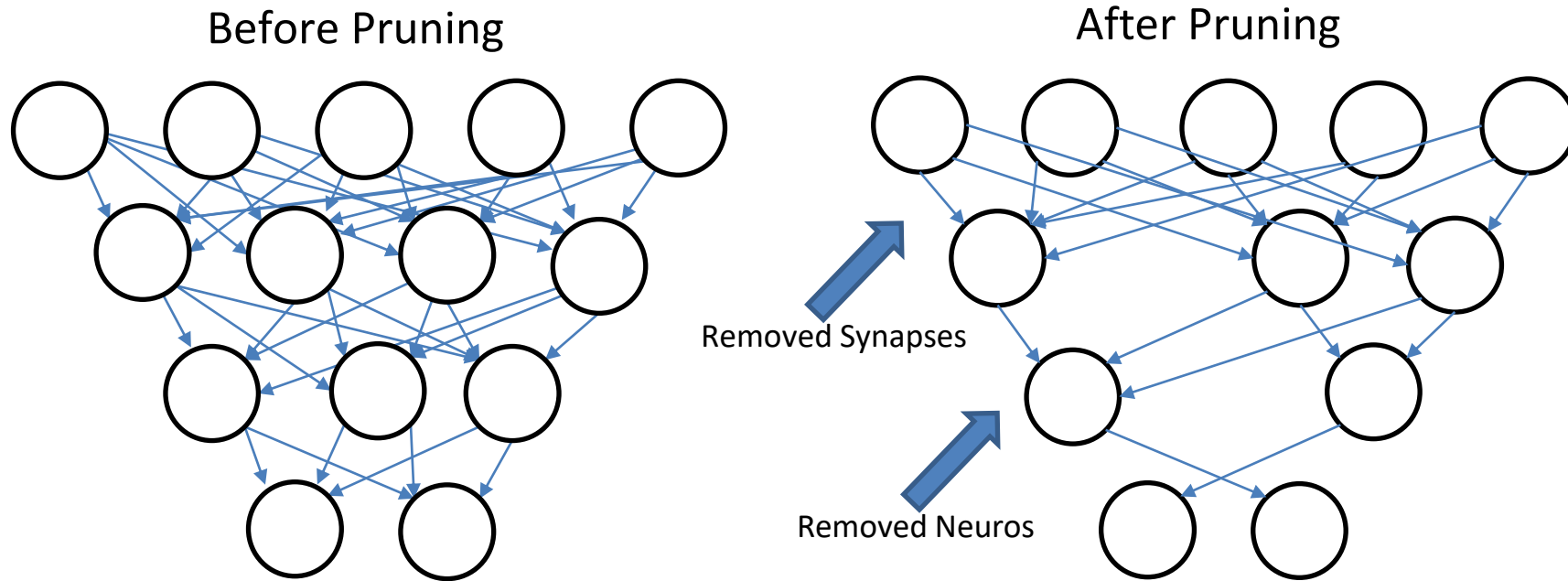
Permasalahan pada Decision Tree Learning

- Berapa ukuran pohon keputusan yang tepat?
- Selain information gain, adakah ukuran pemilihan atribut yang lain?
- Bagaimana jika atributnya bernilai kontinyu?
- Bagaimana menangani sampel data yang atributnya bernilai kosong?
- Dan bagaimana menangani atribut-atribut yang memiliki biaya berbeda?

Berapakah ukuran pohon yang tepat?



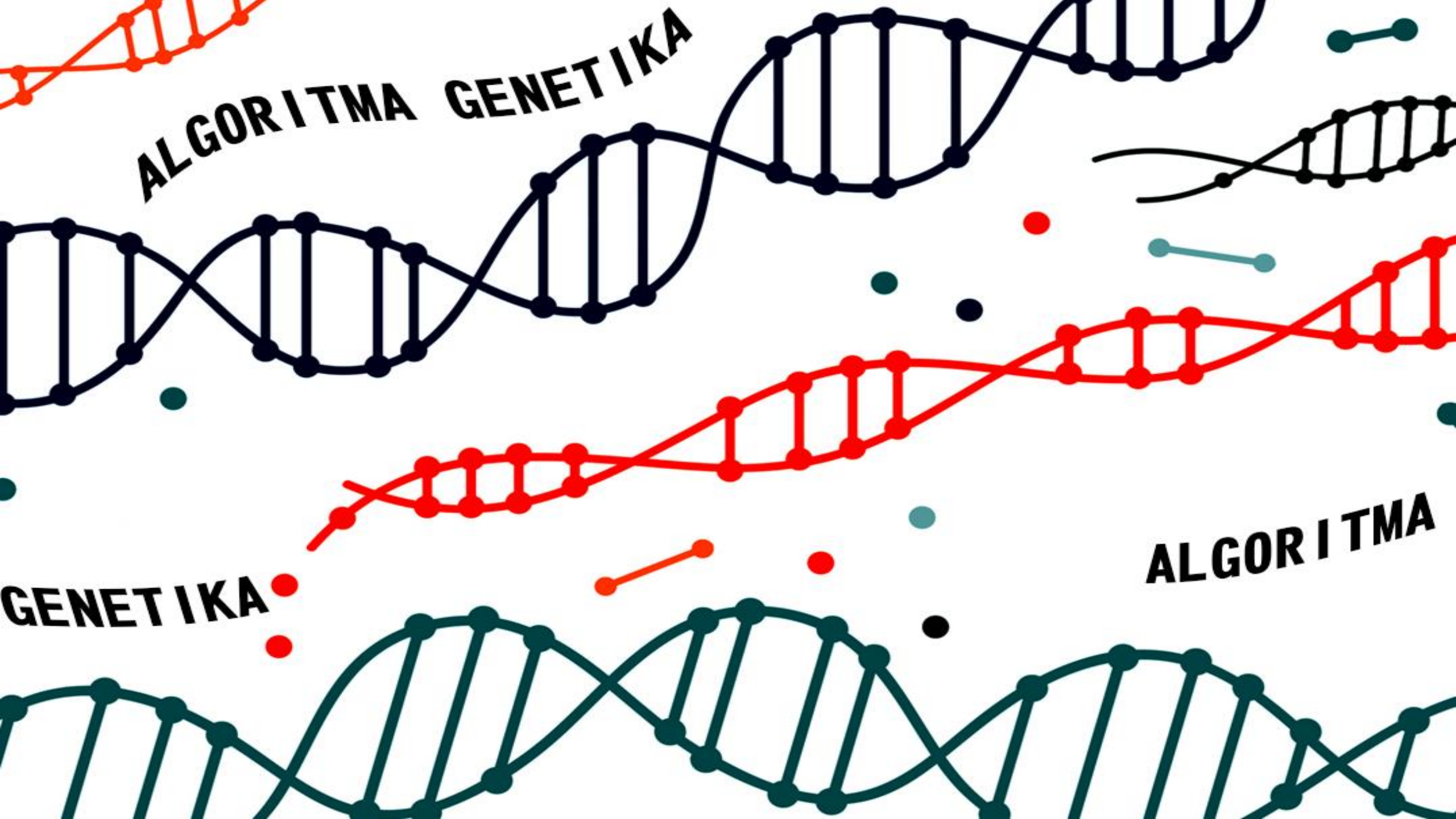
REP



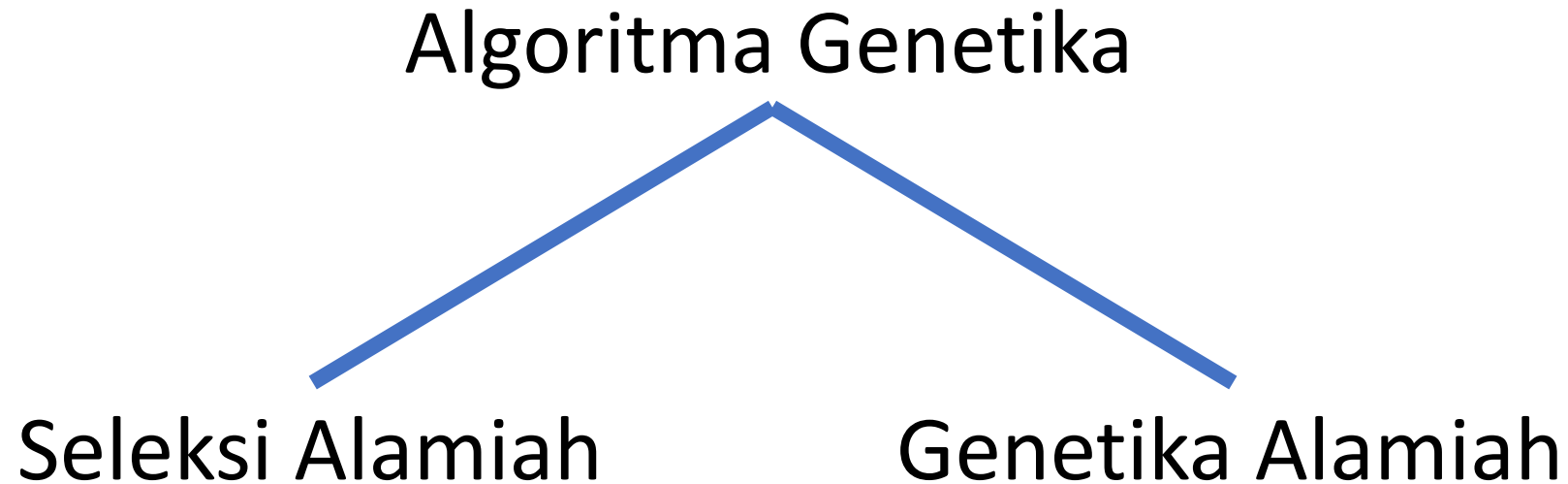
RPP

1. Dengan menggunakan sampel-sampel data pada *training set*, bangun pohon keputusan di mana yang paling sesuai. Biarkan *overfitting* terjadi.
2. Ubah pohon keputusan yang dihasilkan menjadi sekumpulan aturan, di mana satu aturan mempresentasikan satu jalur pohon dari simpul akar sampai simpul daun.
3. Ubah pohon keputusan yang dihasilkan menjadi sekumpulan aturan, di mana satu aturan mempresentasikan satu jalur pohon dari simpul akar sampai simpul daun.
4. Urutkan aturan-aturan hasil pemangkasan berdasarkan akurasi perkiraannya. Pilih aturan-aturan hasil pemangkasan berdasarkan urutan tersebut.

