**CASE BASED 1 PEMBELAJARAN MESIN CII3C3 [IZA]**

**Implementasi dan Analisis Algoritma Supervised Learning Dari Data Audit Risk**

**Artificial Neural Network With Python**



Oleh:

Zalfaa Putri Ayudhia

1301200301 - IF4412

**FAKULTAS INFORMATIKA**

**TELKOM UNIVERSITY**

**2022**

**Tugas ini dikerjakan dengan tidak melanggar aturan perkuliahan dan kode etik akademisi.**

# DAFTAR ISI

[**DAFTAR ISI**](#_ppc3ipnbhir) **1**

[**A. STRUKTUR PENELITIAN**](#_25g5i2l8lqhe) **2**

[1. Dataset](#_x352cjioer9g) 2

[2. Flowchart](#_w32ijf9rpj0e) 3

[**B. DATA PREPROCESSING**](#_vz17v7mch7z9) **4**

[1. Library](#_tvgx7jtwtf48) 4

[2. Read Data](#_22jadjydf7n6) 4

[3. Cek dan Isi Data NULL](#_3bdonv9ij4) 5

[4. Penyesuaian Tipe Data](#_yjsuafbtyzs) 6

[5. Outlier](#_jizmyqhnpupi) 6

[6. Skewness](#_6coinf88ts1n) 9

[7. Penggabungan Data Audit dan Trial](#_n7ty3rx333k0) 10

[8. Korelasi](#_tex98ldrbp33) 12

[**C. MODELLING**](#_env7z0m890dn) **12**

[1. Library](#_odd0wjmrkq6p) 12

[2. Split Data](#_s7du82v5qmjb) 12

[3. Scaling](#_jhj9ytjs2htn) 13

[4. Reduksi Dimensi](#_8a043b3bgrhg) 13

