

**Klasifikasi Data *Microarray* pada Penyakit Diabetes Melitus Tipe  
II dengan Metode *Functional Link Neural Network***

**SKRIPSI**



*Oleh:*

**LUKI DWIANTO SAPUTRA**

**H13114312**

**PROGRAM STUDI ILMU KOMPUTER DEPARTEMEN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN**

**2019**



Optimization Software:  
[www.balesio.com](http://www.balesio.com)

# **Klasifikasi Data *Microarray* pada Penyakit Diabetes Melitus Tipe II dengan Metode *Functional Link Neural Network***

## **SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Ilmu Komputer Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar



2019



## LEMBAR PERYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

### **Klasifikasi Data *Microarray* pada Penyakit Diabetes Melitus Tipe II dengan Metode *Functional Link Neural Network***

adalah benar hasil karya saya sendiri bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 17 Januari 2019

Luki Dwianto Saputra

NIM. H13114312



Optimization Software:  
[www.balesio.com](http://www.balesio.com)

**Klasifikasi Data *Microarray* pada Penyakit Diabetes Melitus Tipe  
II dengan Metode *Functional Link Neural Network***

**Disetujui Oleh :**

**Pembimbing Utama**

**Pembimbing Pertama**

**Dr. Eng. Armin Lawi, M.Eng**  
NIP.197204231995121001

**Sri Astuti Thamrin, S.Si,M.Stat.,Ph.D**  
NIP.197407131999032001



**Optimization Software:**  
[www.balesio.com](http://www.balesio.com)

## HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Luki Dwianto Saputra  
NIM : H13114312  
Program Studi : Ilmu Komputer  
Judul Skripsi : Klasifikasi Data Microarray pada Penyakit Diabetes Melitus Tipe II dengan Metode Functional Link Neural Network

Telah berhasil dipertahankan dihadapan dewan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

### DEWAN PENGUJI

#### Tanda Tangan

1. Ketua : Dr. Eng. Armin Lawi, M.Eng (.....)
2. Sekretaris : Sri Astuti Thamrin, S.Si, M.Stat., Ph.D (.....)
3. Anggota : Dr. Hendra, S.Si, M.Kom (.....)
4. Anggota : Prof. Dr. Moh. Ivan Azis, M.Sc. (.....)

Ditetapkan di : Makassar

Tanggal : 17 Januari 2019



## KATA PENGANTAR

Alhamdulillah, segala ptesting syukur kehadiran Allah SWT, atas segala nikmat dan karunia yang diberikan sehingga penulis dapat menyelesaikan tugas akhir ini. Salam dan shalawat senantiasa tercurahkan kepada Baginda Rasulullah Muhammad SAW, para sahabat dan orang-orang yang senantiasa mengikuti beliau hingga akhir zaman.

Dalam penyelesaian skripsi ini tidak sedikit hambatan dan kesulitan yang dialami penulis, namun berkat bantuan serta bimbingan dari berbagai pihak, skripsi ini dapat terselesaikan meski dengan segala kekurangannya. Karenanya pada kesempatan yang berharga ini, penulis secara khusus menyampaikan terima kasih kepada bapak **Dr. Hendra, S.Si.**, selaku Penasehat Akademik, bapak **Dr. Eng. Armin Lawi, M.Eng** sebagai pembimbing I, dan ibu **Sri Astuti Thamrin, S.Si, M.Stat. PhD.** selaku Pembimbing II, terima kasih atas segala curahan kasih sayang, bimbingan dan arahan yang diberikan kepada penulis selama mengikuti proses perkuliahan di program studi ilmu computer dan juga dalam penyusunan skripsi ini sampai selesai. Semoga Allah SWT membalas kebaikan Bapak dan Ibu dengan berlipat-lipat kebaikan, Aamiin.

Rasa terima kasih juga tercurahkan kepada seluruh keluarga penulis, utamanya orang tua tercinta Ayahanda **Roy Djaja Tanzil** dan Ibunda **Haryati Gali Tanzil** yang tak pernah berhenti memberi do'a, nasihat, semangat, dukungan, kasih sayang serta cinta yang tulus untuk kesuksesan anak-anaknya. Serta kakak tersayang **Cindy Monica Tanzil** dan adik tersayang **Devi Wulandari Anggreni Tanzil** yang senantiasa mendoakan dan memberi dukungan. Semoga Allah Yang Maha Pengasih, senantiasa memberikan rahmat-Nya atas kalian, orang-orang yang paling kucintai.

Tak lupa pula penulis menghaturkan ucapan terima kasih kepada :



**Rektor Universitas Hasanuddin** beserta jajarannya, Bapak **Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam** beserta jajarannya, dan seluruh staf birokrasi atas pengetahuan dan kemudahan-kemudahan yang diberikan, baik dalam bidang akademik maupun bidang kemahasiswaan.