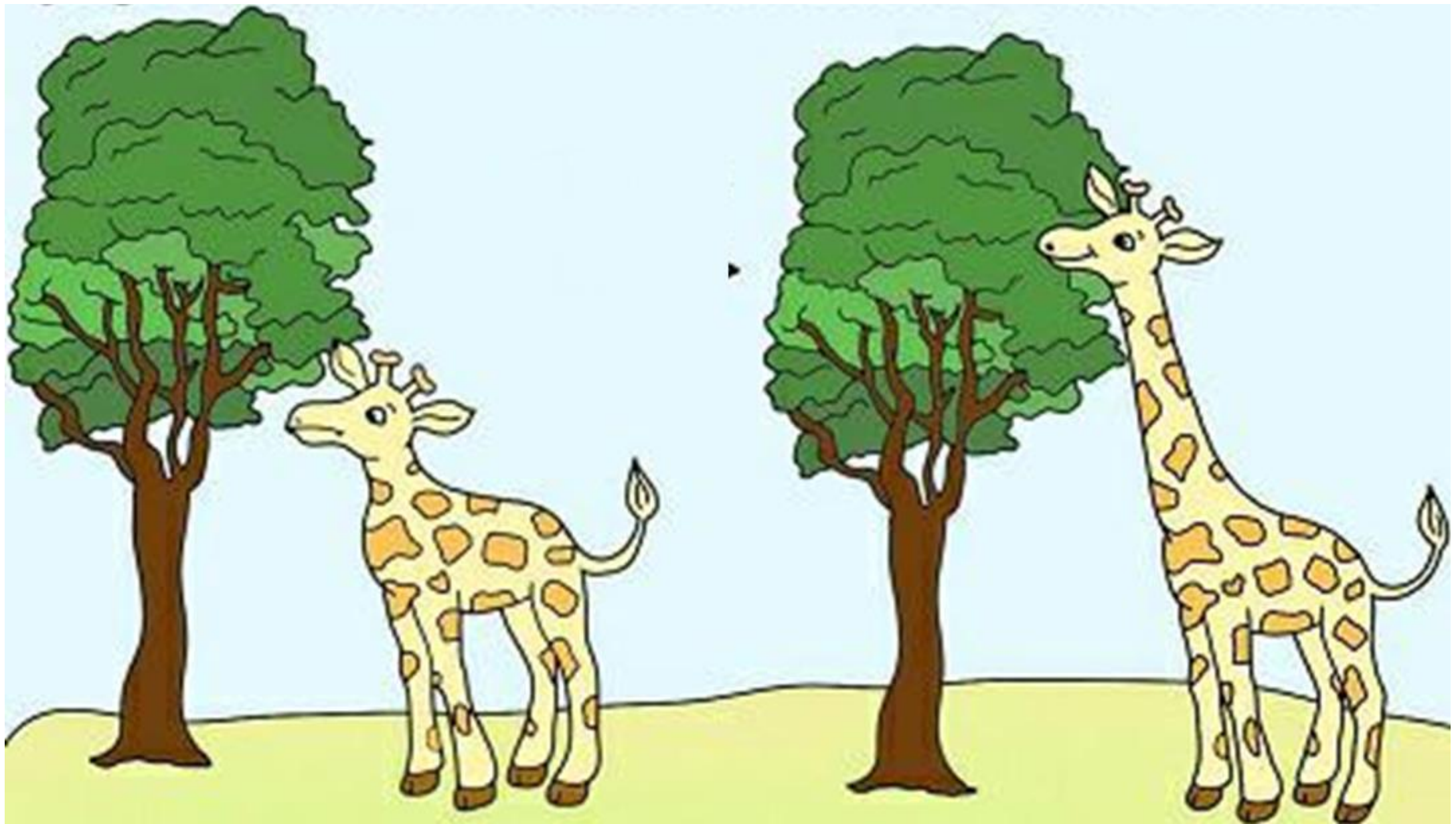
Bagaimana jika atributnya bernilai kontinyu?



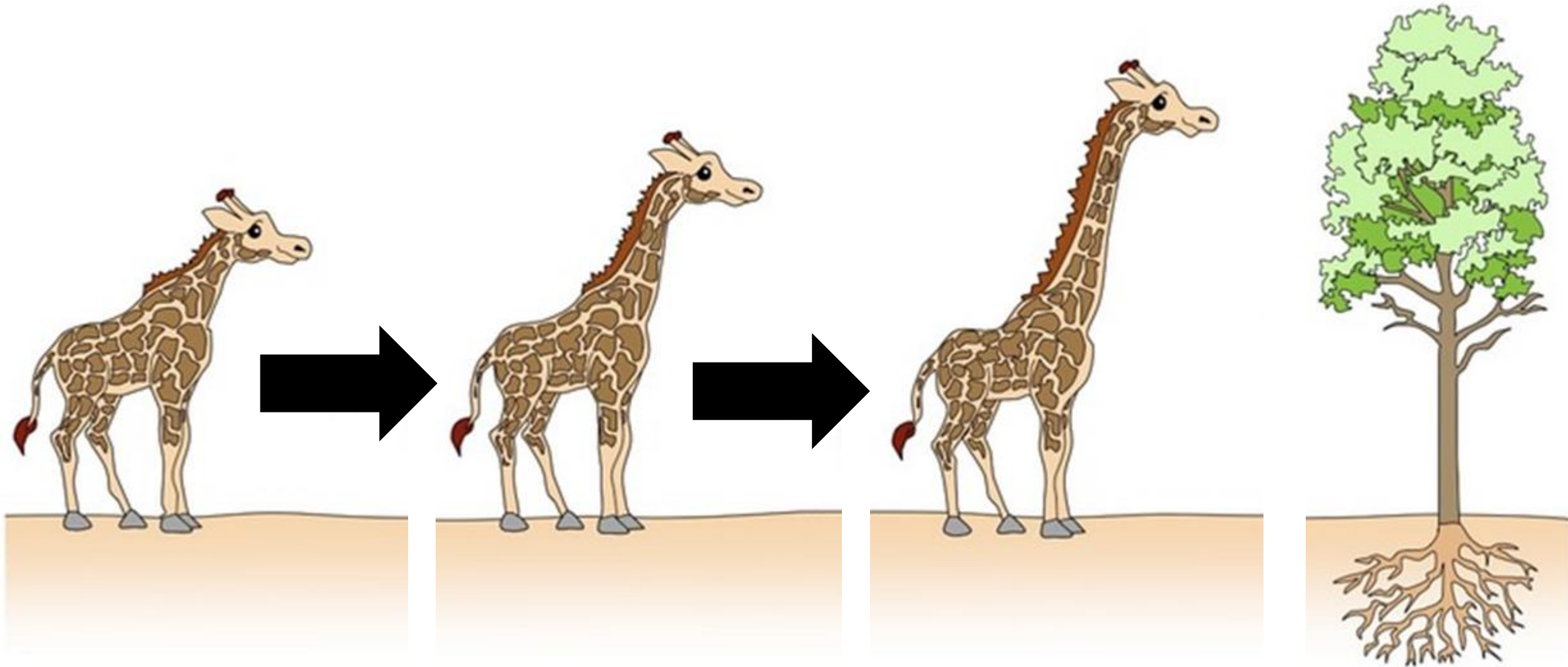
Algoritma Genetika (AG) adalah algoritma pencarian yang didasarkan pada parameter-parameter optimal.