[5. Artificial Neural Network Model](#_xgmjpg2jugxu) 13

[6. Train Data](#_u8vkj1315v8r) 14

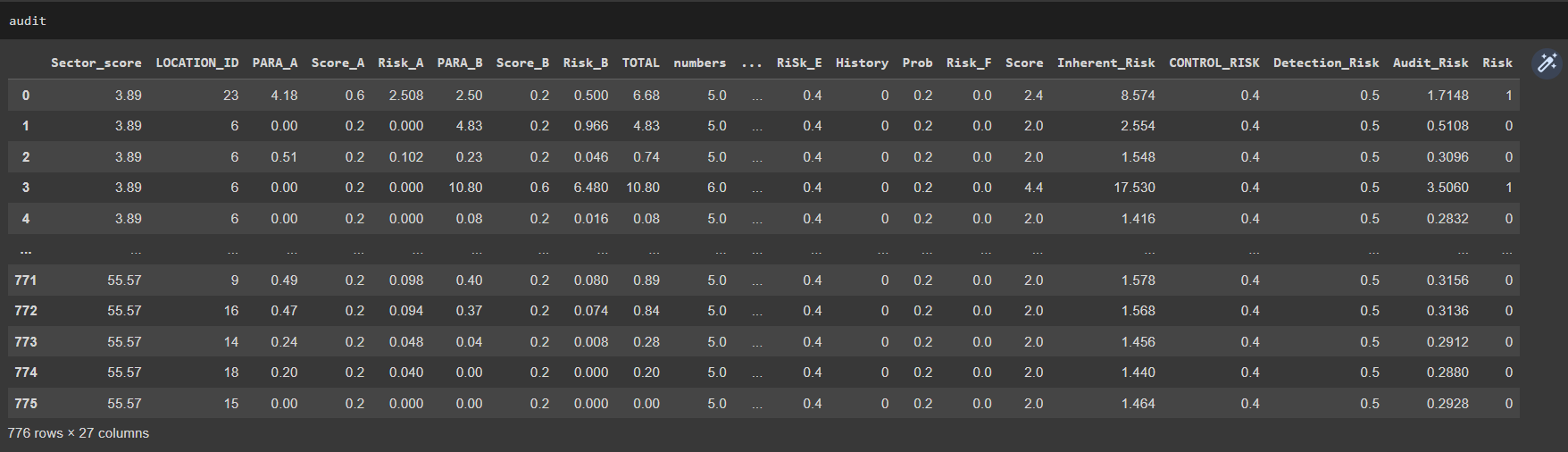
[7. Test data](#_rpbdklivo5eh) 14

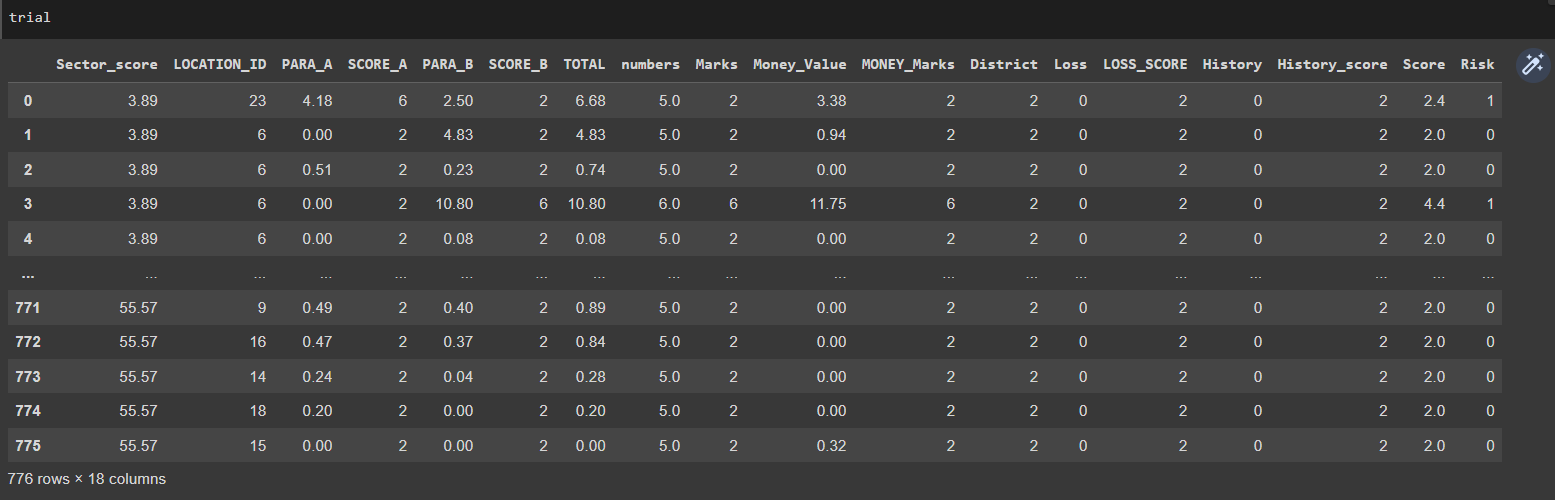
[**D. EVALUASI HASIL**](#_mb3u2wcdtfym) **14**

[**E. LINK COLABORATORY DAN PRESENTATION VIDEO**](#_pk3u42um08qk) **14**

# STRUKTUR PENELITIAN

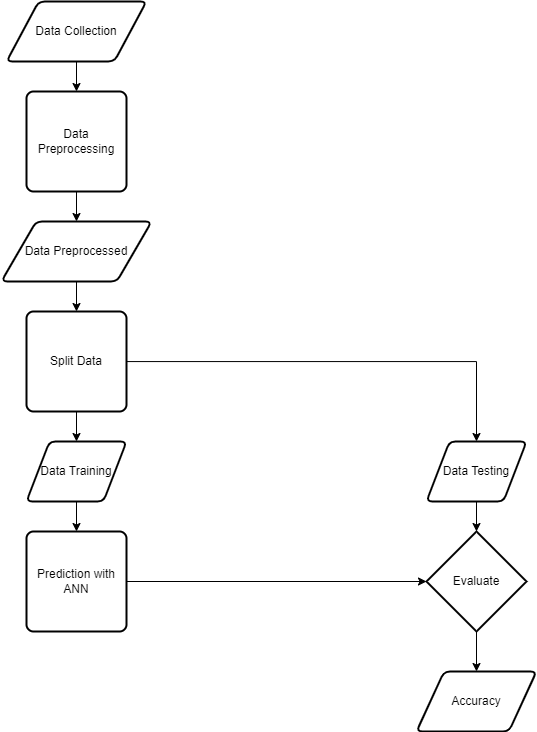
## Dataset





Kumpulan data yang dipilih dalam penelitian ini adalah data audit yang merupakan data non-rahasia selama satu tahun lengkap pada tahun 2015 sampai 2016 dimana perusahaan dikumpulkan dari kantor auditor India untuk membangun prediktor untuk mengklasifikasikan perusahaan yang mencurigakan. Banyak faktor resiko diperiksa dari berbagai bidang seperti catatan masa lalu kantor audit, audit-paras, laporan kondisi lingkungan, ringkasan reputasi perusahaan, laporan masalah yang sedang berlangsung, catatan nilai laba, catatan nilai kerugian, laporan tidak lanjut, dan lain-lain. Setelah wawancara mendalam dengan auditor, faktor-faktor risiko penting dievaluasi dan probabilitas keberadaannya dihitung dari catatan masa kini dan masa lalu.