2. Bapak Prof. **Dr. Amir Kamal Amir, M.Sc.**, selaku Ketua Jurusan Matematika, dan Bapak **Dr. Amran, S.Si., M.Si.**, selaku Sekretaris Jurusan, serta Bapak **Dr. Diaraya, M.Ak.**, selaku Kepala Program Studi Ilmu Komputer yang telah memberikan banyak bantuan selama penulis menjalani pendidikan. Terima kasih juga untuk segenap jajaran Pegawai Akademik Jurusan Matematika atas bantuannya dalam pengurusan akademik selama ini.
3. Bapak **Dr. Eng. Armin Lawi, M.Eng.**, selaku pembimbing utama, atas segala ilmu, nasehat, dan kesabaran dalam membimbing penulis serta meluangkan waktu di sela-sela rutinitas yang begitu padat hingga skripsi ini dirampungkan, dan Ibu **Sri Astuti Thamrin, S.Si, M.Stat. PhD.**, selaku pembimbing pertama, untuk segala ilmu, nasehat, dan kesabaran dalam membimbing dan mengarahkan penulis, serta bersedia meluangkan waktunya untuk mendampingi penulis sejak awal penyusunan hingga akhir perampungan skripsi ini.
4. Bapak **Dr. Hendra, S.Si., M.Kom.** , selaku penguji sekaligus Penasehat Akademik (PA) penulis untuk segala ilmu, nasehat, saran dan motivasi yang diberikan kepada penulis mulai dari perkuliahan hingga penyusunan skripsi ini, dan bapak **Prof. Dr. Moh. Ivan Azis, M.Sc.** selaku penguji yang telah bersedia meluangkan waktu untuk memberikan saran dan arahan kepada penulis dalam penyusunan skripsi ini.
5. Kepada Senior **Kak Octa, Kak Yudha, Kak Edy,** dan **Kak Akas** yang telah menghibur penulis ketika depresi, dengan bermain Dota bersama – sama.
6. Kepada Ibu **Nurhilal Syahrir,** dan Bapak **Supri Bin Amir** atas arahannya dan bimbingannya kepada penulis dalam penyusunan skripsi ini
7. Teman-teman seperjuangan **Ilmu Komputer 2014 (Nadya, Sukma, Darul, Jo, Firda, Iyam, Anita Ola, Ayu, Yaumil, Fuad, Dila, Danti, Yuyu, Nuhi, Nanda, Ica, Syam, Aspar, Nurul, dll)** yang membantu dan memberi support penulis dalam penyusunan skripsi ini.
8. Kepada teman – teman KKN gelombang 96 reguler kecamatan **Segeri** yang telah memberi kesan dan momen membahagiakan selama KKN

pada **Winda Dewi Mulia Setiawaty** yang telah memberi semangat, memberi support, dan meluangkan waktunya untuk membantu penulis dalam penyusunan skripsi ini.

k-Adik **Ilmu Komputer 2015, 2016, 2017, dan 2018.**



11. Semua pihak yang telah banyak berpartisipasi, baik secara langsung maupun tidak langsung dalam penyusunan skripsi ini yang tak sempat penulis sebutkan satu per satu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan, dan karena itu kritikan dan saran yang sifatnya membangun sangat penulis harapkan demi kesempurnaan skripsi ini.

Makassar, 17 Januari 2019

Luki Dwianto Saputra



**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK  
KEPENTINGAN AKADEMIK**

---

Sebagai civitas akademik Universitas Hasanuddin saya yang bertanda tangan di bawah ini:

Nama : Luki Dwianto Saputra

NIM : H 131 14 312

Program Studi : Ilmu Komputer

Departemen : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Prediktor Royalti Non-eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

**“Klasifikasi Data *Microarray* pada Penyakit Diabetes Melitus Tipe II dengan  
Metode *Functional Link Neural Network*”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal, 17 Januari 2019

Yang menyatakan,

Luki Dwianto Saputra



## ABSTRAK

Jumlah pasien diabetes mellitus tipe 2 meningkat setiap tahun. Sebagian besar kasus diabetes melibatkan banyak gen, yang menyumbang sedikit pengaruh pada peningkatan kemungkinan diabetes mellitus tipe 2. Penggunaan data microarray sangat berguna untuk melakukan penelitian pada suatu penyakit. Namun, data microarray dimensi tinggi dapat mengarah pada metode yang digunakan untuk proses klasifikasi, membutuhkan waktu perhitungan yang sangat lama, dan kurang akurat. Dalam tulisan ini, metode pemilihan fitur akan digunakan, yaitu uji-t untuk memilih gen informatif. Gen informatif yang dipilih kemudian diklasifikasikan menggunakan metode *Functional Link Neural Network*. Hasil klasifikasi dapat dilihat dari nilai akurasi yaitu 98,3%.

**Kata Kunci :** *Diabetes Melitus Tipe 2, Microarray, Functional Link Neural Network, t-test, Legendre Polynomial*



## ABSTRAK

The number of Type 2 diabetes mellitus patients increases every year. Most of diabetes cases involve many genes, which contribute such a little influence on increasing the probability of type 2 diabetes mellitus. The use of microarray data is very useful for conducting research in a disease. However, the high dimensions of microarray data can lead to methods used for classification process, requires a very long computation time, and less accuracy. In this paper, the feature selection method will be used, namely the t-test to select informative genes. The informative gene selected was then classified using the *functional link neural network* method. The classification results can be seen from the value of accuracy is 98,3%.

**Kata Kunci :** *Diabetes Melitus Tipe 2, Microarray, Functional Link Neural Network, t-test, Support Vector Machine*



## DAFTAR ISI

LEMBAR PERNYATAAN KEOTENTIKAN.....	iii
HALAMAN PENGESAHAN.....	<b>Error! Bookmark not defined.</b>
KATA PENGANTAR.....	vi
PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	ix
ABSTRAK.....	x
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	1
1.1    LATAR BELAKANG.....	1
1.2    RUMUSAN MASALAH.....	2
1.3    BATASAN MASALAH.....	2
1.4    TUJUAN PENELITIAN.....	2
1.5    MANFAAT PENELITIAN.....	2
BAB II TINJAUAN PUSTAKA.....	3
2.1    Landasan Teori.....	3
2.1.1    Diabetes Melitus Tipe II.....	3
2.1.2    Microarray Data.....	3
2.1.3    Min-Max Standardization.....	5
2.1.4    Quantile Normalization.....	5
2.1.5    t-test.....	6
2.1.6    Artificial Neural Network (ANN).....	7
2.1.7    Functional Link Neural Network.....	13
2.1.8    Cross Validation.....	14
2.1.9    Confusion Matrix.....	15
2.1.10    ROC (Receiver Operating Characteristic) Curve.....	16
2.2    Kerangka Konseptual.....	18
BAB III METODOLOGI PENELITIAN.....	19
Sumber Data.....	19
Identifikasi Variabel.....	19



3.3	Metode Analisis.....	19
3.4	Diagram Alir .....	20
BAB IV .....		21
HASIL DAN PEMBAHASAN .....		21
4.1	Seleksi fitur terbaik pada data MicroArray.....	21
4.1.1	Pengambilan Data MicroArray .....	21
4.1.2	Preprocessing .....	21
4.1.3	Seleksi Fitur .....	24
4.2	Peforma Functional Link Neural network .....	25
4.2.1	Membuat Model Functional Link Neural Network.....	25
4.2.2	Uji model Functional Link Neural Network .....	27
4.2.3	Running Time dan Epoch .....	28
4.2.4	Hasil Klasifikasi.....	28
4.2.5	Cross Validation.....	29
4.3	Pembahasan.....	30
BAB V KESIMPULAN DAN SARAN.....		33
5.1	Kesimpulan.....	33
5.2	Saran .....	33
DAFTAR PUSTAKA .....		34