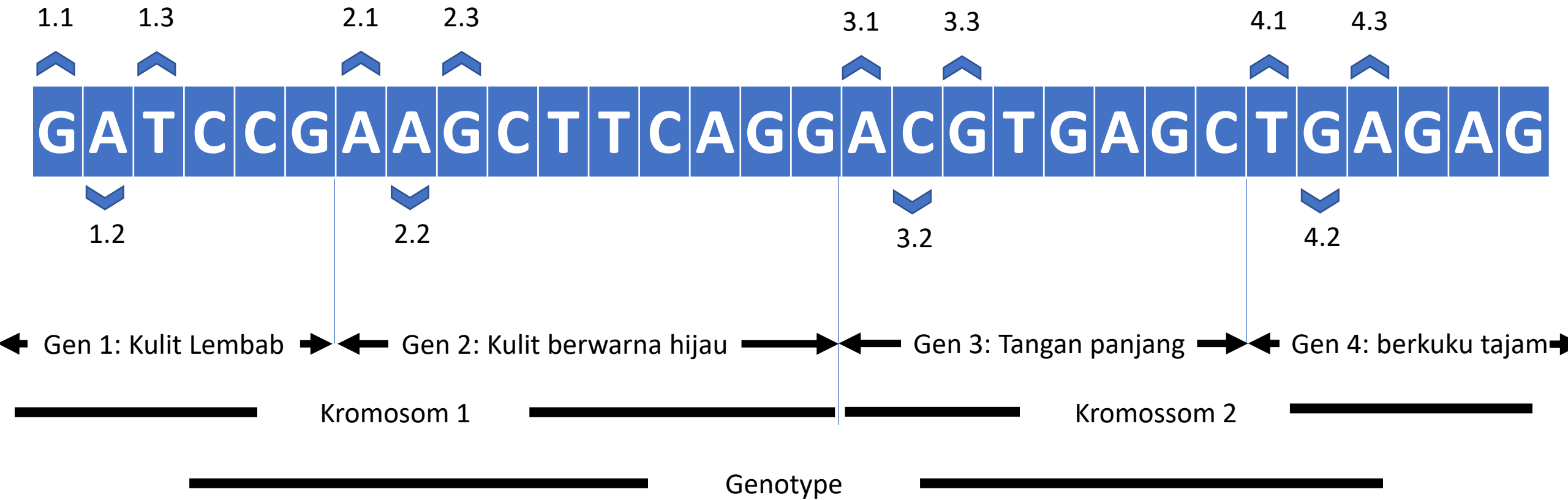


Seleksi Alamiah

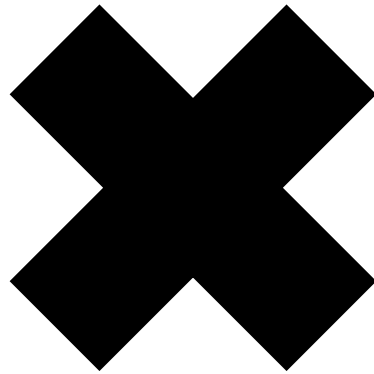


Genetika Alamiah

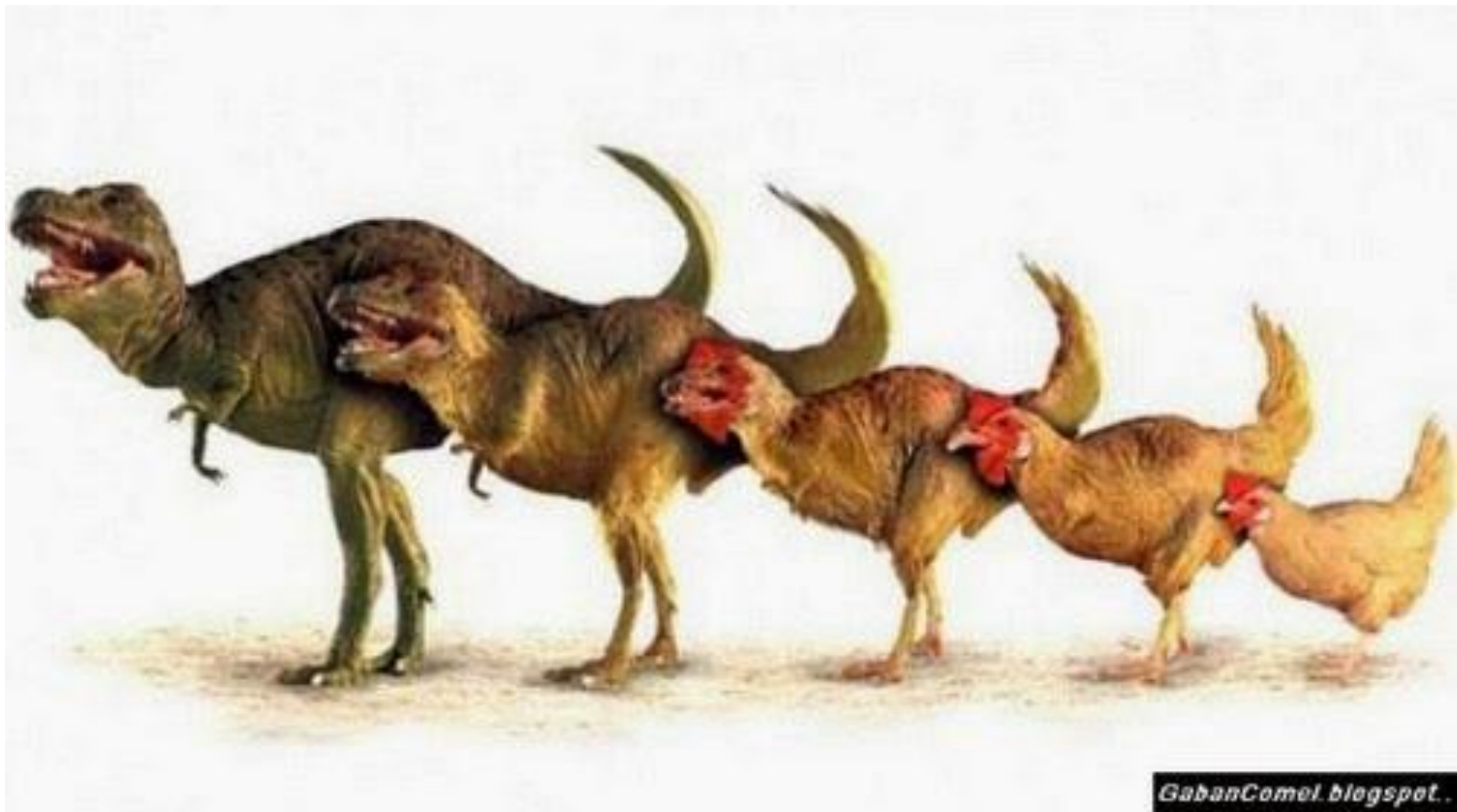




Pada setiap iterasi, proses evolusi yang terjadi adalah sebagai berikut:









Variasi Algoritma Genetika

Steady State

Bangkitkan populasi awal, P individu

Loop untuk P individu

 Dekodekan Individu

 Evaluasi individu

End

Loop sampai kondisi berhenti

 Pilih dua individu sebagai parent1 dan parent2

If perlu crossover **then**

 Offspring = Crossover(parent1, parent2)

End

If perlu mutasi **then**

 Offspring = Mutation(offspring)

End

If offspring lebih baik **then**

 Population = Replacement(population,offspring)

End

End

Genarational Replacement

```
Bangkitkan populasi awal, P individu
Loop sampai kondisi berhenti
    Loop untuk P individu
        Dekodekan Individu
        Evaluasi individu
    End
    Buat satu atau dua individu terbaik
    Loop sampai didapatkan P individu baru

        Pilih dua individu sebagai parent1 dan parent2
        If perlu crossover then
            Offspring = Crossover(parent1, parent2)
        End
        If perlu mutas then
            Offspring = Mutation(offspring)
        End
    End
End
```

Skema Pengkodean

Untuk dapat diproses menggunakan AG, suatu permasalahan harus dikonversi dulu ke dalam bentuk individu yang diwakili oleh satu atau lebih kromosom dengan kode tertentu. AG merepresentasikan gen (buatan), secara umum, sebagai bilangan real, desimal atau biner, yaitu:

- Real-number encoding.
- Discrete decimal encoding.
- Binary encoding.

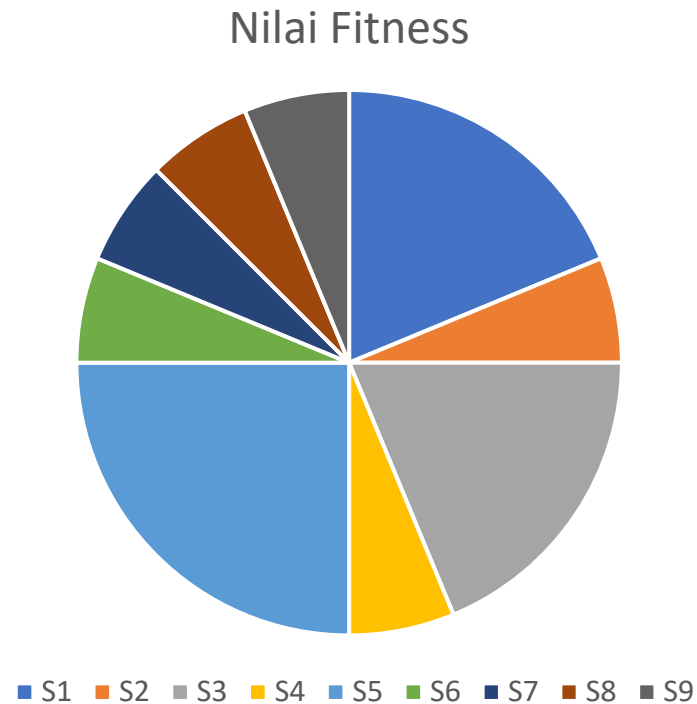
Nilai Fitness

Pada evolusi di dunia nyata, individu bernilai fitness tinggi akan bertahan hidup. Sedangkan individu bernilai fitness rendah akan mati. Pada AG, suatu individu dievaluasi berdasarkan suatu fungsi tertentu sebagai ukuran nilai fitness-nya.

Seleksi Orang Tua

Proses pemilihan dua individu sebagai orang tua biasanya dilakukan secara proporsional berdasarkan nilai-nilai fitness-nya. Salah satu metode seleksi yang umum digunakan adalah roulette-wheel.

String	Nilai Fitness
S1	0.6
S2	0.2
S3	0.6
S4	0.2
S5	0.8
S6	0.2
S7	0.2
S8	0.2
S9	0.2
Jumlah	3.2



**Anak tetangga
Pengusaha sukses**



**Anak tetangga
PNS**



Pindah Silang

Di dunia nyata, tidak mungkin ada dua individu yang sama. Hal ini disebabkan adanya evolusi yang di dalamnya terdapat proses pindah silang (crossover). Pada proses pindah silang terjadi kombinasi pewarisan gen-gen dari induknya, gen-gen dari kedua induk dapat bercampur sehingga dihasilkan susunan kromosom yang baru. Dari proses tersebut akan dihasilkan variasi genetik.

Mutasi

Mutasi diperlukan untuk mengembalikan informasi bit yang hilang akibat cross over. Mutasi diterapkan dengan probabilitas yang sangat kecil. Jika mutasi dilakukan terlalu sering, maka akan menghasilkan individu yang lemah karena konfigurasi gen pada individu yang unggul akan dirusak.

Elitisme

Karena seleksi dilakukan secara acak, maka tidak ada jaminan bahwa suatu individu bernilai fitness tertinggi akan selalu terpilih. Walaupun individu bernilai fitness tertinggi terpilih, mungkin saja individu tersebut akan rusak karena proses pindah silang. Untuk menjaga agar individu bernilai fitness tertinggi tersebut tidak hilang selama evolusi, perlu dibuat satu atau dua kopinya. Prosedur ini dikenal sebagai elitisme. Prosedur ini hanya digunakan pada AG berjenis generational replacement.

Penggantian Populasi

Pada AG berjenis generational replacement, N individu pada suatu generasi digantikan sekaligus oleh N individu baru hasil pindah silang dan mutasi. Untuk mempertahankan individu terbaik, diperlukan skema elitisme seperti dijelaskan di atas.