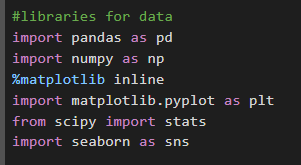
## Flowchart



Dari flowchart yang sudah diperlihatkan di atas, kumpulan data yang telah didapatkan akan melalui proses pra pemrosesan untuk mendapatkan data yang lebih berkualitas untuk diteliti. Setelah mendapatkan data yang telah di preprocess, data tersebut akan dibagi menjadi data training yang akan diproses dengan model Artificial Neural Network dan data testing yang akan dievaluasi untuk mendapatkan keakuratan.

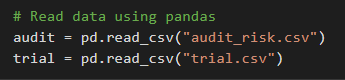
# DATA PREPROCESSING

## Library



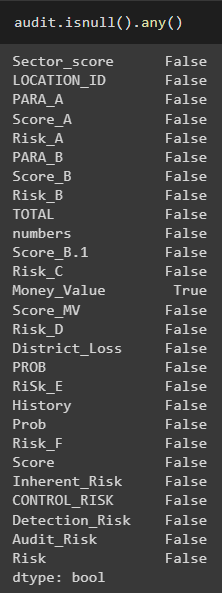
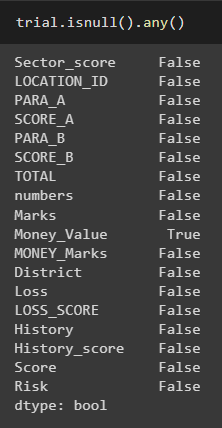
Library yang digunakan dalam praproses data ini adalah pandas untuk menampilkan dataframe dari data, numpy untuk proses fungsi matematika dan list, matplotlib dan seaborn untuk memvisualisasikan data, dan stats dari scipy untuk uji statistik pada data.

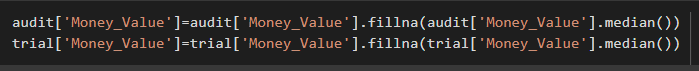
## Read Data



Fungsi read\_csv() pada pandas untuk membuat dataframe dari data audit dan trial.

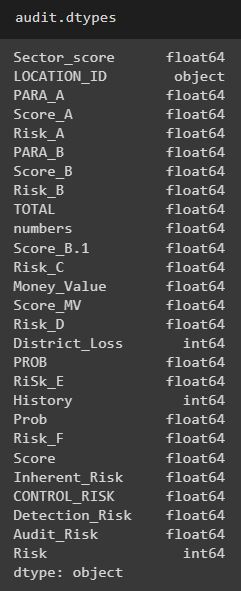
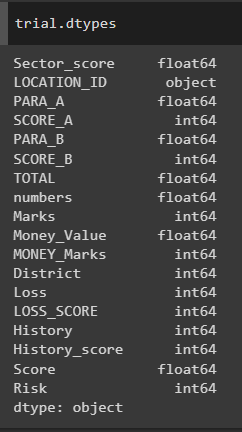
## Cek dan Isi Data NULL

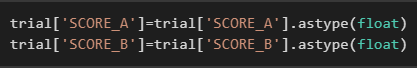
 



Pengecekan data null pada data audit dan trial memperlihatkan atribut Money\_Value memiliki nilai null, kemudian nilai atribut Money\_value yang null diisi dengan median dari data.

## Penyesuaian Tipe Data

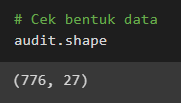
 

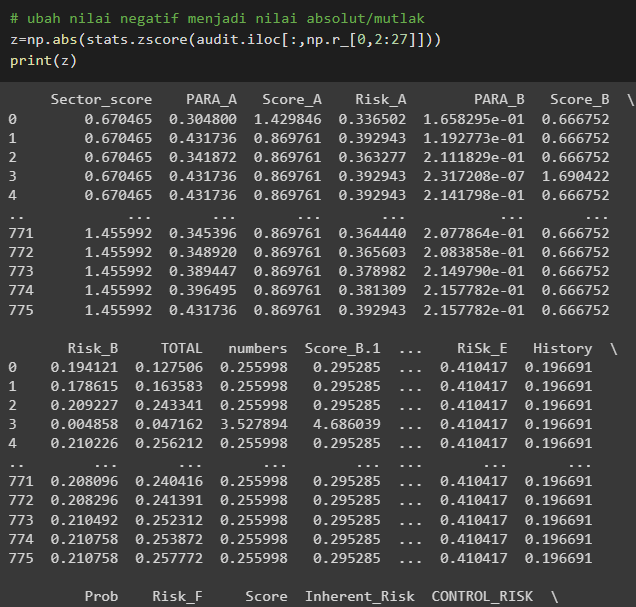


Pada pengecekan tipe data audit dan trial, terlihat tipe atribut score\_A dan score\_B di masing-masing data berbeda. Dalam hal ini, atribut tersebut pada data trial disesuaikan dengan tipe pada data audit yaitu float.

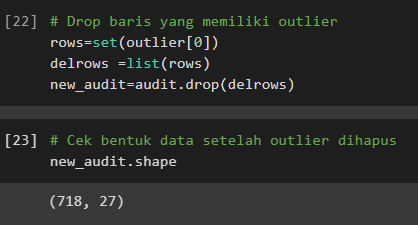
## Outlier

1. Data Audit



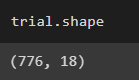


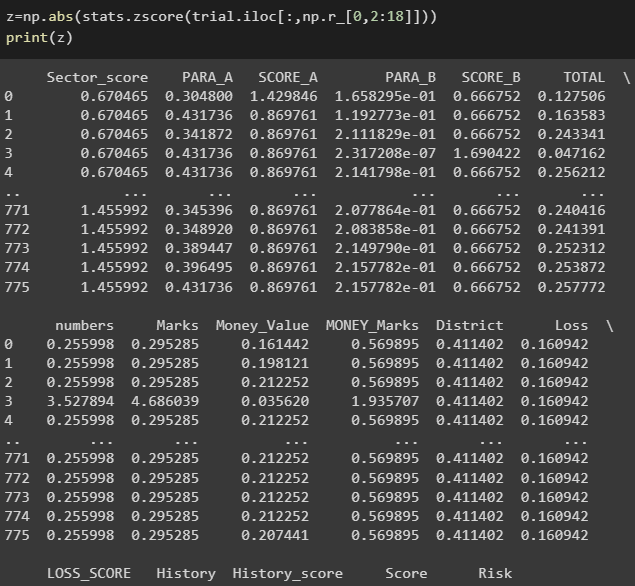


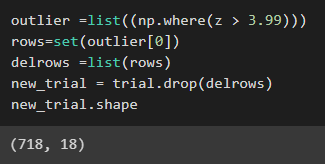


Baris data yang merupakan outlier dari data didrop/dihapus.

1. Data Trial



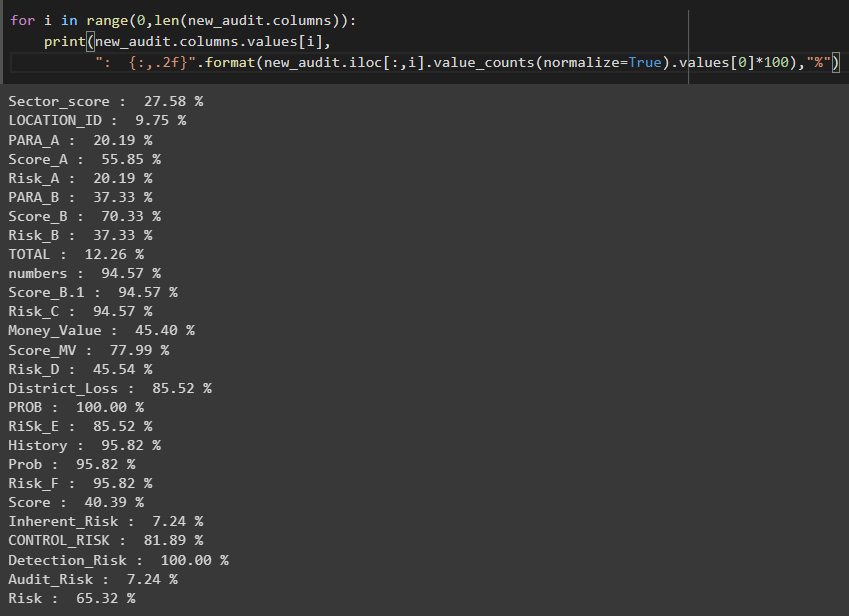


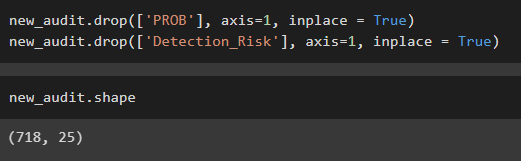


Sama dengan data audit, outlier pada data trial juga didrop/dihapus.