## DAFTAR GAMBAR

Gambar 1	Cara Kerja Teknologi Microarray .....	4
Gambar 2	Hasil Interpretasi Teknologi DNA Microarray[15].....	4
Gambar 3	Data sebelum Normalisasi Quantile .....	5
Gambar 4	Data setelah Normalisasi Quantile .....	6
Gambar 5	Prinsip jaringan saraf tiruan [14].....	8
Gambar 6	Jaringan dengan lapisan tunggal (single layer net).....	9
Gambar 7	Jaringan dengan banyak lapisan (multilayer net).....	10
Gambar 8	Jaringan dengan lapisan kompetitif (competitive layer net) .....	10
Gambar 9	Functional Link Neural Network.....	13
Gambar 10	Skema 10 k <i>fold cross validation</i> .....	15
Gambar 11	Confusion Matrix .....	16
Gambar 12	Contoh distribusi salah satu fitur tidak normal .....	22
Gambar 13	Contoh disribusi salah satu fitur setelah standarisasi .....	23
Gambar 14	Grafik distribusi pada setiap fitur .....	24
Gambar 15	Running time model Functional Link Neural Network .....	31
Gambar 16	Nilai akurasi Functional Link Neural Network.....	32



## DAFTAR TABEL

Tabel 1 Nilai P-Value 50 fitur terbaik .....	24
Tabel 2 Running time pada 10 model Functional Link Neural Network.....	28
Tabel 3 Hasil Klasifikasi dengan 70% data latih dan 30% data uji.....	28
Tabel 4 Hasil Klasifikasi dengan 80% data latih dan 20% data uji.....	29
Tabel 5 Hasil Cross Validasi pada setiap dataset .....	29
Tabel 6 Simbol Gen .....	30



## DAFTAR LAMPIRAN

Lampiran 1	Ambil package .....	37
Lampiran 2	Proses ambil data .....	37
Lampiran 3	Uji normalitas .....	37
Lampiran 4	Proses standarisasi Min-Max .....	37
Lampiran 5	Normalisasi quantile.....	38
Lampiran 6	Seleksi Fitur .....	38
Lampiran 7	Ekspansi.....	38
Lampiran 8	Klasifikasi data 70% data latih dengan 30% data uji .....	41
Lampiran 9	Proses klasifikasi 80% data latih dengan 20% data uji .....	42
Lampiran 10	Plot grafik hasil klasifikasi .....	43
Lampiran 11	Cross Validasi .....	44



# BAB I

## PENDAHULUAN

### 1.1 LATAR BELAKANG

DM tipe 2 yang dahulu disebut diabetes melitus tidak tergantung insulin (*non-insulin-dependent diabetes melitus/NIDDM*) atau diabetes onset dewasa merupakan kelainan metabolik yang ditandai dengan kadar glukosa darah yang tinggi dalam konteks resistensi insulin dan defisiensi insulin relatif [1]. Sebagian besar kasus diabetes melibatkan banyak gen yang masing-masing menyumbangkan pengaruh yang kecil terhadap meningkatnya kemungkinan terjadi diabetes tipe 2. Bila salah satu dari pasangan kembar identik menderita diabetes maka peluang seumur hidup saudara kembarnya terkena diabetes adalah lebih dari 90% sedangkan untuk pasangan kembar tidak identik hanya 25-50% . Secara global pada tahun 2010 diperkirakan terdapat 285 juta penderita diabetes tipe 2 yang mencakup 90% dari kasus diabetes. Untuk mengurangi peningkatan penderita diabetes melitus (DM), maka perlu mendiagnosa lebih awal dengan deteksi penyakit DM untuk penanganan lebih cepat [2].

Data microarray sering digunakan dalam beberapa penelitian untuk mengetahui hasil klasifikasi atau mendiagnosis suatu kelas penyakit. Banyaknya atribut yang terdapat dalam data microarray akan mempersulit dalam proses klasifikasi, dan memerlukan waktu yang relatif lama dalam proses komputasi oleh karena itu perlu dilakukan pemilihan feature terbaik untuk memperoleh hasil klasifikasi yang lebih tinggi [3].

Penelitian mengenai deteksi penyakit dengan data *microarray*, berkaitan dengan penelitian-penelitian lain. Parry dkk.[5] menggunakan metode *K-Nearest Neighbour* (KNN) dalam deteksi penyakit kanker payudara. Adapun Kumar [4] menggunakan metode *Functional Link Neural Network* dengan metode seleksi fitur yang digunakan adalah *t-test* . Zhang dkk [6] menggunakan metode *discrimantive area* untuk identifikasi biomarker dari penyakit diabetes melitus tipe 2 maka diperoleh keakuratan biomarker dengan 10 kali lipat cross-validasi

ata 84.79% dari 5 dataset berbeda dan variansi 0.054.



*Functional Link Neural Network* merupakan salah satu model *Artificial Neural Network* (ANN) yang mampu menyelesaikan tugas klasifikasi dengan keberhasilan yang cukup signifikan karena menggunakan arsitektur datar atau layar tunggal, yang mampu mengoptimasi proses, meningkatkan kemampuan diskriminasi terhadap masukan yang irrelevan, insignifkan, dan *noisy*. Sebelumnya, dalam penelitian yang dilakukan oleh Kumar dkk [4] dengan Akurasi 97.22%,98.42%,85.57%.

## 1.2 RUMUSAN MASALAH

1. Bagaimana Mendapatkan Fitur Terbaik pada data microarray?
2. Bagaimana Peforma Metode *Functional Link Neural Network* pada data microarray?

