**IMPLEMENTASI JARINGAN SYARAF TIRUAN UNTUK MEMPREDIKSI KEMUNGKINAN TERJANGKIT PENYAKIT DIABETES**



Disusun oleh

Ivan Pakpahan – 152018008

Kelas – A

**Program Studi Informatika**

**Fakultas Teknologi Industri**

**Institut Teknologi Nasional Bandung**

**2021**

# Kata Pengantar

Segala syukur dan puji hanya bagi Tuhan Yesus Kristus, oleh karena anugerah-Nya yang melimpah, kemurahan dan kasih setia yang besar, sehingga saya dapat menyelesaikan Laporan Tugas Akhir Machine Learning yang berjudul “IMPLEMENTASI JARINGAN SYARAF TIRUAN UNTUK MEMPREDIKSI KEMUNGKINAN TERJANGKIT PENYAKIT DIABETES” ini dengan tujuan untuk melengkapi tugas akhir mata kuliah Machine Learning.

Laporan ini disusun tidak terlepasa dengan adanya bantuan dari beberapa pihak sehingga dapat memperlancar pembuatan laporan ini. Dengan berhasil disusunnya laporan ini, saya ingin menyampaikan ucapan terimakasih yang sedalam-dalamnya kepada pihak terkait.

Terlepas dari semua itu, Saya menyadari sepenuhnya bahwa masih ada kekurangan baik dari segi susunan kalimat maupun tata bahasanya. Oleh karena itu Saya sangat mengharapkan saran dan kritik yang membangun dari pihak pembaca demi kesempurnaan laporan ini.

Bandung, 24 Mei 2021

# Daftar Isi

[Kata Pengantar 2](#_Toc73653739)

[Daftar Isi 3](#_Toc73653740)

[BAB I Pendahuluan 4](#_Toc73653741)

[A. Latar Belakang 4](#_Toc73653742)

[B. Rumusan Masalah 4](#_Toc73653743)

[C. Tujuan 4](#_Toc73653744)

[BAB II Landasan Teori 5](#_Toc73653745)

[1. Diabetes 5](#_Toc73653746)

[2. Pima Indians Diabetes Database 5](#_Toc73653747)

[3. Backward Elimination 5](#_Toc73653748)

[4. Jaringan Syaraf Tiruan 6](#_Toc73653749)

[BAB III Implementasi dan Analisa 11](#_Toc73653750)

[BAB IV Kesimpulan dan Saran 16](#_Toc73653751)

# BAB I Pendahuluan

## Latar Belakang

Diabetes adalah penyakit yang ditandai dengan kadar gula darah yang tinggi yang disebabkan oleh gangguan pada sekresi dan kinerja insulin yaitu tidak dapat memproduksi atau tidak dapat merespon hormon insulin yang dihasilkan oleh organ pankreas, sehingga menyebabkan kadar gula darah meningkat dan dapat menyebabkan komplikasi jangka pendek maupun jangka panjang. Penyakit ini membutuhkan perawatan seumur hidup dan tidak ada pengobatan yang pasti untuk menyembuhkannya. Insulin adalah hormon yang diproduksi oleh organ terletak di belakang perut yang disebut pankreas yang berfungsi mengatur metabolisme glukosa menjadi energi serta mengubah kelebihan glukosa menjadi glikogen yang disimpan di dalam hati dan otot.

Berbagai macam penelitian *machine learning* telah dilakukan, salah satunya adalah penelitian pemanfaatan *machine learning* untuk memprediksi penyakit Diabetes. Machine Learning adalah bagian dari penelitian tentang kecerdasan buatan, proses pembelajaran mesin dilakukan bertujuan untuk memberi pengetahuan kepada komputer melalui data, pengamatan dan interaksi dengan dunia. Pengetahuan yang diperoleh memungkinkan komputer untuk melakukan generalisasi dengan benar ke pengaturan model yang baru.

Oleh karena itu, berdasarkan latar belakang dan hasil penelitian sebelumnya, penelitian ini bertujuan untuk meningkatkan akurasi dalam studi kasus prediksi penyakit diabetes dengan cara melakukan seleksi fitur dan mengimplementasikan metode jaringan syaraf tiruan (ANN). Pengambilan data pada penelitian ini berupa data sekunder penyakit diabetes yang didapat dari website Kaggle. Hasil dari penelitian ini akan menampilkan keluaran berupa hasil prediksi berdasarkan kriteria tertentu.