Kriteria Penghentian

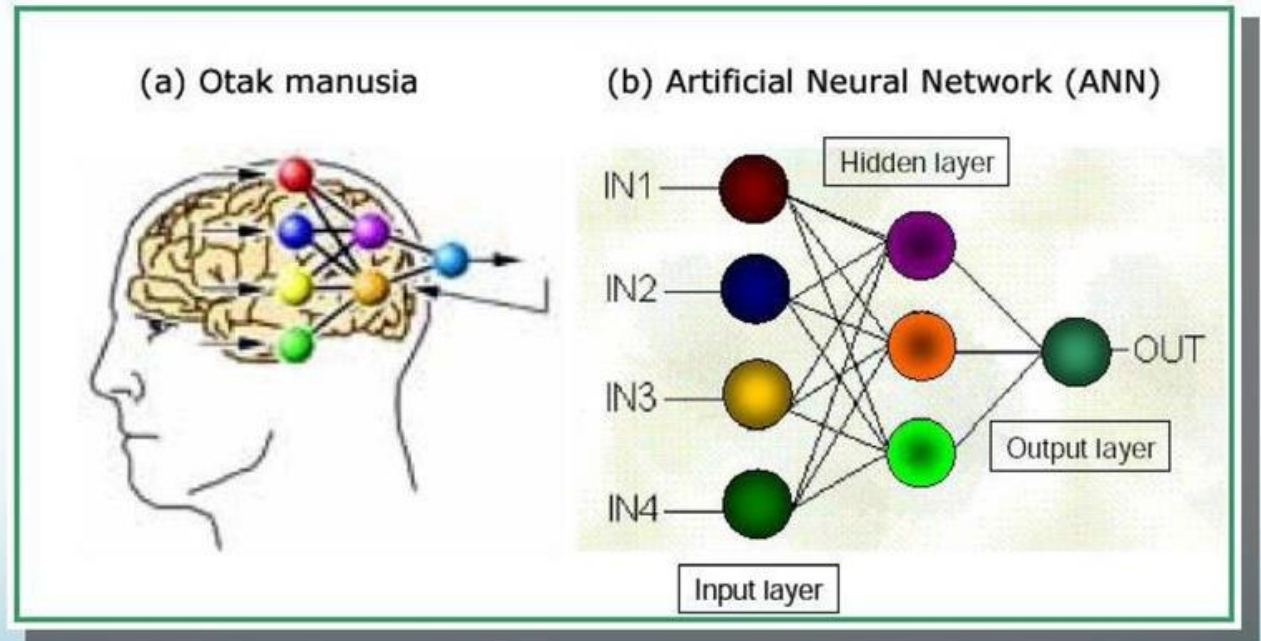
Terdapat berbagai macam kriteria penghentian yang bisa digunakan, tige diantaranya adalah:

- Memberikan batasan jumlah iterasi.
- Memberikan batasan waktu proses AG.
- Menghitung kegagalan penggantian anggota populasi yang terjadi secara berurutan sampai jumlah tertentu.

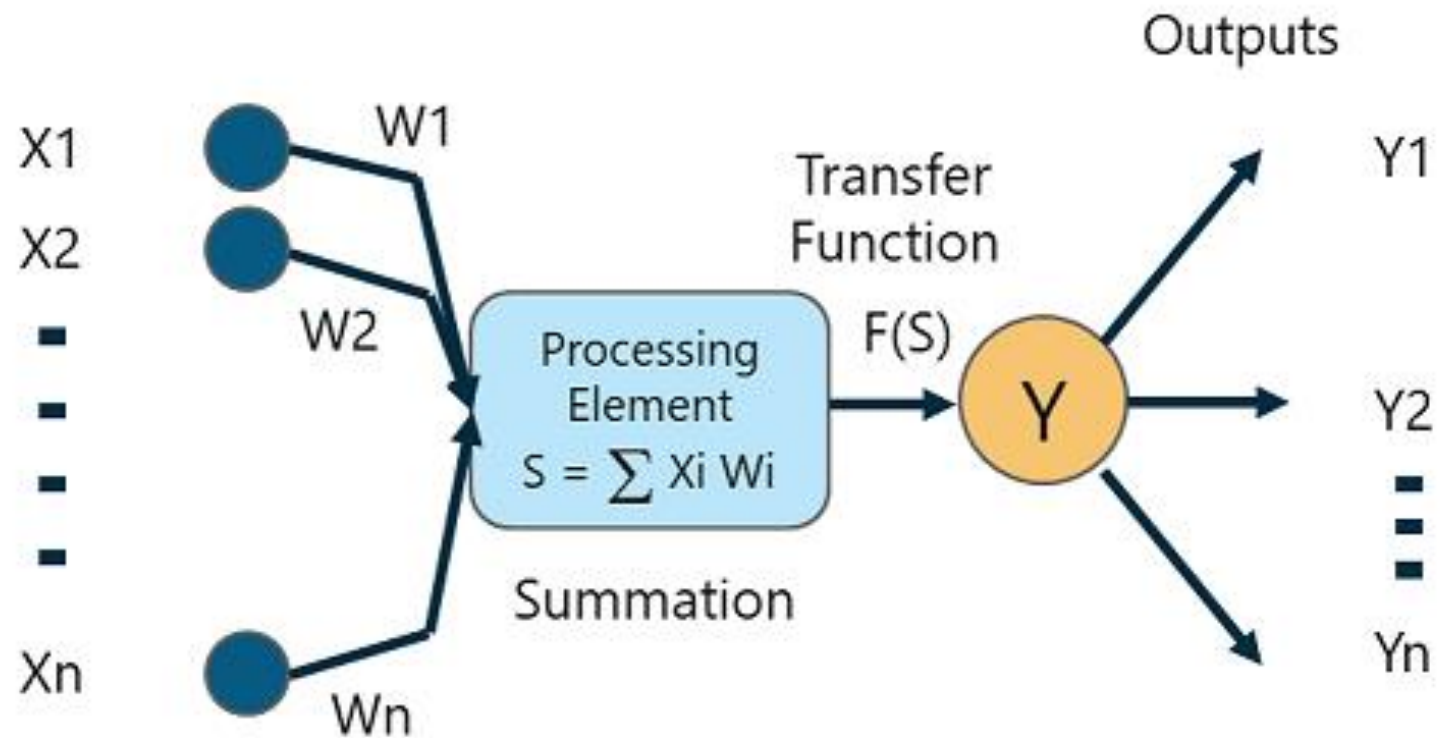
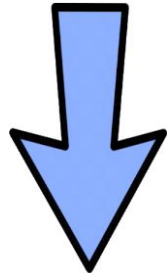
JARINGAN SYARAF TIRUAN?



JST (Jaringan Syaraf Tiruan)

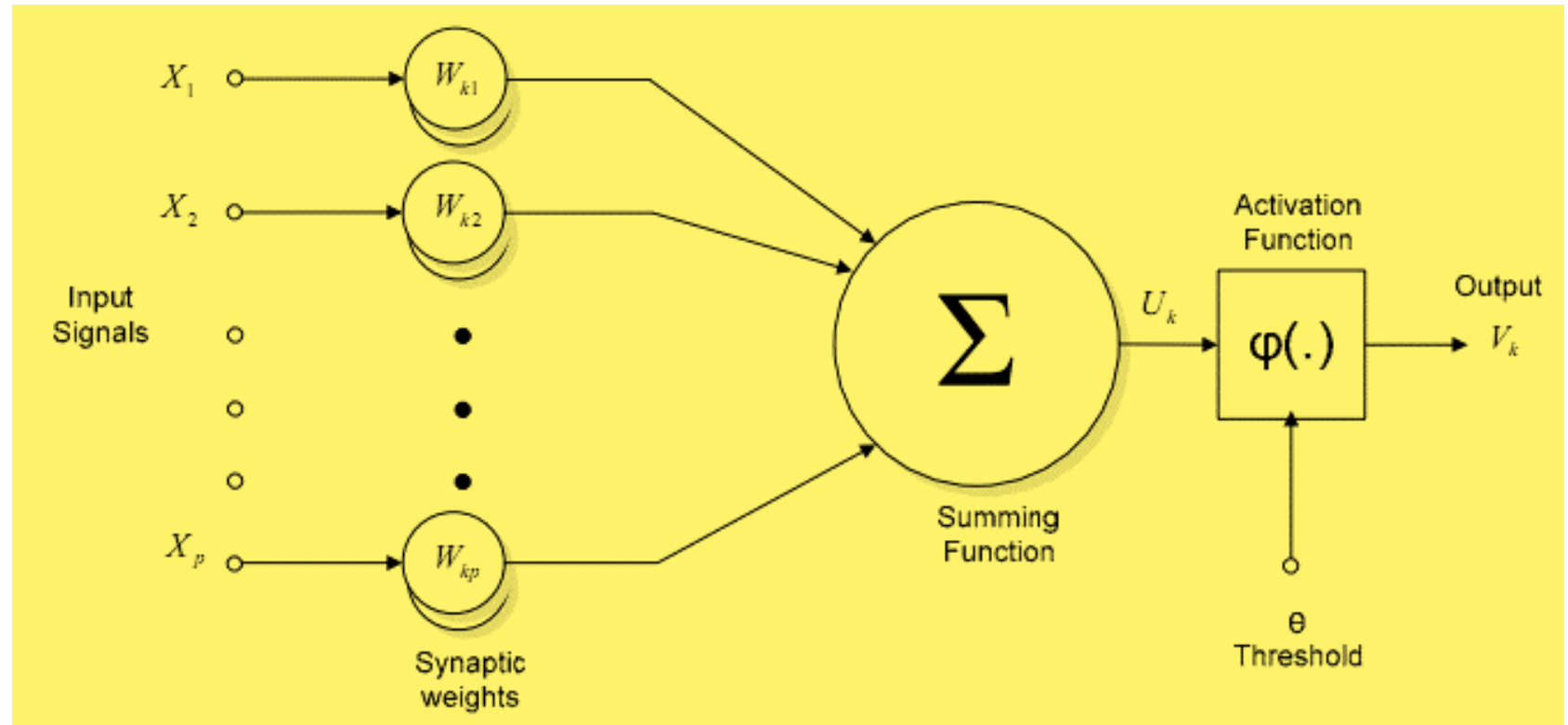
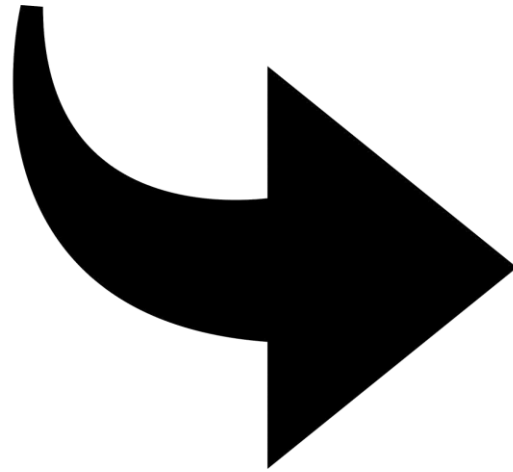