## 1.3 BATASAN MASALAH

Ruang lingkup penelitian adalah mengelola data *microarray* untuk deteksi penyakit diabetes melitus tipe 2 pada data GSE18732. Metode yang digunakan untuk seleksi fitur adalah *t-test*. Banyaknya fitur yang diseleksi pada penelitian ini adalah sebanyak 50 fitur. Metode yang digunakan untuk klasifikasi adalah *Functional Link Neural Network*.

## 1.4 TUJUAN PENELITIAN

Dengan memperhatikan latar belakang dan rumusan masalah di atas maka tujuan dari penelitian ini adalah:

1. Mendapatkan fitur terbaik pada data microarray.
2. Mengetahui performa metode *Functional Link Neural Network* untuk data microarray diabetes melitus tipe 2.

## 1.5 MANFAAT PENELITIAN

Hasil penelitian ini diharapkan berguna untuk dapat menambah pemahaman mengenai data *microarray* dan metode *Functional Link Neural Network* bagi pembaca dan dapat dikembangkan. Selain itu, penelitian ini juga dapat menjadi masukan dalam penyakit diabetes melitus tipe 2.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Landasan Teori

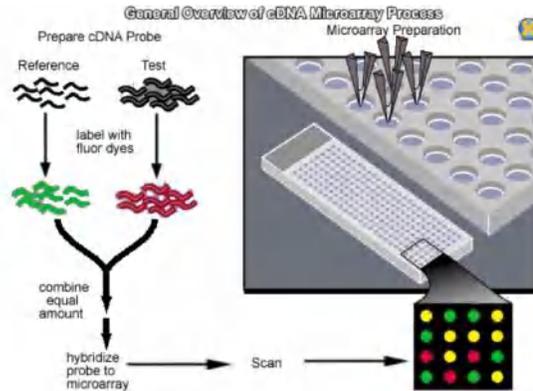
##### 2.1.1 Diabetes Melitus Tipe II

Diabetes Melitus (DM) merupakan penyakit kronik yang terjadi ketika pankreas tidak dapat lagi memproduksi insulin dalam jumlah yang cukup atau dapat juga disebabkan oleh berkurangnya kemampuan tubuh untuk merespon kerja insulin secara efektif. Insulin adalah hormon yang berfungsi untuk mengatur kadar gula darah. Peningkatan kadar gula dalam darah atau hiperglikemia merupakan gejala umum yang terjadi pada diabetes dan seringkali mengakibatkan kerusakan-kerusakan yang cukup serius pada tubuh, terutama pada sel saraf dan pembuluh darah. DM diklasifikasikan menjadi DM tipe I yang dikenal sebagai *insulin-dependent* atau *childhood onset diabetes*, ditandai dengan kurangnya produksi insulin dan DM tipe II, yang dikenal dengan *non-insulin-dependent* atau *adult-onset diabetes*, disebabkan ketidakmampuan tubuh menggunakan insulin secara efektif [1].

##### 2.1.2 Microarray Data

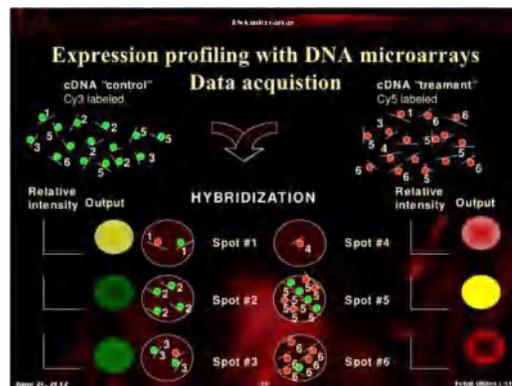
*Microarray* merupakan suatu alat yang dapat digunakan untuk mengukur tingkat ekspresi ribuan gen dari sebuah sampel (pasien) secara simultan dalam sebuah eksperimen. Metode ini menggunakan alat berupa *slide* yang terbuat dari kaca dan terdiri dari ribuan bahkan puluhan ribu blok. Kemampuan teknologi *microarray* tersebut menjadikannya banyak digunakan dalam bidang biologi molekuler. Pola ekspresi gen dari sejumlah pasien yang telah diketahui jenis penyakitnya, dapat digunakan untuk memprediksi jenis penyakit pada pasien yang baru. Permasalahan tersebut disebut sebagai klasifikasi [7,8].





Gambar 1 Cara Kerja Teknologi Microarray

Berdasarkan Gambar 1 setelah mengambil sampel berupa darah, kemudian mRNA ini diisolasi. mRNA ditransformasi menjadi cDNA, cDNA tersebut dilabeli dengan dua dyes yang berbeda. DNA pertama diberi label dengan fluorescent green Cy3 dan DNA kedua diberi label fluorescent red Cy5. cDNA tersebut kemudian dicampur dan dimasukkan kedalam *slide microarray* yang sebelumnya telah diberi probe. Setelah itu proses hibridisasi, yaitu melakukan inkubasi terhadap *chip microarray* selama satu malam dalam suhu 60°. Langkah selanjutnya adalah *scanning* dengan sinar laser, tujuannya adalah untuk mendeteksi intensitas fluorescent yang divisualisasikan pada komputer. Langkah terakhir adalah melakukan analisis data yang menampilkan ekspresi gen yang berbeda [9].



Gambar 2 Hasil Interpretasi Teknologi DNA Microarray.

Berdasarkan pada Gambar 2 hasil interpretasi teknologi DNA *microarray* ekspresi gen tertentu lebih tinggi maka akan tampak merah. Sebaliknya, jika ekspresi gen dalam sampel percobaan lebih rendah, maka akan tampak hijau.



Akhirnya, jika ada dua ekspresi yang sama dalam sebuah sampel, maka akan muncul kuning. Sebuah titik hitam menunjukkan bahwa tidak ada cDNA yang terikat pada DNA pada gen yang terletak di tempat tersebut. Hal ini menunjukkan bahwa gen tidak aktif (semua gen dalam percobaan aktif)[10].

Data yang dikumpulkan melalui *microarray* dapat digunakan untuk membuat profil ekspresi gen, yang menunjukkan perubahan ekspresi banyak gen secara simultan dalam menanggapi suatu kondisi atau pengobatan tertentu.