## Rumusan Masalah

* Bagaimana cara implementasi seleksi fitur backward elimination pada dataset?
* Bagaimana cara implementasi algoritma machine learning untuk prediksi penyakit diabetes?

## Tujuan

* Mengetahui cara implementasi seleksi fitur backward elimination pada dataset
* Mengetahui cara implementasi algoritma machine learning untuk prediksi penyakit diabetes

# BAB II Landasan Teori

### Diabetes

Diabetes adalah penyakit kronis yang ditandai dengan ciri-ciri berupa tingginya kadar gula (glukosa) darah. Glukosa merupakan sumber energi utama bagi sel tubuh manusia. Glukosa yang menumpuk di dalam darah akibat tidak diserap sel tubuh dengan baik dapat menimbulkan berbagai gangguan organ tubuh. Jika diabetes tidak dikontrol dengan baik, dapat timbul berbagai komplikasi yang membahayakan nyawa penderita.

### Pima Indians Diabetes Database

Source dataset *https://www.kaggle.com/uciml/pima-indians-diabetes-database*

Spesifikasi dataset:

* Variabel Independen (X) :

1. Pregnancies : Number of times pregnant (mempresentasikan berapa kali seorang wanita hamil selama hidupnya)
2. Glucose : Plasma glucose concentration a 2 hours in an oral glucose tolerance test (mempresentasikan konsentrasi glukosa plasma pada 2 jam dalam tes toleransi glukosa).
3. Blood Pressure : Diastolic blood pressure (Tekanan darah adalah cara yang sangat terkenal untuk mengukur kesehatan jantung seseorang, ada juga ukuran tekanan darah yaitu diastolik dan sistolik. Dalam data ini, kita memiliki tekanan darah diastolik dalam (mm / Hg) ketika jantung rileks setelah kontraksi).
4. Skin Thickness : Triceps skin fold thickness (nilai yang digunakan untuk memperkirakan lemak tubuh (mm) yang diukur pada lengan kanan setengah antara proses olecranon dari siku dan proses akromial skapula).
5. Insulin : 2-Hour serum insulin (tingkat insulin 2 jam insulin serum dalam satuan mu U/ml)
6. BMI : Body mass index (Indeks Massa Tubuh (berat dalam kg / (tinggi dalam meter kuadrat), dan merupakan indikator kesehatan seseorang).
7. Diabetes Pedigree Function : Diabetes pedigree function (indikator riwayat diabetes dalam keluarga)
8. Age : umur wanita suku indian pima (years)

* Variabel Dependen (Y) :

1. Outcome: Class variable (0 or 1), 0 untuk tidak mengidap diabetes, dan 1 mengidap diabetes.

### Backward Elimination

Backward Elimination merupakan metode yang dapat menghilangkan atribut yang tidak signifikan dari model. Untuk prosedur Backward Elimination, model dimulai dengan semua atribut yang ada di dalamnya, dan atribut dengan persial statistic terkecil dihapus. Atribut yang digunakan akan diimplementasikan dengan Backward Elimination, sebelumnya data yang digunakan dijadikan data numerik dan diregresikan, sebelum mendapatkan nilai F terlebih dahulu mencari nilai prediksi, rata-rata prediksi, MSR dan MSE

Metode Backward Elimination dilakukan dengan cara memasukkan semua prediktor kemudian mengeliminasi satu persatu hingga tersisa prediktor yang signifikan saja. Eliminasi ini didasarkan pada prediktor yang memiliki nilai sig F yang di atas 0.1.

Berikut ini adalah prosedur melakukan backward elimination:

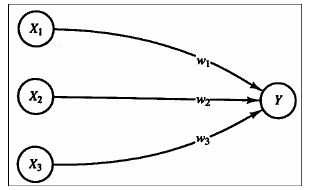
* Pertama-tama masukkan semua variabel independen ke dalam model.
* Pada tiap langkahnya, jika variabel independen yang menghasilkan probabilitas dari F terbesar, lebih besar dibandingkan dengan probability to remove maka variabel independen tersebut dikeluarkan dari dalam model.
* Proses dapat dihentikan jika semua probability dari F lebih kecil dibandingkan dengan probability to remove.