Model Sel Syaraf (Neuron)



Artificial Neural Network

Contoh Model Sel Syaraf (Neuron)



Gambar 5-19 : Model Matematis Nonlinier Dari Suatu Neuron



$$u_k = \sum_{j=1}^p w_{kj} x_j \quad (5.19)$$

$$y_k = \varphi(u_k - \theta_k), \quad (5.20)$$

$$v_k = u_k - \theta_k \quad (5.21)$$

$$v_k = \sum_{j=0}^p w_{kj} x_j \quad (5.22)$$

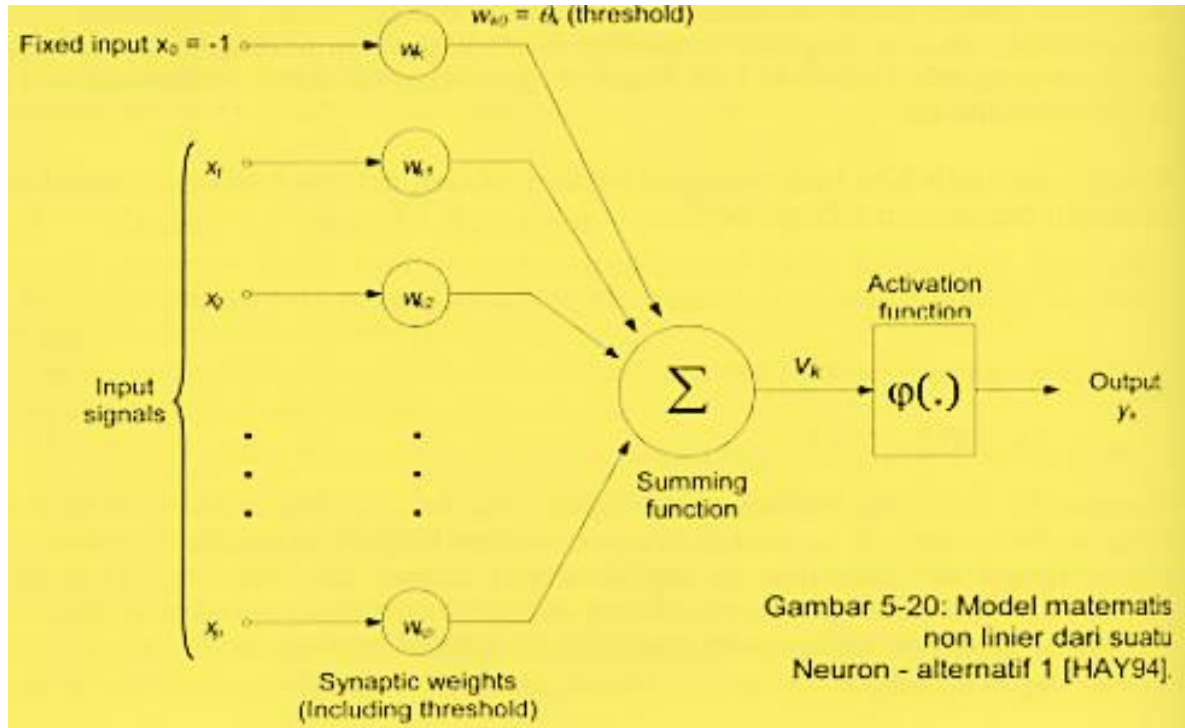
$$y_k = \varphi(v_k) \quad (5.23)$$

$$x_0 = -1. \quad (5.24)$$

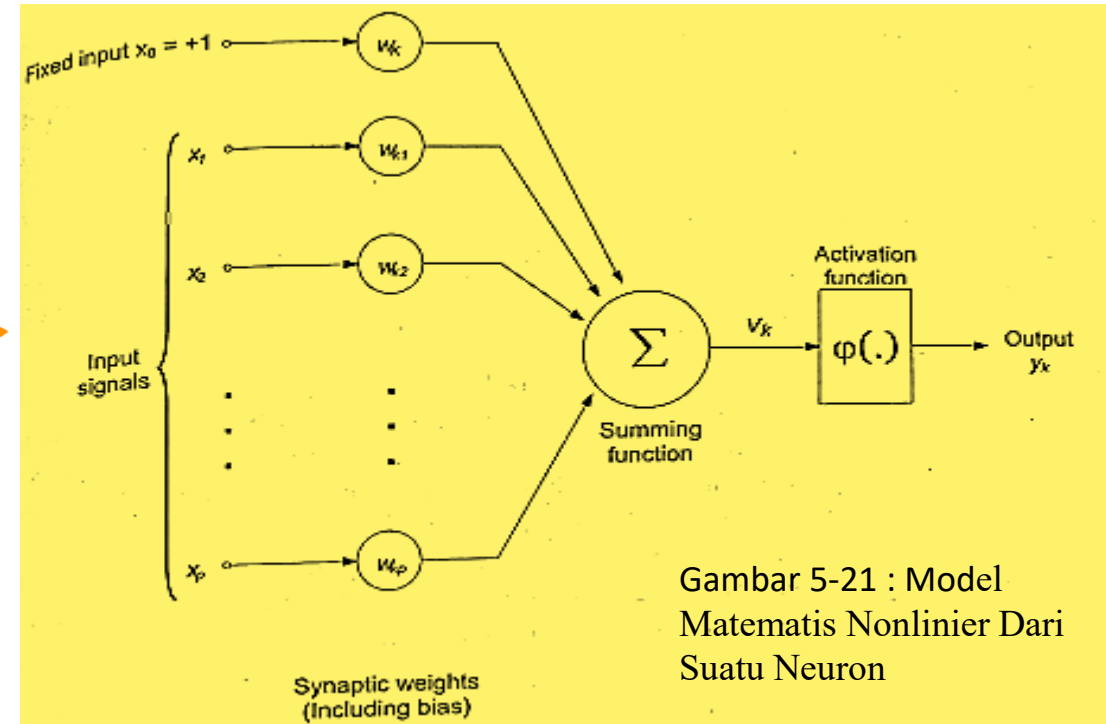
dan bobotnya adalah

$$w_{k0} = \theta_k \quad (5.25)$$

Model Matematis Nonlinier Dari Suatu Neuron

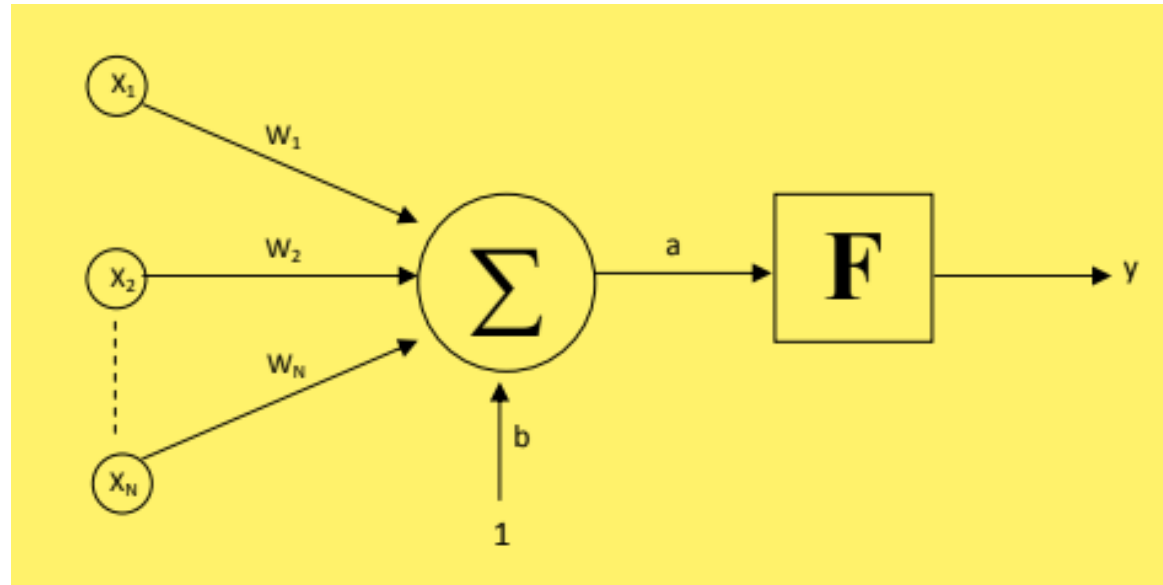


Gambar 5-20: Model matematis non linier dari suatu Neuron - alternatif 1 [HAY94].



Gambar 5-21 : Model Matematis Nonlinier Dari Suatu Neuron

Fungsi Aktivasi



1. *Threshold function*

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases}$$

(5.26)

2. *Piecewise-linear function*

$$\varphi(v) = \begin{cases} 1 & v \geq \frac{1}{2} \\ v & \frac{1}{2} > v > -\frac{1}{2} \\ 0 & v \leq -\frac{1}{2} \end{cases}$$

(5.27)