### 2.1.3 Min-Max Standarditation

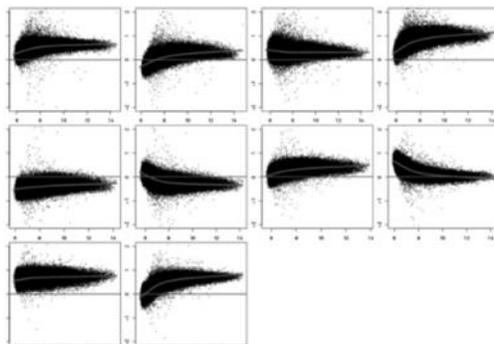
Min-Max normalization merupakan salah satu teknik dalam standarisasi data, dimana skala nilainya antara 1 dan 0. Untuk standarisasi dengan teknik Min-Max standaritation menggunakan persamaan berikut.

$$\text{Standarisasi}(x) = \frac{x - \text{Min}(X_i)}{\text{Max}(X_i) - \text{Min}(X_i)} \quad (1)$$

$\text{Min}(X_i)$  dan  $\text{Max}(X_i)$  merupakan nilai minimum dan nilai maksimum dari kolom fitur ( $X_i$ ) pada data. Setelah dilakukan standarisasi data, maka diperoleh nilai masukan data yang berada dalam jangkauan nilai [0,1],[4].

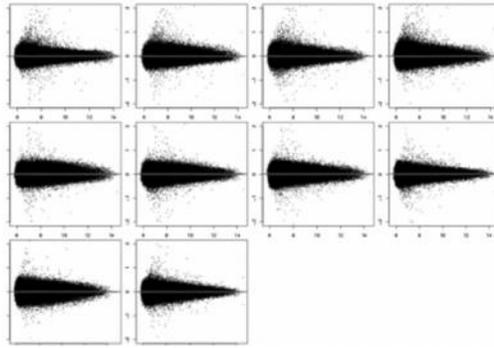
### 2.1.4 Quantile Normalization

*Quantile normalization* merupakan salah satu teknik dalam normalisasi data dimana mengubah setiap variabel pada suatu data memiliki distribusi yang sama.



**Gambar 3** Data sebelum Normalisasi Quantile





**Gambar 4** Data setelah Normalisasi Quantile

*Quantile normalization* biasanya digunakan pada data *MicroArray* [11]. Untuk normalisasi dengan teknik *Quantile normalization* dengan langkah – langkah berikut.

1. Diberikan sebuah data dalam bentuk Matrix X.
2. Urutkan setiap kolom pada Matrix X sehingga membentuk Matrix  $X_{Sort}$ .
3. Hitung *mean* setiap baris pada Matrix  $X_{Sort}$  , Sehingga membentuk Matrix  $X'_{Sort}$ .
4. Untuk mendapatkan matrix  $X_{normalized}$ . Ganti nilai setiap elemen kolom pada X dengan  $X'_{Sort}$  sesuai dengan urutan pada X.

### 2.1.5 t-test

Untuk mengatasi masalah tingginya dimensi suatu data, dapat dilakukan dengan seleksi fitur .Dalam penelitian ini, pendekatan *filter* seperti *t-test* akan dilakukan. Hal ini bertujuan untuk mendapatkan hasil klasifikasi terbaik hanya dengan beberapa fitur-fitur yang penting [4]. Nilai *t-test* (TS) pada fitur ke-i dihitung menggunakan pada persamaan berikut.

ada persa

$$TS(i) = \frac{\bar{X}_1 - \bar{X}_2}{s_{X1X2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2)$$

$$s_{X1X2}^2 = \frac{(n_1 - 1) s_{X1}^2 + (n_2 - 1) s_{X2}^2}{n_1 + n_2 - 2} \quad (3)$$



Dimana  $SX_1X_2$  merupakan standar deviasi dari 2 sampel  $X_{ik}$  representasi rata – rata dari fitur ke – i dari class  $k \in \{1,2\}$  dan  $s$  representasi dari standar deviasi.

Dalam dua kelas, uji untuk hipotesis nol ( $H_0$ ) menunjukkan bahwa rata – rata dari kelas satu dan dua adalah sama. Berarti tidak ada perbedaan signifikan kedua rata – rata tersebut, sehingga fitur tersebut tidak banyak berkontribusi dalam klasifikasi. Oleh karena itu, perlu menolak hipotesis nol ( $H_0$ ) dan menerima hipotesis alternatif ( $H_1$ ). Sehingga, fitur yang memiliki kesamaan yang signifikan dapat dibuang, dan fitur yang memiliki perbedaan signifikan dapat diterima.

Dalam penelitian ini *t-test* akan diterapkan pada setiap fitur, dan diukur dari nilai *p-Value* yang sesuai. Fitur yang paling informatif akan diidentifikasi dengan menyortir masing-masing fitur berdasarkan nilai *p-Value* , yaitu  $p < 0.05$ . Fitur – fitur yang memiliki nilai *p-Value* paling kecil dapat digunakan untuk klasifikasi.[4]

### 2.1.6 Artificial Neural Network (ANN)

Jaringan Saraf Tiruan (JST) adalah prosesor yang terdistribusi besar-besaran secara parallel yang dibuat dari unit proses sederhana, yang mempunyai kemampuan untuk menyimpan pengetahuan berupa pengalaman dan dapat digunakan untuk proses lain [12].

Jaringan saraf tiruan tidak diprogram untuk menghasilkan keluaran tertentu. Semua keluaran atau kesimpulan yang ditarik oleh jaringan didasarkan pengalamannya selama mengikuti proses pembelajaran. Pada proses pembelajaran, ke dalam jaringan saraf tiruan dimasukkan pola-pola masukan (dan keluaran) lalu jaringan akan diajari untuk memberikan jawaban yang bisa diterima [17].

Jaringan saraf tiruan dibentuk sebagai generalisasi model matematika dari saraf biologis manusia, dengan asumsi bahwa :

prosesan informasi terjadi pada banyak elemen sederhana (neuron).