### Jaringan Syaraf Tiruan

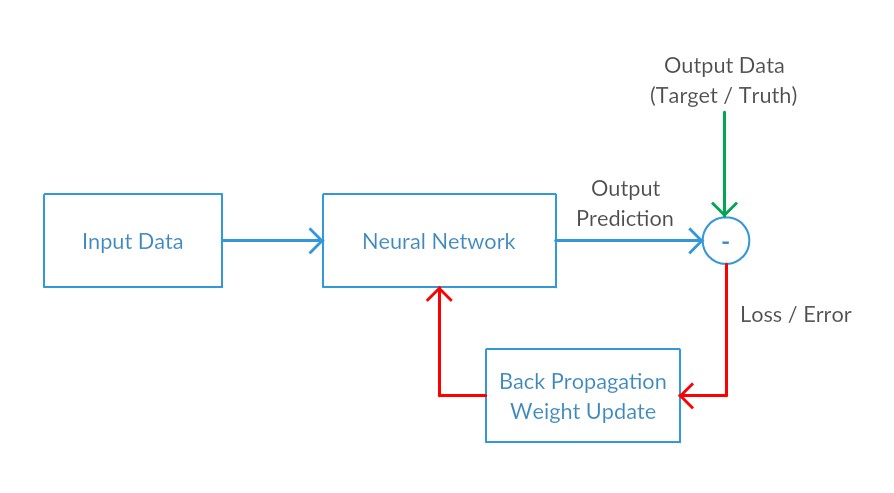
* *Timeline* Sejarah
  + Permodelan aktivitas biologis manusia dimulai pada abad ke-17 (studi tentang kerja jantung)
  + Studi dasar mengenai jaringan neural terjadi pada akhir abad ke-19 dan awal abad 20.
  + Studi bersifat interdisipliner, melibatkan fisika, psikologi, dan neuropsikologi.
  + Ilmuwan yang terlibat : Hermann von Helmhotz, Ernst Mach, dan Ivan Pavlov
  + Pandangan modern mengenai jaringan neural muncul di tahun 1940-an melalui karya Warren McCulloch dan Walter Pitts, yang memperlihatkan bahwa jaringan neuron artifisial, secara prinsipil, dapat menghitung segala fungsi aritmatik dan logika.
  + Donald Hebb mengikuti dengan mengajukan konsep mekanisme belajar pada neuron biologis.
  + Aplikasi JST praktis pertama muncul di akhir tahun 1950-an, melalui penemuan jaringan perseptron dan aturan belajar terasosiasi dari Frank Rosenblatt. Di tahun yang sama, Bernard Widrow dan Ted Hoff memperkenalkan algoritma belajar baru, yang masih dipakai hingga sekarang.
  + Pada tahun 1969, Marvin Minsky dan Seymour Papert menunjukkan sejumlah keterbatasan inherent pada jaringan Rosenblatt dan Widrow. Pada saat itu, sejumlah orang meyakini bahwa penelitian JST sudah berakhir.
  + Keraguan pada masa itu (tahun 60-an) muncul karena minimnya ide baru dan kinerja komputer
  + Penelitian JST ternyata berlanjut pada tahun 70-an, ketika Teuvo Kohonen dan James Anderson secara terpisah mengembangkan JST baru yang dapat berfungsi sebagai memori.
  + Tahun 80-an muncul komputer-komputer berkinerja tinggi (untuk saat itu) dan penelitian JST meningkat secara dramatis.
  + Stephen Grossberg (1976) juga sangat aktif meneliti jaringan self-organizing.
  + John Hopfield (1982) memperkenalkan konsep mekanika statistik, yang digunakan untuk menerangkan jaringan recurrent, yang dapat difungsikan sebagai memori.
  + Kunci kemajuan kedua di tahun 1980-an adalah ditemukannya algoritma backpropagation untuk melatih perseptron multi-layer oleh sejumlah peneliti. Publikasi algoritma yang paling berpengaruh ditulis oleh David Rumelhart dan James McClelland. Algoritma ini merupakan jawaban terhadap kritik Minsky dan Papert (1960)
  + Penelitian terus berlanjut. Konsep, algoritma, dan arsitektur JST baru bermunculan.
* Konsep

JST adalah sistem pemrosesan informasi yang memiliki karakteristik mirip dengan jaringan syaraf biologi. JST dibentuk sebagai generalisasi model matematika dari jaringan syaraf biologi, dengan asumsi bahwa:

* Pemrosesan informasi terjadi pada banyak elemen sederhana (neuron)
* Sinyal dikirimkan diantara neuron-neuron melalui penghubung2
* Penghubung antar neuron memiliki bobot yang akan memperkuat atau memperlemah sinyal
* Untuk menentukan output, setiap neuron menggunakan fungsi aktivasi (biasanya bukan fungsi linier) yang dikenakan pada jumlahan input yang diterima. Besarnya output ini selanjutnya dibandingkan dengan suatu batas ambang (threshold)

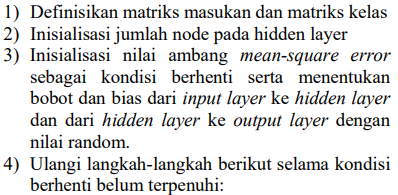


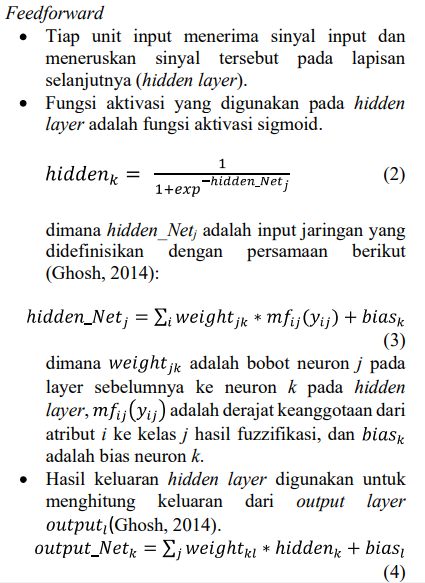
Gambar 1 Hubungan antar neuron

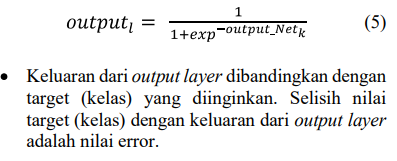


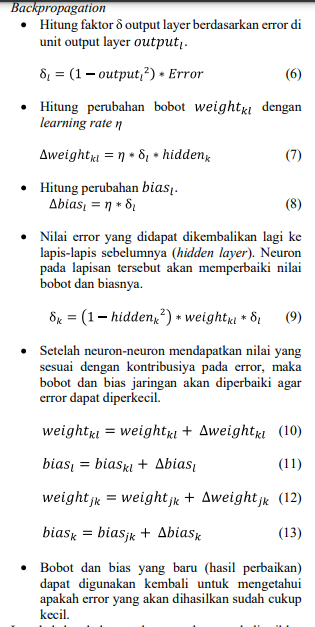
Gambar 2 Arsitektur jaringan syaraf tiruan

Algoritma :





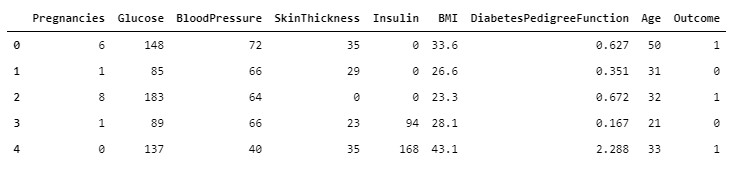




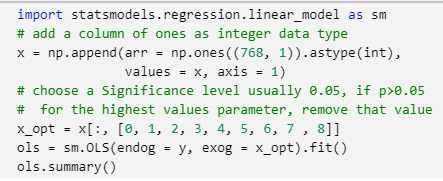
# BAB III Implementasi dan Analisa

* Seleksi Variabel Backward Elimination

1. Semua variabel X dan Y, dimasukkan kedalam model



1. Menghitung P value

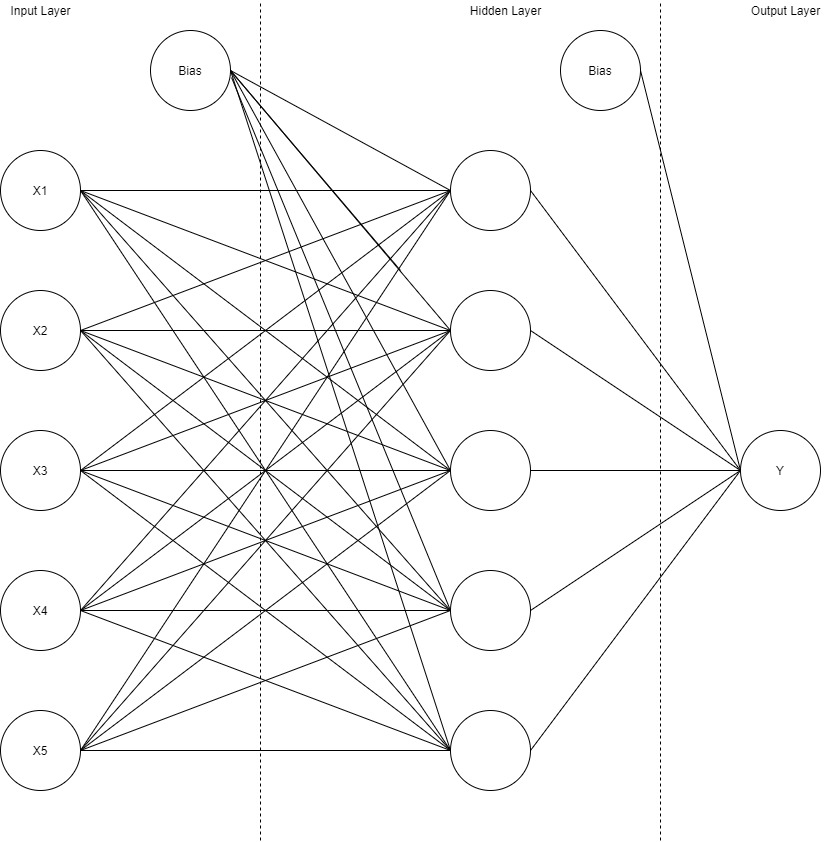


1. Apabila terdapat variabel yang lebih dari level signifikan model, maka variabel X tersebut dihilangkan dari model, hingga variabel X optimum tercapai



* Multilayer Perceptron

Rancangan arsitektur jaringan syaraf yang digunakan



Gambar 3 Arsitektur JST Dataset

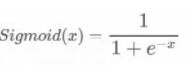
1. Feedforward

Input = 5 variabel

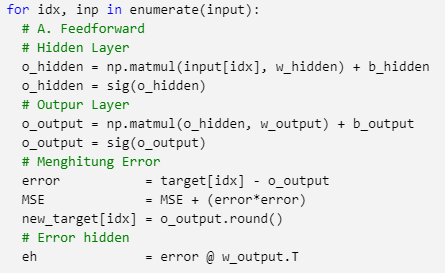
Hidden Layer = 5

Output = 1 variabel

Bobot pertama yang digunakan adalah nilai random, dengan range nilai minimal -1 dan maksimal 1. Untuk menghasilkan nilai output, maka setiap nilai input perlu di aktivasi dengan fungsi sigmoid biner.

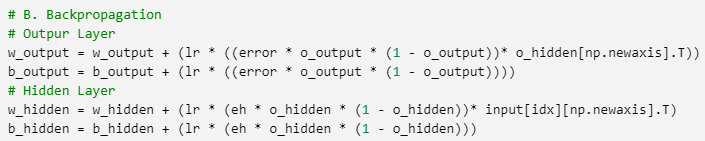


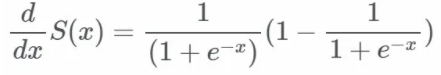
Gambar 4 Fungsi Sigmoid



1. Backpropagation

Pada tahap backpropagation, nilai bobot untuk output layer dan hidden layer di update berdasarkan nilai error.