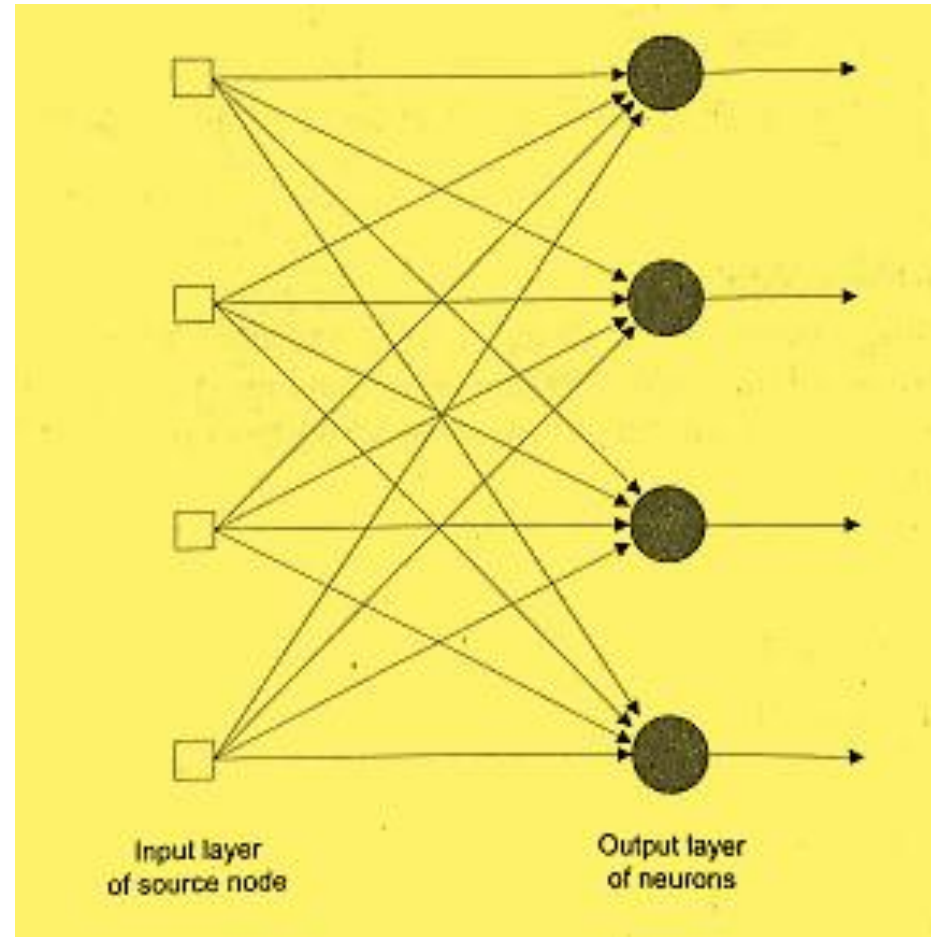
3. *Sigmoid function*

$$\varphi(v) = \frac{1}{1 + \exp(-av)}$$

(5.28)

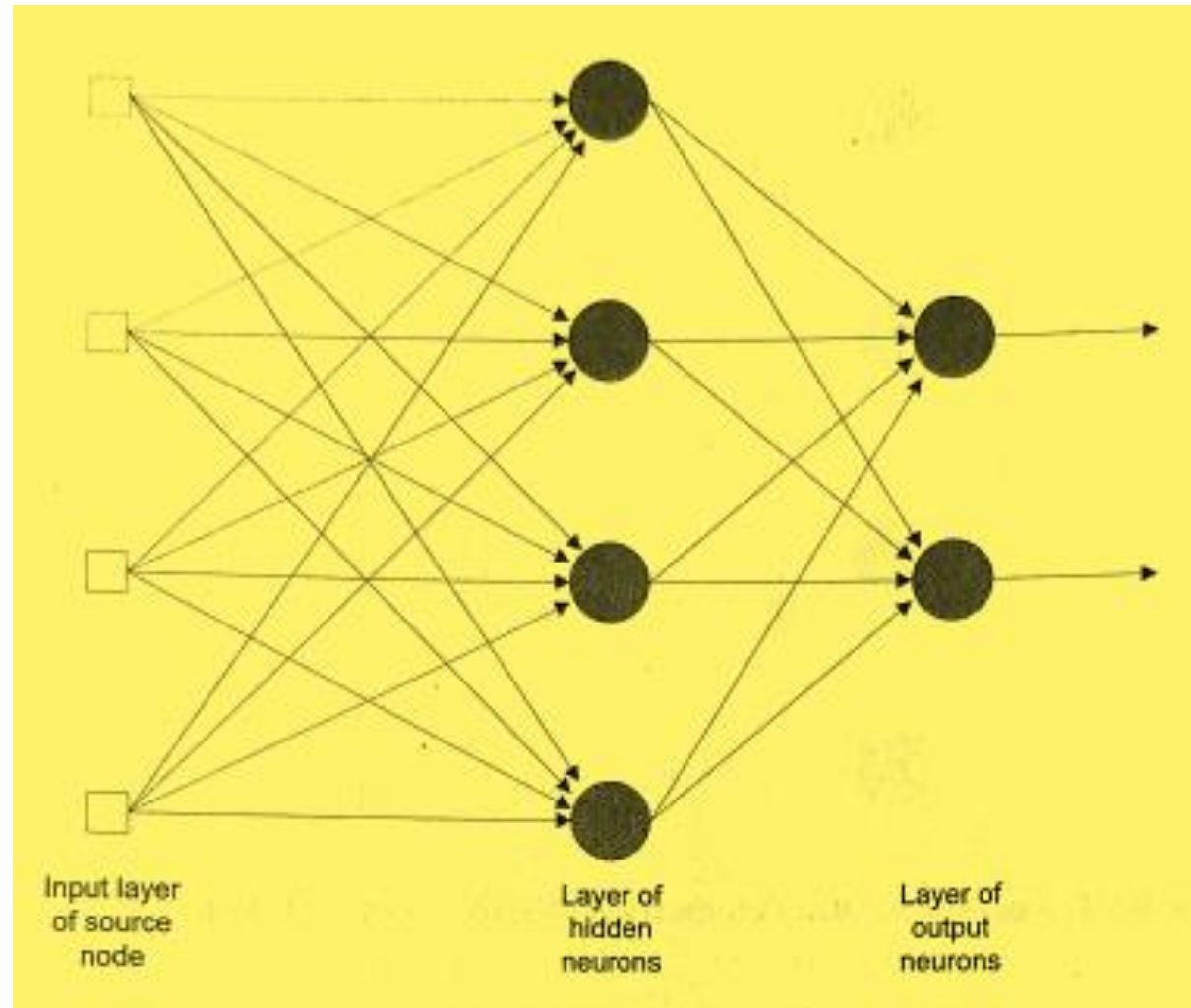
Arsitektur Jaringan

1. Single-Layer Feedforward Networks



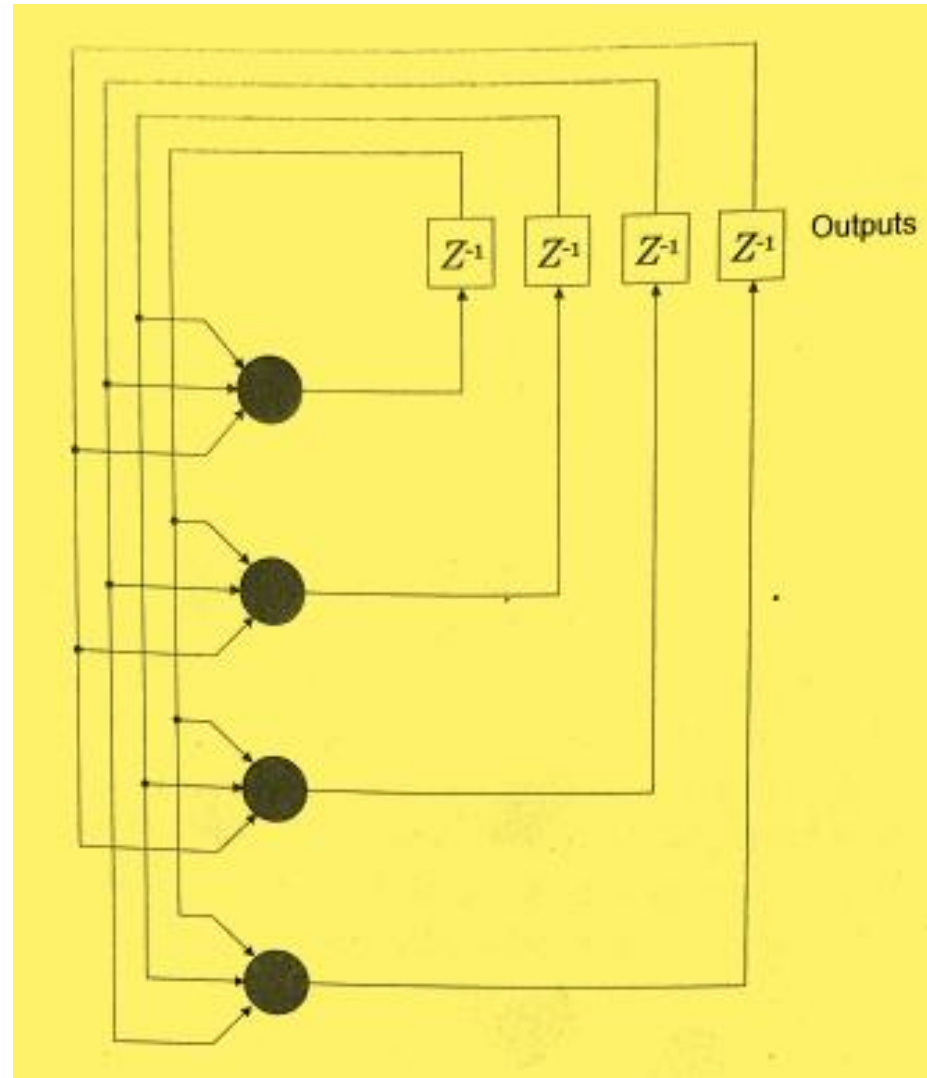
Gambar 5-22: Feedforward Network dengan satu lapisan neuron tunggal.