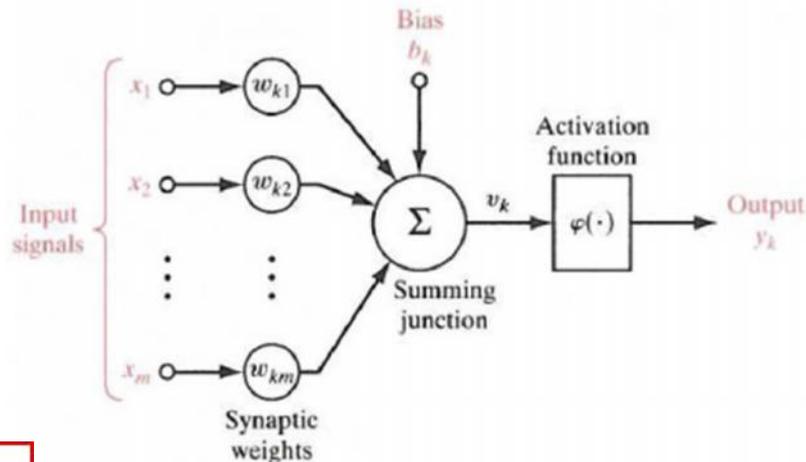
- Sinyal dikirimkan di antara neuron-neuron melalui penghubung-penghubung.
- Penghubung antar neuron memiliki bobot yang akan memperkuat atau memperlemah sinyal.
- Untuk menentukan keluaran, Setiap neuron menggunakan fungsi aktivasi (biasanya bukan fungsi linier) yang dikenakan pada jumlah masukan yang diterima. Besarnya keluaran ini selanjutnya dibandingkan dengan suatu batas ambang.

### 1. Prinsip Jaringan Syaraf Tiruan

Prinsip jaringan saraf tiruan (JST) ditentukan oleh tiga elemen dasar model saraf, yaitu:

- 1.) Satu set dari sinapsis, atau penghubung yang masing-masing digolongkan oleh bobot atau kekuatannya.
- 2.) Sebuah penambah untuk menjumlahkan sinyal-sinyal input. Ditimbang dari kekuatan sinaptik masing-masing neuron.
- 3.) Sebuah fungsi aktivasi untuk membatasi amplitudo output dari neuron. Fungsi ini bertujuan membatasi jarak amplitudo yang diperbolehkan oleh sinyal output menjadi sebuah angka yang terbatas

Prinsip jaringan saraf tiruan secara sederhana digambarkan di bawah ini:



Gambar 5 Prinsip jaringan saraf tiruan [14]



Pada gambar diatas, Y menerima masukan dari neuron  $x_1, x_2, x_3, \dots, x_n$  dengan bobot hubungan masing-masing adalah  $w_1, w_2, w_3, \dots, w_n$ . Impuls neuron yang ada dijumlahkan menjadi:

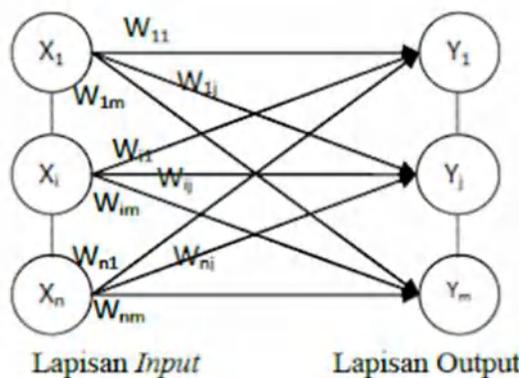
$$Net = \sum_{i=1}^n w_i x_i \quad (4)$$

Besarnya impuls yang diterima oleh Y mengikuti fungsi aktivasi  $y=f(net)$ . Apabila nilai fungsi aktivasi cukup kuat, maka sinyal akan diteruskan. Nilai fungsi aktivasi (keluaran model jaringan) juga dapat dipakai sebagai dasar untuk mengubah bobot [15].

## 2. Arsitektur Jaringan Saraf Tiruan

Pada jaringan saraf tiruan, neuron-neuron akan dikumpulkan dalam sebuah lapisan yang disebut dengan lapisan neuron (neuron layers). Neuron-neuron pada satu lapisan akan dihubungkan dengan lapisan-lapisan lainnya. Informasi yang didapatkan pada sebuah neuron akan disampaikan ke semua lapisan-lapisan yang ada, mulai dari lapisan masukan sampai dengan lapisan keluaran melalui lapisan tersembunyi (hidden layer). Pada jaringan saraf tiruan ini tiga lapisan bukanlah sebuah struktur umum karena beberapa jaringan saraf ada yang tidak memiliki lapisan tersembunyi. Menurut [12], secara umum ada tiga jenis arsitektur dari Jaringan Saraf Tiruan yaitu:

### a. Jaringan dengan lapisan tunggal (single layer net)



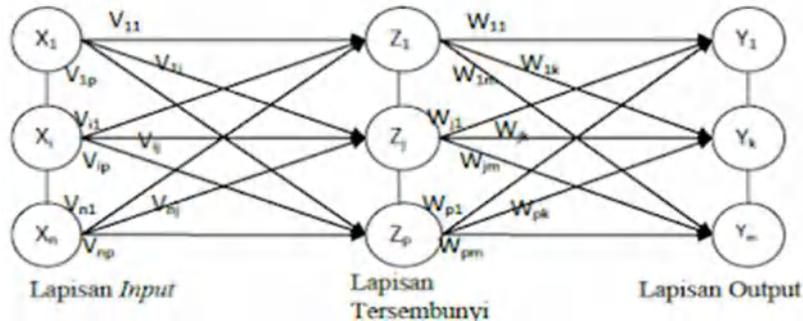
Gambar 6 Jaringan dengan lapisan tunggal (single layer net)

alam Jaringan Saraf Tiruan dengan satu layer, neuron-neuron i dalam bentuk layer-layer. Dalam bentuk paling sederhana dari



Jaringan Saraf Tiruan dengan satu layer, kita mempunyai sebuah input layer dari node sumber di mana informasi diproyeksikan ke output layer dari neuron tapi tidak bisa sebaliknya. Dengan kata lain, jaringan ini adalah tipe feed forward. Input layer dari node sumber tidak dihitung karena tidak ada perhitungan yang dilakukan.