## Skewness

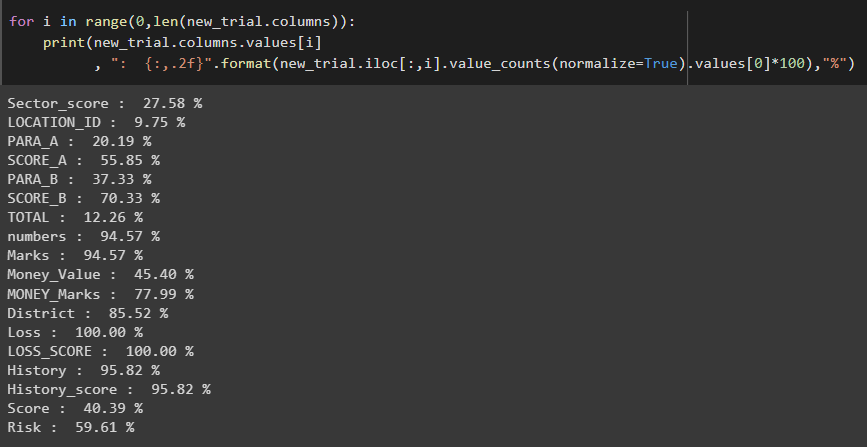
1. Data Audit

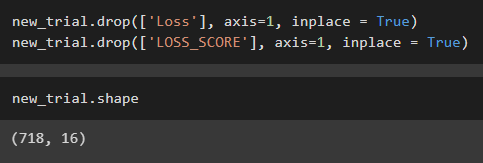




Data yang atributnya memiliki skewness sebesar 100% juga dihapus/didrop.

1. Data Trial

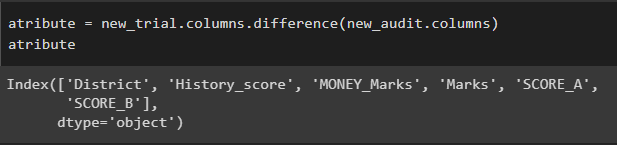




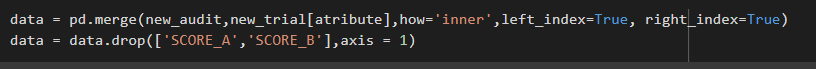
Sama dengan data audit, data trial yang atributnya memiliki skewness sebesar 100% didrop/dihapus.

## Penggabungan Data Audit dan Trial

1. Cek atribut yang sama



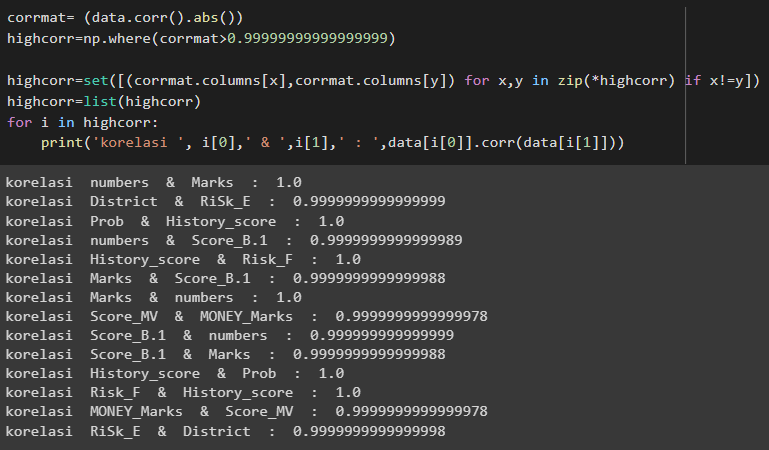
1. Merge



1. Info Data



## Korelasi

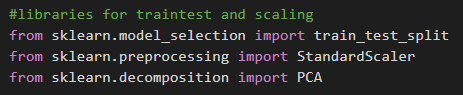




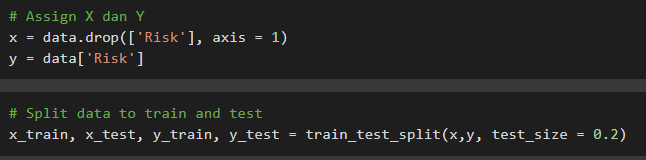
Atribut yang memiliki korelasi terlalu tinggi didrop/dihapus.

# MODELLING

## Library

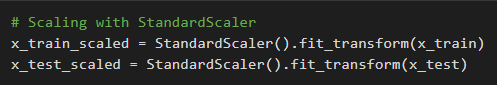


## Split Data



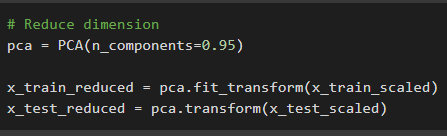
Assign data x dan y dimana x merupakan semua atribut kecuali atribut Risk yang merupakan targetnya (y). Setelah itu, data x dan y dibagi lagi menjadi data training dan data testing menggunakan fungsi train\_test\_split() dari library sklearn. Pembagian data adalah data testing sebesar 20% dan data training sebesar 80%.

## Scaling



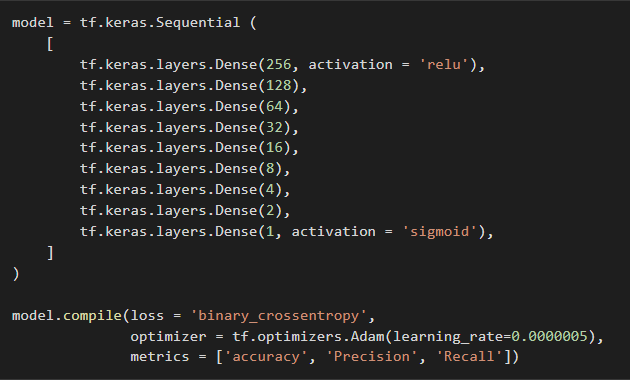
Scaling data menggunakan Standard Scaler.

## Reduksi Dimensi



Reduksi dimensi pada data x menggunakan PCA (Principal Component Analysis).

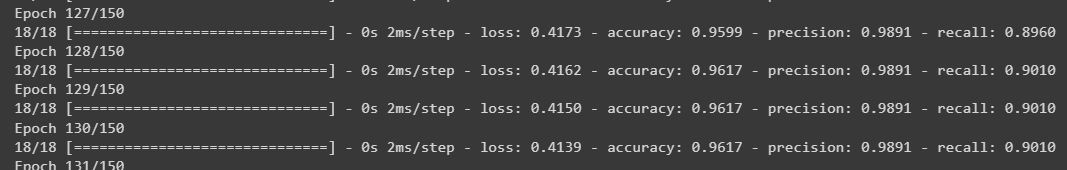
## Artificial Neural Network Model



Model Artificial Neural Network yang dibuat terdiri dari satu input layer, tujuh hidden layer, dan satu output layer dengan aktivasi input layer menggunakan relu dan aktivasi output layer menggunakan sigmoid. Dalam model tersebut, akan dicompile dengan mengeluarkan nilai accuracy, precision, dan recall.

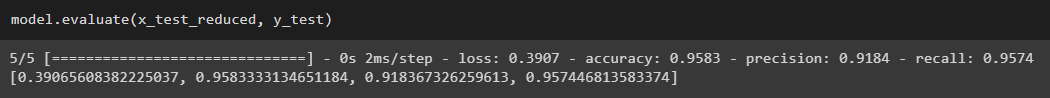
## Train Data





Proses data training mendapatkan akurasi maksimal 96% dan presisi 98% di epoch ke-128 dari 150 epoch.

## Test data



Evaluasi data testing mendapatkan presisi sebesar 91%.

# EVALUASI HASIL

Dengan menggunakan model Artificial Neural Network yang penulis buat, berhasil mendapatkan presisi sebesar dengan 98% pada data training dan presisi 91% pada evaluasi data testing. Dalam hal ini, baik atau tidaknya model ANN yang dibuat juga bergantung hidden layernya.

# LINK COLABORATORY DAN PRESENTATION VIDEO

Colaboratory:

<https://colab.research.google.com/drive/1qEpJ2QjFZyXoMNcfrr8iv2wDiWDzujSO?usp=sharing>

Presentation Video:

<https://drive.google.com/file/d/1luAP3P9tLOL0riPU-hr_FvlibVr3FHSm/view?usp=sharing>