2. Multi-Layer Feedforward Networks

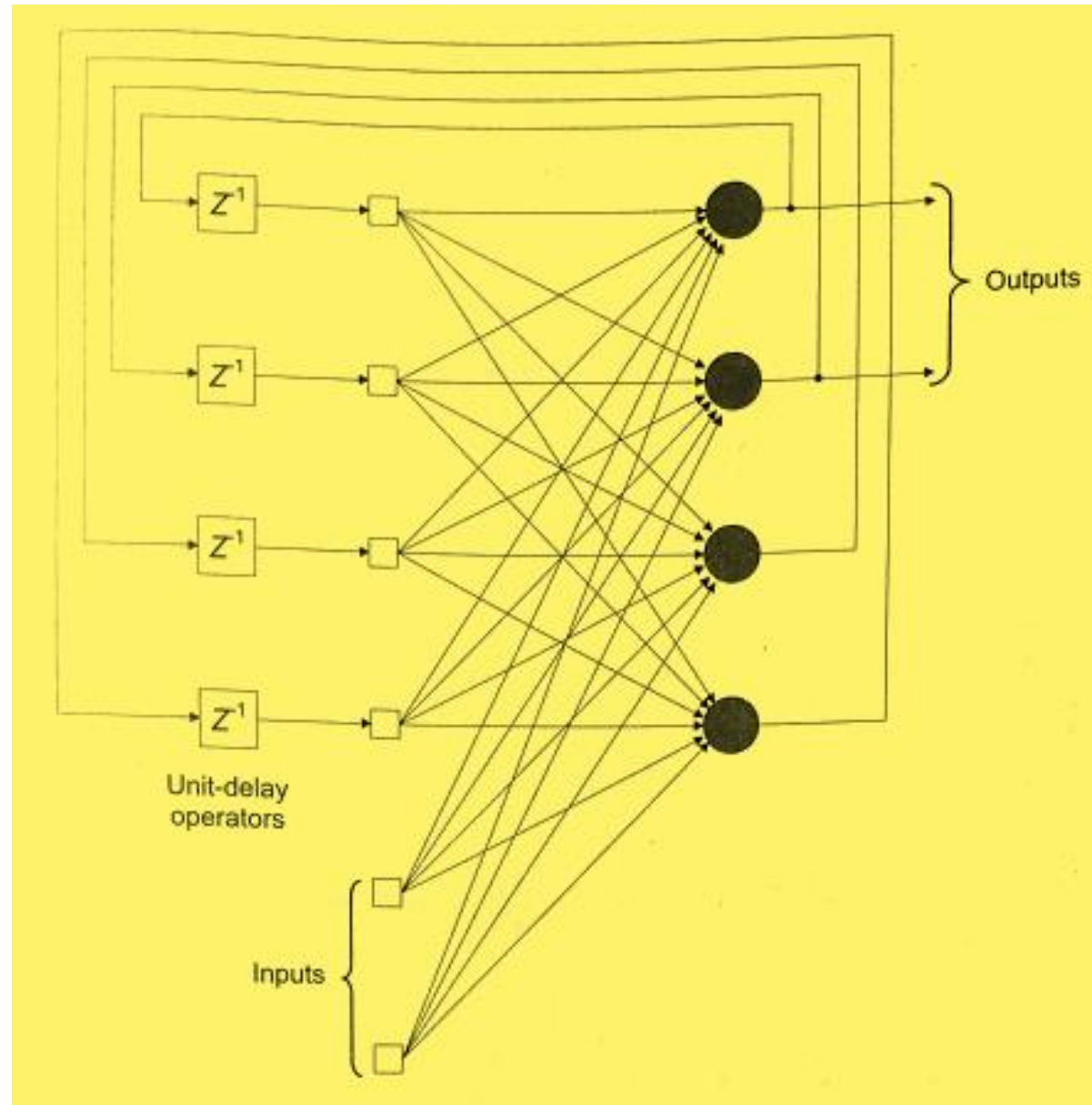


Gambar 5-23: Fully connected feedforward network dengan satu hidden layer dan satu output layer (jaringan 4-4-2)

3. Recurrent Networks

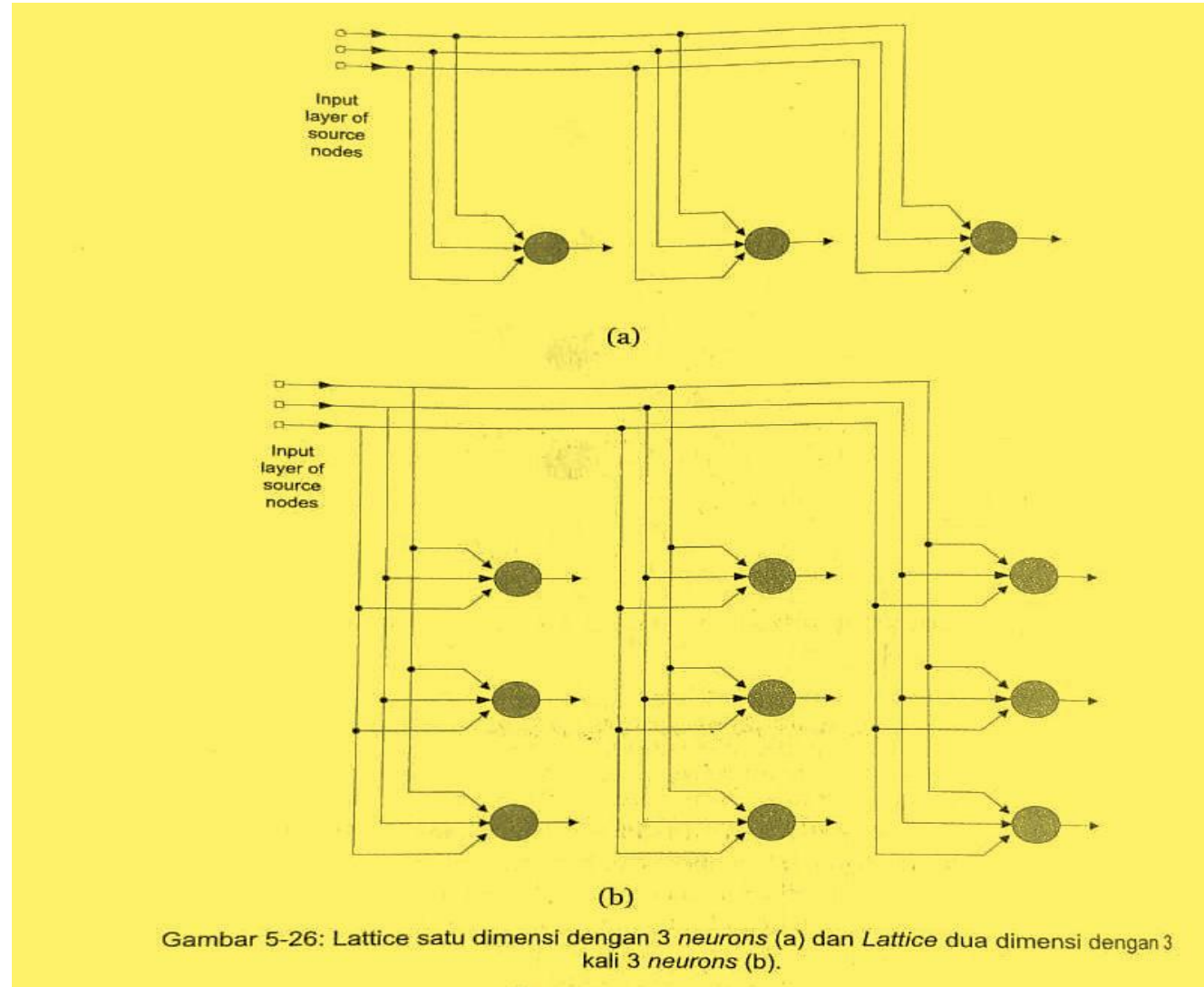


Gambar 5-24: Recurrent network tanpa self-feedback loop dan tanpa hidden neurons.



Gambar 5-25: Recurrent network dengan hidden neurons.

4. Lattice Structure



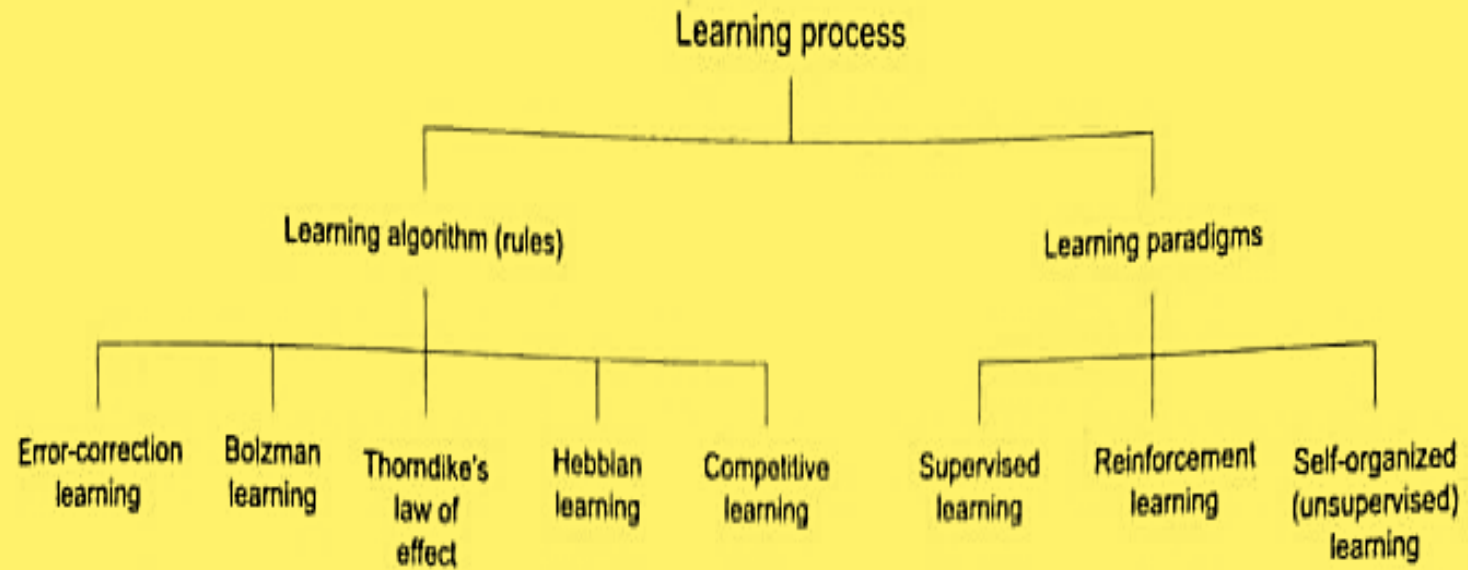
Proses Belajar

Defenisi belajar dalam konteks JST sebagai berikut :

Belajarliah adalah suatu proses dimana parameter-parameter bebas JST diadaptasikan melalui suatu proses perangsangan berkelanjutan oleh lingkungan dimana jaringan berada. Jenis belajar ditentukan oleh pola dimana penngubahan parameter dilakukan.

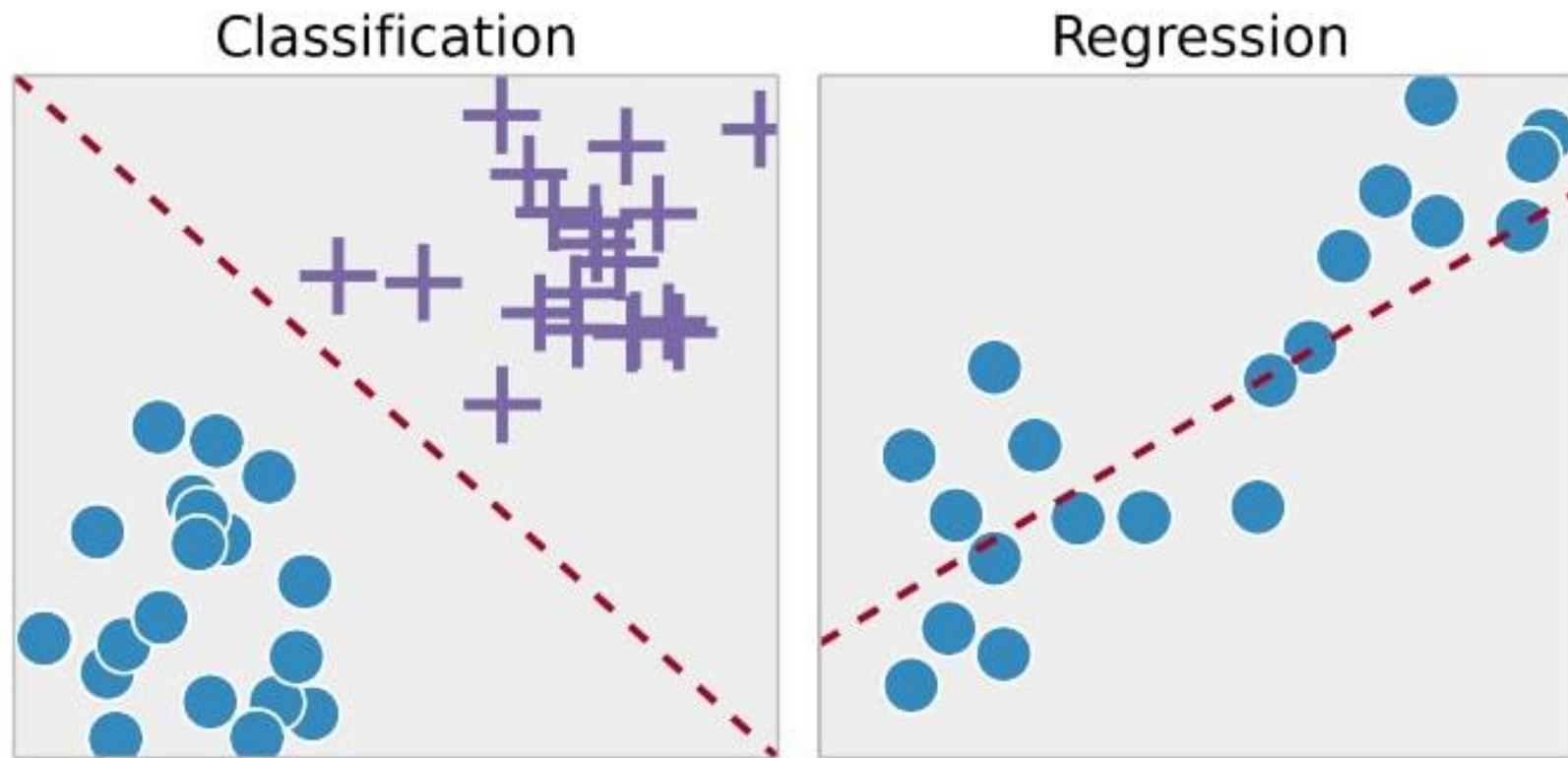
Dalam proses belajar terdapat kejadian-kejadian sebagai berikut :

- JST dirangsag oleh lingkungan
- JST mengubah dirinya sebagai hasil rangsangan ini.
- JST memberikan respon dengan caara yang baru kepada lingkungan, disebabkan perubahan yang terjadi dalam struktur internalnya sendiri.



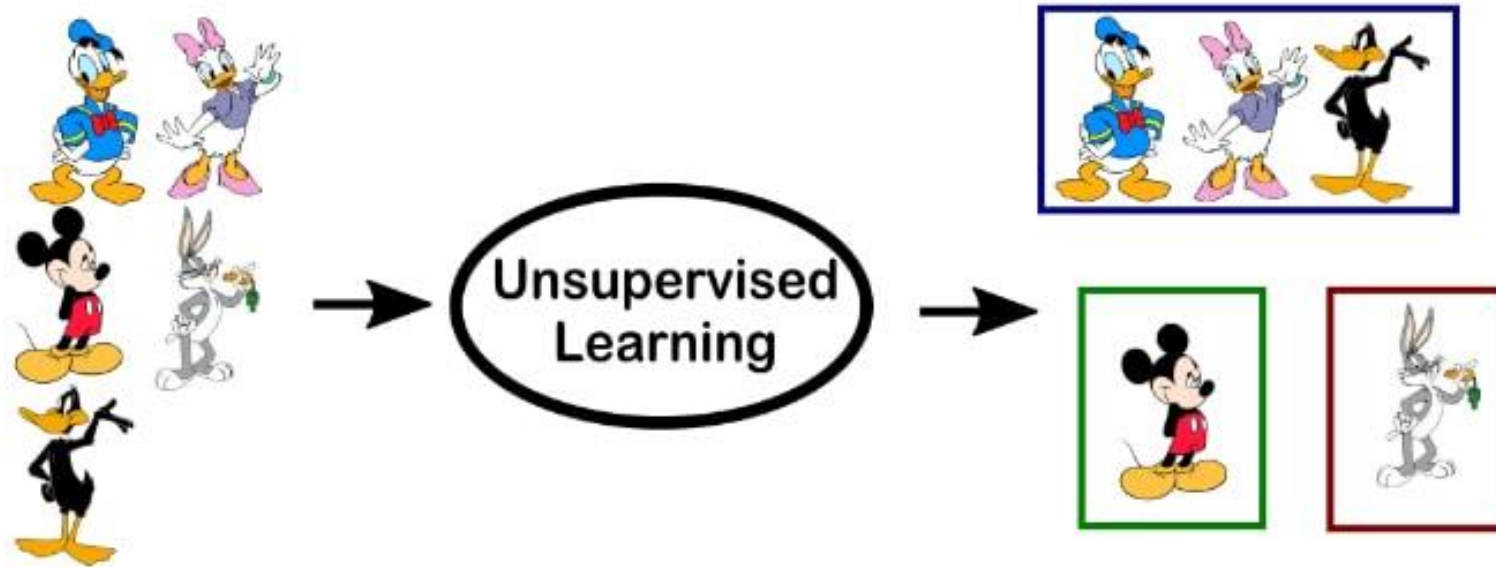
Gambar 5-27: Taksonomi proses belajar.

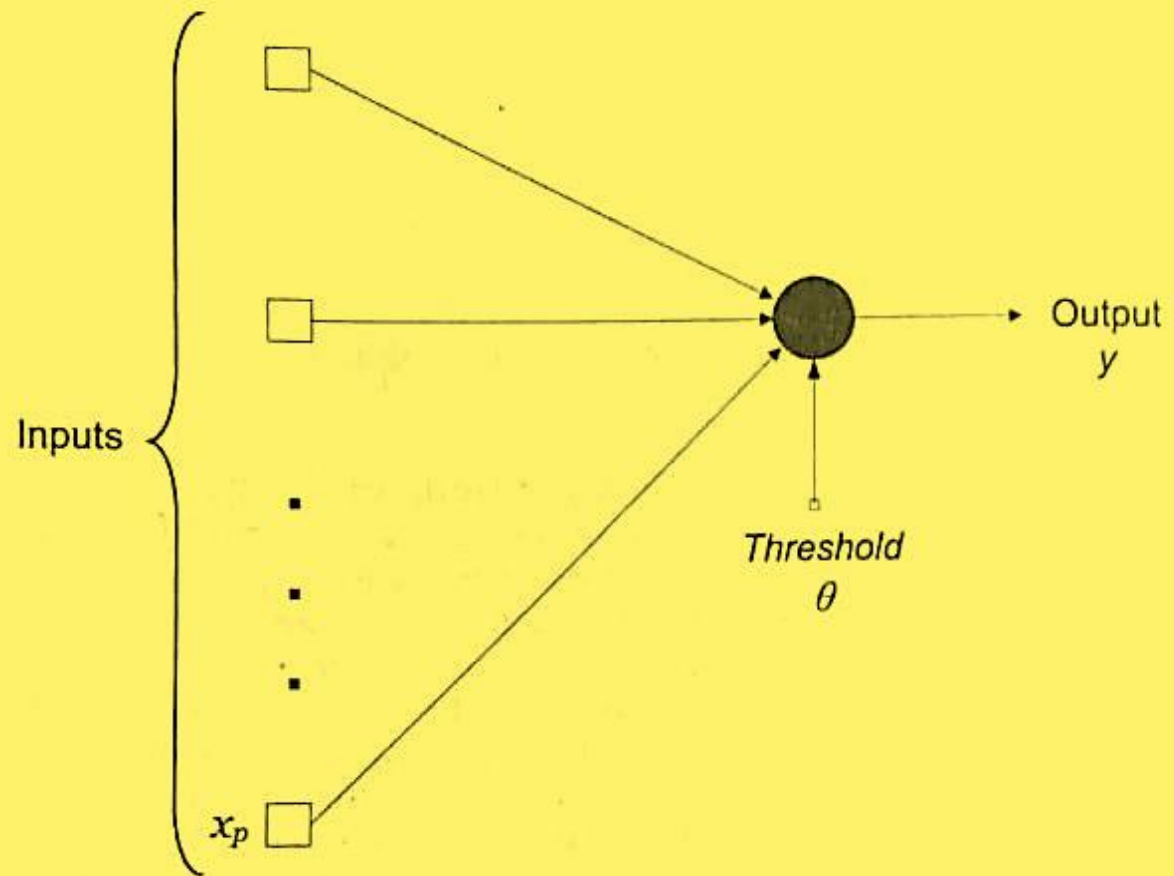
Supervised Learning:



Unsupervised Learning:

Unsupervised Learning





Gambar 5-28: *Single-layer perceptron* [HAY94].