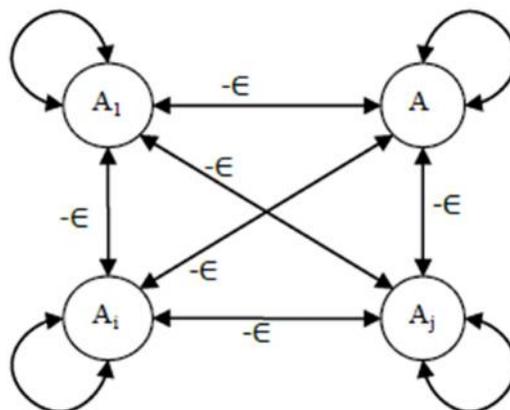
b. Jaringan dengan banyak lapisan (multilayer net)



Gambar 7 Jaringan dengan banyak lapisan (multilayer net)

Merupakan jaringan dengan satu atau lebih lapisan tersembunyi (hidden layer). Jaringan multi lapis ini memiliki kemampuan lebih dalam memecahkan masalah bila dibandingkan dengan jaringan lapis tunggal, namun pelatihannya mungkin lebih rumit. Pada beberapa kasus, pelatihan pada jaringan ini lebih baik karena memungkinkan bagi jaringan untuk memecahkan masalah yang tidak dapat diselesaikan jaringan berlapis tunggal karena jaringan tidak bisa dilatih untuk menampilkan secara benar.

c. Jaringan dengan lapisan kompetitif (competitive layer net)



Gambar 8 Jaringan dengan lapisan kompetitif (competitive layer net)



Bentuk lapisan kompetitif merupakan jaringan saraf tiruan yang sangat besar. Interkoneksi antar neuron pada lapisan ini tidak ditunjukkan pada arsitektur seperti jaringan yang lain. Pada jaringan ini sekumpulan neuron bersaing untuk mendapatkan hak menjadi aktif atau sering pula disebut dengan prinsip winner takes all atau yang menanglah yang mengambil semua bagiannya.

### 3. Mode Pelatihan

Selain arsitektur, metode pengaturan nilai bobot (training) merupakan karakteristik yang penting dalam jaringan Neural Network. Metode pelatihan pada Neural network dibagi menjadi dua jenis, yaitu :

#### a. Pelatihan Terawasi (*Supervised Learning*)

Pelatihan ini dilakukan dengan adanya urutan vektor pelatihan, atau pola yang masing-masing terkait dengan vektor target output. Bobot kemudian disesuaikan untuk algoritma pembelajaran. Proses ini dikenal sebagai pelatihan terawasi.

#### b. Pelatihan tak Terawasi (*Unsupervised Learning*)

Pada pelatihan ini jaring saraf mengatur segala kinerja dirinya sendiri, mulai dari masukan vektor hingga menggunakan data training untuk melakukan pembelajaran

### 2. Fungsi Aktivasi

Fungsi aktivasi yang akan menentukan apakah sinyal dari input neuron akan diteruskan atau tidak. Ada beberapa fungsi aktivasi yang sering digunakan dalam Neural Network, antara lain:

#### a. Fungsi undak biner (Threshold)

Fungsi undak biner dengan menggunakan nilai ambang sering juga disebut dengan fungsi nilai ambang (Threshold) atau fungsi Heaviside.

Fungsi undak biner (dengan nilai ambang  $\theta$ ) dirumuskan sebagai:

$$y = \begin{cases} 0, & x < \theta \\ 1, & x \geq \theta \end{cases} \quad (5)$$

#### b. Fungsi Linier (Identitas)

Fungsi linear memiliki nilai output yang sama dengan nilai inputnya.

Fungsi ini dirumuskan sebagai:



(6)

$$y = x$$

c. Fungsi Sigmoid Biner

Fungsi ini digunakan untuk jaringan syaraf yang dilatih dengan menggunakan metode Backpropagation. Fungsi sigmoid biner memiliki nilai pada range 0 sampai 1. Oleh karena itu, fungsi ini sering digunakan untuk jaringan syaraf yang membutuhkan nilai output yang terletak pada interval 0 sampai 1. Namun, fungsi ini bisa juga digunakan oleh jaringan syaraf yang nilai outputnya 0 atau 1. Fungsi sigmoid biner dirumuskan sebagai:

$$y = f(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

$$f'(x) = f(x)[1 - f(x)] \tag{8}$$

d. Fungsi Sigmoid Bipolar

Fungsi sigmoid bipolar hampir sama dengan fungsi sigmoid biner, hanya saja output dari fungsi ini memiliki range antara 1 sampai -1. Fungsi sigmoid bipolar dirumuskan sebagai:

$$y = f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \tag{9}$$

$$f'(x) = \frac{1}{2} \left[ 1 + \frac{1 - e^{-x}}{1 + e^{-x}} \right] \tag{10}$$

Fungsi ini sangat dekat dengan fungsi hyperbolic tangent. Keduanya memiliki range antara -1 sampai 1. Untuk fungsi hyperbolic tangent, dirumuskan sebagai:

$$y = f(x) = \frac{e^{-x} - e^x}{e^{-x} + e^x} \tag{11}$$

$$y = f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \tag{12}$$

$$f'(x) = \frac{1}{2} \left[ 1 + \frac{1 - e^{-x}}{1 + e^{-x}} \right] \tag{13}$$

