

### PREDIKSI PASIEN PENYAKIT JANTUNG MENGGUNAKAN JARINGAN SYARAF TIRUAN MULTI LAYER PERCEPTRON DAN PYTHON PADA BASIS DATA PENYAKIT JANTUNG DI CLEVELAND

#### PENULIS

<sup>1)</sup>Ari Kurniawan, <sup>2)</sup>Astried Silvanie

#### ABSTRAK

Penelitian ini menggunakan Jaringan Syaraf Tiruan *Multi-Layer Perceptron* (JST MLP) untuk memprediksi apakah pasien berpenyakit jantung berdasarkan kondisi medis mereka. Sebagai prediktor fitur terdiri atas 13 atribut. Kelas target berupa nilai biner, dimana 1 = penyakit jantung dan 0 = bukan penyakit jantung. JST MLP yang diusulkan terdiri atas tiga lapisan yaitu lapisan masukan (13 neuron), satu lapisan tersembunyi (12 neuron) dan lapisan keluaran (1 neuron). Implementasi menggunakan *python* dan pustaka *tensorflow*. Pelatihan dilakukan sebanyak 300 *epochs* dan 10 *batch*. Hasil berupa bobot untuk setiap neuron dalam JST. Nilai metrik akurasi model didapatkan sebesar 81.19% dengan nilai sensitivitas 99.39% dan spesifisitas 94.93%.

#### Kata Kunci

Jaringan Syaraf Tiruan, Prediksi Penyakit Jantung, *Machine Learning*, *Multilayer Perceptron*, Implementasi ML Dengan *Python*

#### AFILIASI

Prodi, Fakultas  
Nama Institusi  
Alamat Institusi

Teknik Informatika, Fakultas Ilmu Komputer  
Institut Bisnis dan Informatika (IBI) Kosgoro 1957  
Jl. M. Kahfi II No. 33, Jagakarsa, Jakarta Selatan, DKI Jakarta

#### KORESPONDENSI

Penulis  
Email

Astried Silvanie  
astried@ibi-k57.ac.id

#### LICENSE



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

## I. PENDAHULUAN

### 1.1 Latar Belakang

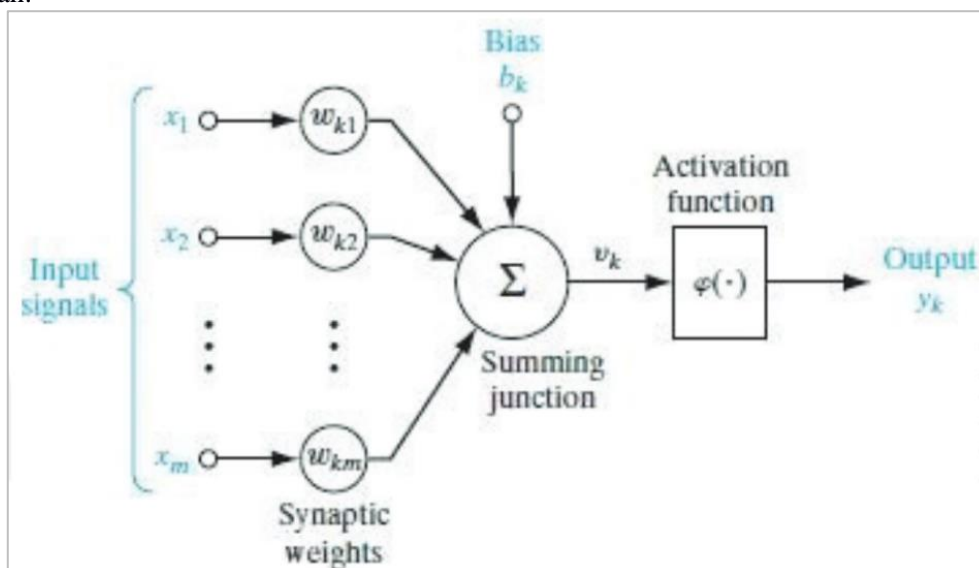
Menurut Bala, Rajni dkk (2017), Klasifikasi adalah teknik penambangan data yang digunakan untuk prediksi kelas objek dan merupakan contoh pembelajaran terawasi. Klasifikasi memprediksi label kategoris (diskrit, berurutan). Klasifikasi data melibatkan dua langkah. Langkah pertama adalah langkah pembelajaran (langkah pelatihan) di mana *classifier* dibangun untuk menggambarkan sekumpulan kelas data yang telah ditentukan. Pada langkah kedua model yang dibangun pada langkah pertama digunakan untuk klasifikasi data yang tidak diketahui yaitu data uji digunakan untuk memperkirakan akurasi *classifier*.

Dalam beberapa tahun terakhir, ada peningkatan minat dalam penggunaan pembelajaran mesin di bidang diagnosis medis. K, Uma Maheswari dkk (2017) mengombinasikan logistik analisis regresi dan jaringan saraf sebagai pendekatan baru dalam memprediksi penyakit jantung. Ballinger dkk (2018) memprediksi orang berisiko tinggi kena jantung pada perangkat jam pintar menggunakan *multi-task LSTM*. Sistem prediksi penyakit yang dibuat Chauhan dkk (2020) menganalisis gejala yang diberikan oleh pengguna sebagai *input* dan memberikan probabilitas penyakit sebagai *output*. Prediksi Penyakit dilakukan dengan menerapkan pengklasifikasi pohon keputusan. Hal ini sebagai bukti pembelajaran mesin selalu berkembang dan digunakan dalam proses diagnosis medis. Dalam penelitian ini peneliti menggunakan jaringan syaraf tiruan (JST) untuk prediksi penyakit jantung berdasarkan kondisi medis seseorang. Proses pelatihan dan pengujian diimplementasikan dengan *python* dan pustaka *tensorflow*.

### 1.2 Jaringan Syaraf Tiruan (JST)

Menurut Zayegh.dkk (2018), Otak manusia terdiri dari sekitar 1011 unit komputer "neuron" yang bekerja secara paralel dan bertukar informasi melalui "sinapsis" penghubungnya; neuron ini merangkum semua informasi yang masuk ke dalamnya, dan jika hasilnya lebih tinggi dari potensial yang diberikan yang disebut potensial aksi, mereka mengirim pulsa melalui akson ke tahap berikutnya.

Menurut Zayegh.dkk (2018), Jaringan syaraf tiruan, disingkat JST, adalah teknik komputasi yang dirancang untuk menyimulasikan metode otak manusia dalam memecahkan masalah. Dengan cara yang sama, jaringan syaraf tiruan terdiri dari unit komputasi sederhana "neuron buatan", dan setiap unit terhubung ke unit lain melalui konektor berat; kemudian, unit-unit ini menghitung jumlah tertimbang dari *input* yang datang dan mencari *output* menggunakan fungsi *squashing* atau fungsi aktivasi. Gambar 1 menunjukkan diagram blok neuron buatan.



Gambar 1. Arsitektur Jaringan Syaraf Tiruan  
(Zayegh. dkk, 2018)

### 1.3 Multilayer Perceptron (MLP)

Menurut Marius dkk (2009), *Multilayer Perceptron*, disingkat MLP, adalah arsitektur jst yang diawasi dengan satu lapisan *input*, satu atau lebih lapisan tersembunyi dan satu lapisan keluaran. Lapisan *input*

menerima *input* eksternal dan mendistribusikannya kembali ke neuron di lapisan tersembunyi. Lapisan tersembunyi terdiri dari neuron komputasi yang mendeteksi fitur tersembunyi dalam data *input*. Fitur yang dideteksi oleh neuron di lapisan tersembunyi kemudian digunakan oleh lapisan keluaran untuk menentukan keluaran akhir dari jaringan. Dengan satu atau dua lapisan tersembunyi, MLP dapat mendekati hampir semua peta masukan-keluaran.

## II. METODE PENELITIAN

Langkah-langkah dalam penelitian ini dirangkum sebagai berikut:

- 1) Identifikasi Masalah. Pada tahap ini, referensi dimana dikumpulkan terkait prediksi penyakit jantung koroner dengan kondisi medis. Masalah yang berhubungan untuk diagnosis penyakit juga dikumpulkan dari penelitian-penelitian sebelum.
- 2) Pengumpulan dan Pra Proses Data. Pada tahap ini sumber data adalah data sekunder untuk diagnosis penyakit jantung diperoleh dari Gudang Pembelajaran Mesin UCI. Data sudah diproses sebelumnya dan dinormalisasi.
- 3) Desain dan implementasi. Desain arsitektur JST dirancang dengan tiga lapisan. Proses pelatihan dan pengujian diimplementasikan menggunakan *Python* dan pustaka *Tensorflow*.
- 4) Hasil: Hasil yang dikeluarkan dari pelatihan pembelajaran mesin adalah arsitektur JST lengkap dengan bobotnya. Diskusi terkait dengan pengukuran akurasi.

## III. HASIL DAN PEMBAHASAN

### 3.1 Data Olah

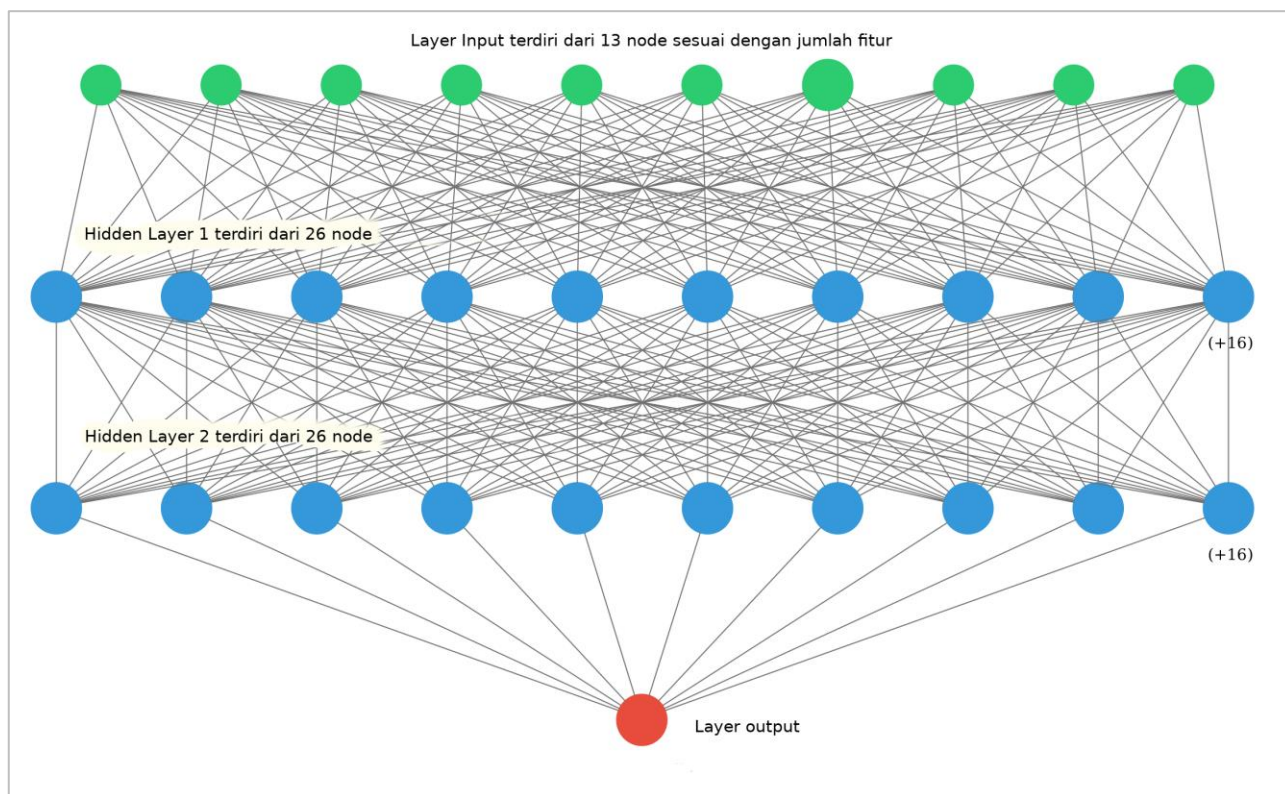
*Database* klinis mengumpulkan sejumlah besar informasi tentang pasien dan kondisi medisnya. Istilah penyakit jantung mencakup beragam kondisi yang mempengaruhi jantung. Penyakit kardiovaskular adalah penyebab utama korban di dunia. Syarat "Penyakit kardiovaskular" terdiri dari berbagai kondisi yang mempengaruhi jantung dan vena darah dan cara darah dipompa dan diedarkan melalui tubuh (Grace, S.L skk, 2004). Data yang digunakan dalam proyek ini diperoleh dari *database* penyakit jantung Cleveland. Sebanyak 303 pasien dengan 13 atribut kondisi medis yang digunakan untuk memprediksi penyakit jantung. Deskripsi *dataset* ditabulasikan dalam Tabel 1 di bawah ini. Kelas target adalah atribut ke empat belas yang bernilai 1 untuk penyakit jantung dan 0 untuk bukan penyakit jantung.

**Tabel 1. Atribut Prediksi dan Kelas Target pada Basis Data Pasien Jantung di Cleveland**

NO	ATRIBUT	TIPE DATA
1	age	int64
2	sex	int64
3	chest pain type (4 values)	int64
4	resting blood pressure	int64
5	serum cholestoral in mg/dl	int64
6	fasting blood sugar > 120 mg/dl	int64
7	resting electrocardiographic results (values 0,1,2)	int64
8	maximum heart rate achieved	int64
9	exercise induced angina	int64
10	oldpeak = ST depression induced by exercise relative to rest	float64
11	the slope of the peak exercise ST segment	int64
12	number of major vessels (0-3) colored by flourosopy	int64
13	thal: 3 = normal; 6 = fixed defect; 7 = reversable defect	int64
14	Target: 1 = Yes, 0 = No	int64

### 3.2 Desain Arsitektur JST yang Diusulkan

Arsitektur JST terdiri dari tiga lapisan yaitu lapisan masukan, satu lapisan tersembunyi dan lapisan keluaran seperti pada gambar di bawah ini. Dimensi masukan untuk lapisan pertama terdiri dari 13 neuron sama dengan jumlah atribut prediksi yang telah dijelaskan pada bab sebelum. Satu lapisan tersembunyi terdiri atas 26 neuron yang merupakan dua kali neuron dari neuron masukan. Lapisan *output* terdiri dari satu neuron keluaran. Arsitektur JST tersebut dapat dilihat pada gambar di bawah ini.



**Gambar 2. Arsitektur JST Usulan Terdiri Atas Tiga Lapisan**

### 3.3 Implementasi dengan menggunakan Python

Pertama kali kita melakukan *import* pustaka yang diperlukan yaitu *numpy*, *pandas*, *asarray* dan *sklearn*. Kemudian data yang berbentuk format comma separated values (.csv) diload ke pengubah dataset *pandas*.

```

1 import numpy as np
2 import pandas as pd
3
4 dataset = pd.read_csv("heart.csv",delimiter=",")

```

Dataset kemudian dikopi dan disini kita akan mengganti nilai data not applicable (NA) menjadi string kosong.

```

10 copy_dataset = dataset
11 copy_dataset.fillna(value="", inplace=True)

```

Kemudian impor pustaka *keras*, *dense* dan *sequential* untuk membuat model *jst*. Pertama kita harus memisahkan mana atribut yang menjadi prediktor dan mana kelas target. Atribut fitur berjumlah 13 dari indeks ke 0 sampai dengan 12. Kelas target merupakan atribut pada indeks terakhir yaitu 13.

```

12 from keras.models import Sequential
13 from keras.layers import Dense
14
15 Fitur = copy_dataset.iloc[:,0:13]
16 Target = copy_dataset.iloc[:,13]

```

Menggunakan *tensorflow*, *keras* dan *sequential*, kita desain arsitektur *jst* dengan tiga lapisan. Lapisan pertama menerima 13 dimensi masukan. Inisialisasi ditentukan dengan bobot acak awal lapisan *Keras*. Fungsi aktivasi yang diterapkan fungsi aktivasi unit linier yang diperbaiki. Dengan nilai default mengembalikan aktivasi *ReLU* standar:  $\max(x, 0)$ , maksimum dari segi elemen 0 dan tensor input. Lapisan tersembunyi terdiri

dari 26 neuron dengan inisialisasi bobot awal secara acak. Fungsi aktivasi yang digunakan adalah unit linier yang diperbaiki. Lapisan keluaran terdiri dari hanya 1 neuron dengan inisialisasi bobot awal secara acak. Fungsi aktivasi yang digunakan adalah sigmoid.

```

17 model = Sequential()
18
19 model.add(Dense(26, input_dim=13, kernel_initializer='random_uniform', activation='relu'))
20 model.add(Dense(26, kernel_initializer='random_uniform', activation='relu'))
21 model.add(Dense(1, kernel_initializer='random_uniform', activation='sigmoid'))
    
```

Konfigurasi model latihan dengan menentukan nilai *loss*, *optimizer* dan *metrics*. Tujuan dari fungsi kerugian (*loss*) adalah untuk menghitung kuantitas yang model harus berusaha untuk meminimalkan selama pelatihan berlangsung. Fungsi yang digunakan yaitu *binary cross entropy* menghitung kerugian lintas-entropi antara label yang benar dan label yang diprediksi. Pengukuran model dengan akurasi, metrik ini menghitung seberapa sering prediksi sama dengan label yang diinginkan. Metrik ini membuat dua variabel lokal, total dan jumlah yang digunakan untuk menghitung frekuensi *y\_prediksi* cocok dengan *y\_benar*.

Bagian dari data latih yang akan digunakan sebagai data validasi ditentukan sebanyak 30%. Model akan memisahkan bagian data pelatihan ini, tidak akan melatihnya, dan akan mengevaluasi kehilangan dan metrik model di akhir setiap epoch. Ukuran batch ditentukan 10, ukuran ini adalah *hyperparameter* yang menentukan jumlah sampel yang harus dikerjakan sebelum memperbarui parameter model internal. Jumlah epoch adalah *hyperparameter* yang mendefinisikan berapa kali algoritma pembelajaran akan bekerja melalui seluruh dataset pelatihan dan disini ditentukan sebanyak 300 kali.

```

22 # Compile model
23 model.compile(loss='binary_crossentropy', optimizer='adam', metrics=[keras.metrics.Precision(),
24 keras.metrics.Recall(), keras.metrics.SpecifictyAtSensitivity(0.5),
25 keras.metrics.SensitivityAtSpecificty(0.5), 'accuracy'])
26
27 # Fit the model
28 history = model.fit(Fitur, Target, validation_split=0.30, epochs=300, batch_size=10)
29
    
```

#### IV. PENUTUP

Ukuran kinerja jaringan saraf adalah dihitung dengan menggunakan berbagai ukuran seperti akurasi, presisi dan spesifisitas dan sensitivitas.

Perhitungan sebagai berikut:

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN)$$

$$\text{Specificity} = TN / (TN+FP)$$

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

Diketahui:

TP = True Positive, adalah jumlah sampel yang diklasifikasikan memiliki penyakit jantung dan mereka sebenarnya memiliki penyakit jantung.

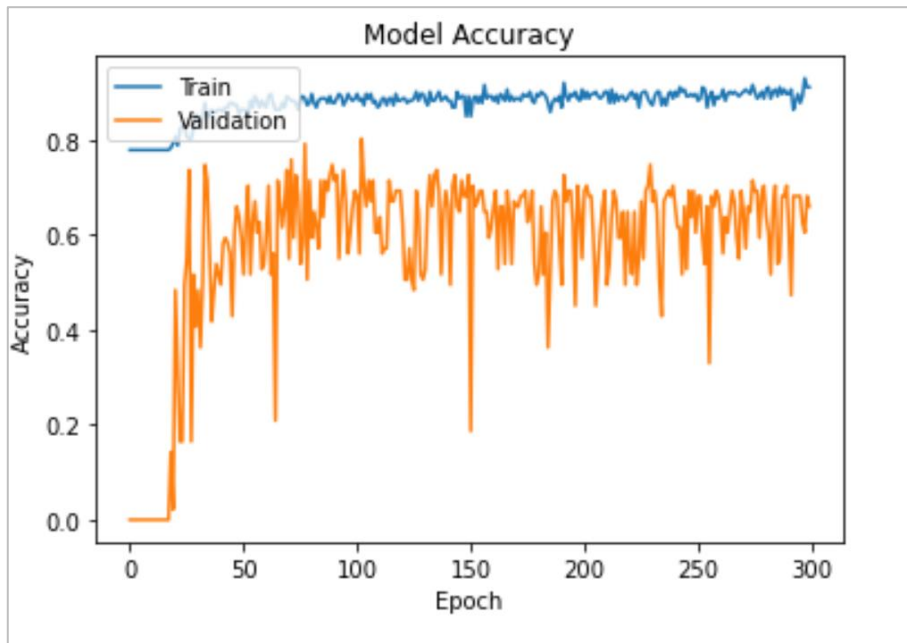
TN = True Negative, adalah jumlah sampel yang diklasifikasikan sebagai tidak memiliki penyakit jantung dan memang benar jika mereka sebenarnya bukan.

FN = False Neagtive, adalah jumlah sampel yang diklasifikasikan sebagai tidak memiliki penyakit jantung padahal mereka sebenarnya memiliki penyakit jantung.

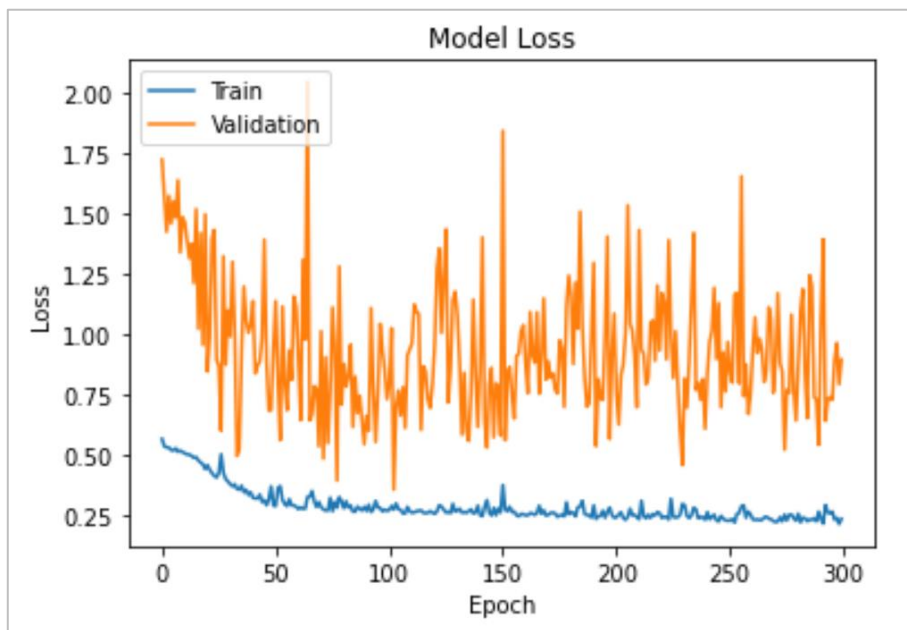
FP = False Positive, adalah jumlah sampel yang diklasifikasikan sebagai memiliki penyakit jantung padahal mereka sebenarnya tidak memiliki penyakit jantung.

Akurasi yang diperoleh dari jaringan saraf adalah 81.19%. Nilai sensitivitas 99.39% dan spesifisitas 94.93%.





Gambar 3. Grafik Metrik Akurasi



Gambar 4. Grafik Metrik Kehilangan

Berikut ini merupakan hasil keluaran yaitu bobot untuk setiap neuron di JST yang sudah dirancang.

Tabel 2. Ringkasan Hasil Bobot Untuk Model JST Prediksi Penyakit Jantung yang Diusulkan

<p><b>Layer 1</b>                  [array([[ -5.04135676e-02, -4.85455170e-02, -5.44698834e-02,                  -8.97674933e-02, -1.51826115e-02, 1.03304870e-02,                  -1.00231878e-02, 7.69542307e-02, 4.39888723e-02,                  7.27043822e-02, -1.00974683e-02, -2.48675104e-02,                  -3.47866081e-02, -1.01632684e-01, -4.94176634e-02,                  4.63768281e-02, -6.61990698e-03, -2.68794894e-02,                  6.56066090e-03, -9.00033023e-03, 3.35142240e-02,                  -2.84368210e-02, -4.92011793e-02, -4.62276824e-02,                  2.93028839e-02, 1.16809830e-02],</p> <p>.....</p>
---

## Layer 2

```
[array([[ -5.44081479e-02, -7.78662115e-02,  6.30459236e-03,
         1.00784943e-01, -5.02638370e-02, -1.16731994e-01,
         7.09891915e-02,  1.17209040e-01, -1.17532752e-01,
         9.35726538e-02,  2.05954425e-02,  7.76523575e-02,
        -2.19414428e-01,  7.22329831e-03, -7.63132721e-02,
         1.14929721e-01,  9.71682295e-02,  5.65704368e-02,
         3.52205299e-02,  9.97119695e-02, -2.93218601e-03,
         3.40375165e-03, -9.42222849e-02,  7.18530174e-03,
         9.15994588e-03,  4.27050404e-02],
        [ 2.26561911e-02,  4.76408862e-02, -2.19793078e-02,
        -2.17169058e-02,  2.01429464e-02,  3.14068533e-02,
         9.83313471e-03, -2.83221956e-02,  2.49758847e-02,
         8.87621194e-04,  4.10182513e-02, -2.15862039e-02,
         3.39695476e-02,  3.45381759e-02, -6.97356462e-03,
         4.36009876e-02,  3.75664718e-02,  4.80983406e-03,
        -3.48760001e-02, -4.26918380e-02, -1.67770013e-02,
         3.30810659e-02,  2.64212154e-02,  1.99007504e-02,
        -1.77179687e-02,  2.05655433e-02],
```

## Layer 3

```
[array([[ 0.02165431],
        [-0.8368734 ],
        [-0.00175756],
        [ 0.16771291],
        [-0.5014123 ],
        [-0.69241714],
        [ 0.18356402],
        [ 0.16224071],
        [-0.407143 ],
        [ 0.18817875],
        [-0.21452232],
        [ 0.09150755],
        [-0.6527461 ],
        [-0.17162275],
        [-0.6847269 ],
        [ 0.17780155],
        [ 0.14952335],
        [ 0.13478632],
        [ 0.15474778],
        [ 0.16390319],
        [-0.21944687],
        [-0.55224675],
        [-0.58785045],
        [-0.07582567],
        [-0.00817954],
        [ 0.13496763]], dtype=float32), array([-0.14039049], dtype=float32)]
```

Pembelajaran mesin adalah area penelitian aktif di mana banyak peneliti bekerja di domain kesehatan untuk mengidentifikasi risiko penyakit. Keuntungan JST adalah membutuhkan lebih sedikit pelatihan statistik formal untuk dikembangkan dan secara implisit dapat mendeteksi hubungan non-linier yang kompleks antara dependen dan variabel independen. Penelitian selanjutnya bisa dikembangkan untuk pembuatan aplikasi dan diharapkan dapat mengembangkan model dengan tingkat akurasi yang lebih tinggi dalam prediksi penyakit jantung.

**DAFTAR PUSTAKA**

- Rufai, Ahmad., U. S. Idriss & Mahmood Umar. (2018). Using Artificial Neural Networks to Diagnose Heart Disease. *International Journal of Computer Applications*. 182. 1-6. 10.5120/ijca2018917938.
- Bala, Rajni. & Dr. Dharmender Kumar. (2017). Classification Using ANN: A Review. *International Journal of Computational Intelligence Research* ISSN 0973-1873 Volume 13, Number 7 (2017), pp. 1811-1820.
- Grace, S.L., Rick Fry , Angela Cheung & Donna E Stewart. (2004). Cardiovascular Disease. *BMC Women's Health* 4,S15. <https://doi.org/10.1186/1472-6874-4-S1-S15>
- Han, Jiawei. dan Michael Kamber. (2006). *Data Mining Concept and Techniques*, 2nd edition, USA: Elsevier, Inc, 2006.
- Marius, Popescu., Valentina Balas., Liliana Perescu-Popescu. & Nikos Mastorakis. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*. 8.
- Zayegh, Amer. & Nizar Bassam. (2018). *Neural Network Principles and Applications*. 10.5772/intechopen.80416.
- Ballinger , Brandon., Johnson Hsieh, Avesh Singh,Nimit Sohoni, Jack Wang, Geoffrey H. Tison, Gregory M. Marcus, Jose M. Sanchez, Carol Maguire,Jeffrey E. Olgin, Mark J. Pletcher. (2018). DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction. *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*.
- Chauhan, Raj H., Daksh N. Naik., Rinal A. Halpati., Sagarkumar J. Patel. & Mr. A.D.Prajapati. (2020). Disease Prediction using Machine Learning. *International Research Journal of Engineering and Technology (IRJET)* Volume: 07 Issue: 05 | May 2020.
- K, Uma Maheswari. & J. Jasmine. (2017). Neural Network based Heart Disease Prediction. *International Journal of Engineering Research & Technology (IJERT)* ISSN: 2278-0181 RTICCT - 2017 Conference Proceedings.