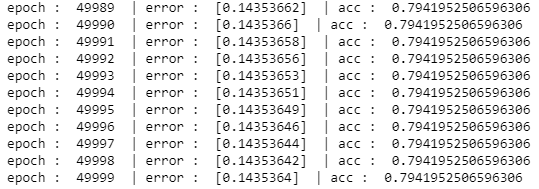


Gambar 5 Turunan Fungsi Sigmoid

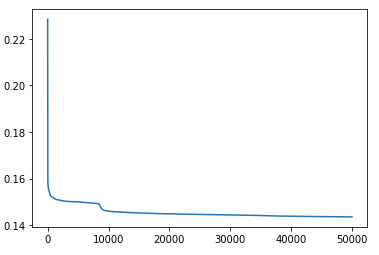
1. Implementasi pada Dataset

Epoch = 50000

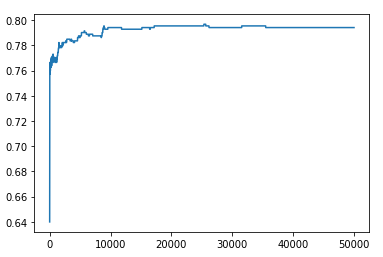
Learning Rate = 0,1



* + Nilai error

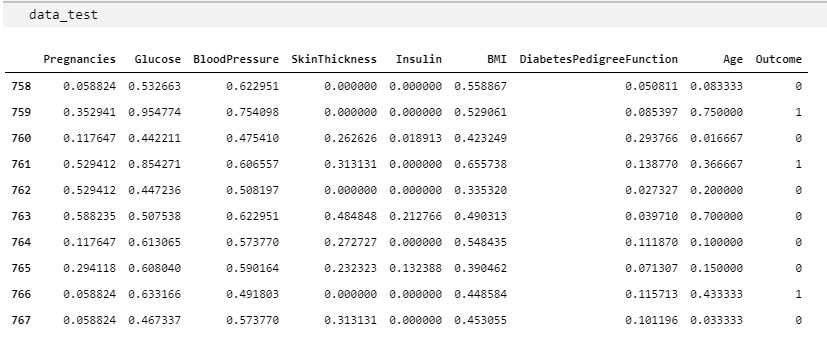


* + Nilai akurasi



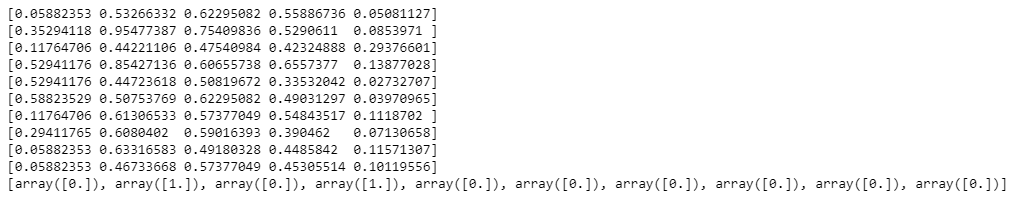
* + Prediksi pada Data Test

Data Test adalah sepuluh data terakhir yang tidak di training pada tahap sebelumnya.





Hasil prediksi data test



|  |  |
| --- | --- |
| Target | Output (Prediksi) |
| 0 | **0** |
| 1 | **1** |
| 0 | **0** |
| 1 | **1** |
| 0 | **0** |
| 0 | **0** |
| 0 | **0** |
| 0 | **0** |
| 1 | **0** |
| 0 | **0** |

Dari 10 data test yang di prediksi terdapat 1 data error, hal ini kemungkinan dikarenakan tingkat akurasi yang digunakan masih 79 %, belum mencapai 90 % atau lebih.

# BAB IV Kesimpulan dan Saran

1. **Kesimpulan**

Dari model yang telah dibentuk berdasarkan dataset, dihasilkan nilai akurasi 79,8 % dan error 0,14. Dari data test yang digunakan untuk menguji model, sebanyak 10 data yang diuji, terdapat 1 nilai error.

1. **Saran**

Sebaiknya terdapat pengembangan pada model, untuk memperbaiki tingkat akurasi prediksi.