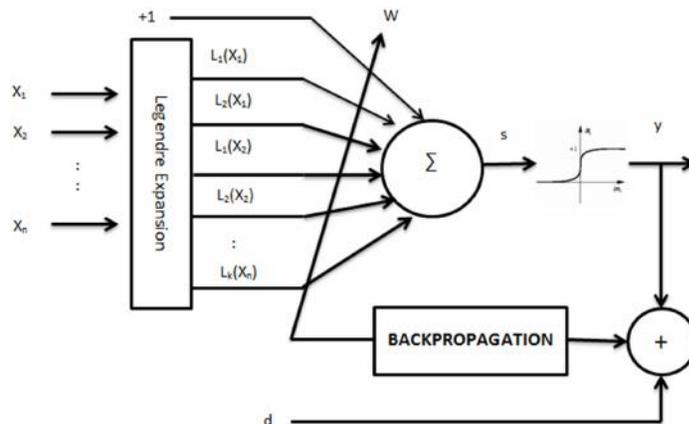


### 2.1.7 Functional Link Neural Network

Functional Link Neural Network (FLNN) merupakan jaringan syaraf tiruan yang memiliki arsitektur layer tunggal, sehingga tidak memiliki *hidden layer* ataupun *hidden node*. Dalam FLNN, vektor masukan diekspansi ke dimensi yang lebih tinggi sesuai blok fungsional atau fungsi basis yang telah ditentukan. Dalam penelitian ini, penulis memilih Legendre Polynomial. Legendre Polynomial dinotasikan dengan  $L_n(X)$ ,  $n$  merupakan orde polinomial, dan  $X$  adalah nilai masukan dari data asli. Untuk membangkitkan nilai masukan ke orde yang lebih tinggi dengan Legendre Polynomial digunakan persamaan rekursif seperti berikut.

$$L_{n+1}(x) = \frac{1}{n+1} [(2n+1)xL_n(x) - nL_{n-1}(x)] \quad (14)$$

Misalkan  $X = [x_1, x_2, \dots, x_n]^T$ . Untuk orde ke-nol dan orde pertama Legendre Polynomial diberikan  $L_0(x) = 1$  dan  $L_1(x) = x$ , maka akan diekspansi ke orde  $k$  sehingga vektor masukan ke *neural network* bukanlah data yang sebenarnya melainkan nilai data yang sudah diekspansi menjadi  $Xe = [1, L_1(x_1), L_2(x_1), L_1(x_2), L_2(x_2), \dots, L_k(x_n)]^T$ .



**Gambar 9** Functional Link Neural Network

Pada gambar 9 merupakan model Functional Link Neural Network dimana nilai input adalah  $X_n$ , kemudian di ekspansi menjadi  $L_k X_n$ . Kemudian nilai keluaran jaringan diaktivasi dengan fungsi sigmoid biner dengan persamaan 8 dan diklasifikasikan ke dalam kelas diabetes. Selanjutnya, untuk evaluasi hasil klasifikasi digunakan fungsi *square error* yang dinotasikan



dengan  $E$  dalam persamaan 16, dimana  $y_i$  merupakan kelas prediksi dan  $d_i$  merupakan target kelas dari masing-masing masukan  $i$  dan sejumlah  $n$  pola pelatihan.

$$s = \sum_{i=1}^n w_i x^i - \theta \quad (15)$$

$$E = \frac{1}{2} \sum_{i=1}^n [d_i - y_i]^2 \quad (16)$$

Selain itu, algoritma pembelajaran yang digunakan pada FLNN adalah *Backpropagation*. *Backpropagation* akan melakukan dua tahap perhitungan, yaitu perhitungan maju untuk menghitung galat antara kelas prediksi dengan target kelas, dan perhitungan mundur untuk mempropagasikan mundur galat tersebut untuk memperbaharui bobot  $w$  dengan persamaan:

$$w_i^{(new)} = w_i^{(old)} + \Delta w_i \quad (17)$$

$$\Delta w_i = \eta (d_i - y_i) L(x_i) \quad (18)$$

dimana  $\eta$  merupakan *learning rate* yang digunakan selama proses pelatihan data terhadap model klasifikasi.[17].

### 2.1.8 Cross Validation

*Cross-validation* (CV) adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi / evaluasi. Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi. Selanjutnya pemilihan jenis CV dapat didasarkan pada ukuran dataset. Biasanya CV K-fold digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi [18].





Classification	Predicted class	
	Class=Yes	Class=No
Class=Yes	a (true positive)	b (false negative)
Class=No	c (false positive)	d (true negative)

Gambar 11 Confussion Matrix

Pada tabel *confusion matrix* di atas, *true positive* (TP) adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positive* (FP) adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negatives* (FN) adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negatives* (TN) adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif. Setelah data uji diklasifikasikan maka akan didapatkan *confusion matrix* sehingga dapat dihitung jumlah sensitivitas, spesifisitas, dan akurasi [20].

Rumus untuk menghitung akurasi pada *confusion matrix* adalah sebagai berikut

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{a+d}{a+b+c+d} \quad (19)$$

### 2.1.10 ROC (Receiver Operating Characteristic) Curve

ROC (*Receiver Operating Characteristic*) *curve* merupakan suatu metode yang dapat digunakan untuk menilai kinerja suatu pengujian [12]. Pengujian tersebut digambarkan dalam suatu kurva dengan axis vertikal merupakan tingkat *true positif (sensitivity)*, artinya ialah suatu kelas yang diprediksi masuk ke kelas positif dan hasilnya benar. Label axis horizontal merupakan tingkat *false positif (1-specificity)*, artinya ialah kelas data yang diprediksi masuk ke kelas positif dan hasilnya salah. Cara tersebut akan memberikan hasil 100% jika tidak ada yang yang salah terprediksi.

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \quad (20)$$

$$Specificity = \frac{TN}{TN+FP} \times 100\% \quad (21)$$



*Area Under Curve*(AUC) adalah ukuran gabungan sensitivitas dan spesifisitas yang merupakan ukuran keseluruhan kinerja tes diagnostik yang diinterpretasikan sebagai rata-rata nilai sensitivitas untuk semua kemungkinan nilai spesifisitas [21]. Nilai AUC terentang antara 0 dan 1 karena sumbu x dan y mempunyai nilai mulai dari 0 sampai 1. Jika nilai AUC mendekati 1, kinerja keseluruhan tes diagnostik semakin baik dan tes dengan nilai  $AUC = 1$  berarti kinerja sangat akurat [22].



## 2.2. Kerangka Konseptual

Kerangka konseptual atau kerangka pikir merupakan model konseptual tentang bagaimana teori berhubungan dengan berbagai faktor yang telah diidentifikasi. Sehingga pada penelitian ini memerlukan kerangka konseptual agar mempermudah dalam proses penelitian. Maka dari itu dapat digambarkan kerangka konseptual sebagai berikut:

