**LAPORAN TUGAS PEMBELAJARAN MESIN**

**MENJELASKAN, IMPLEMENTASI, ANALISIS, SERTA DESAIN TEKNIK PEMBELAJARAN MESIN SUPERVISED LEARNING**

laporan ini dibuat untuk memenuhi kebutuhan nilai mata kuliah Pembelajaran Mesin

Dosen mata kuliah : **Tjokorda Agung Budi Wirayuda, S.M., M.T.**

**CII3C3-IF-44-11**



disusun oleh

**NAUFAL ADRIAN HIDAYAT – 1301204056**

**IF-44-11**

**FAKULTAS INFORMATIKA**

**TELKOM UNIVERSITY**

**2022**

**BAB 1**

**PENDAHULUAN**

**Case-Based 1**

**Scenario**

Mengikuti keberhasilan tugas sebelumnya, Anda diberi kesempatan lebih lanjut untuk mengesankan atasan Anda mengenai kemampuan Anda untuk menganalisis data. Anda diminta untuk melakukan beberapa analisis dan menghasilkan seperangkat aturan yang berguna menggunakan dataset berikut: Kumpulan data berikut tersedia online, tautan ke kumpulan data adalah sebagai berikut:

*https://archive.ics.uci.edu/ml/datasets/Arrhythmia*

Kemudian pilih teknik pra-pemrosesan data apa yang akan dilakukan untuk meningkatkan kualitas data tersebut. Akan ada banyak hal yang harus diuraikan dan kemudian Anda harus mengumpulkan *case-based* ini sebagai karya individu.

Hint: Anda bebas memilih satu dari tiga alat analisis data yaitu Weka, R, atau Python untuk membantu Anda menganalisis data dan menunjukkan pra-pemrosesan data yang diperlukan.

**Tugas Anda**

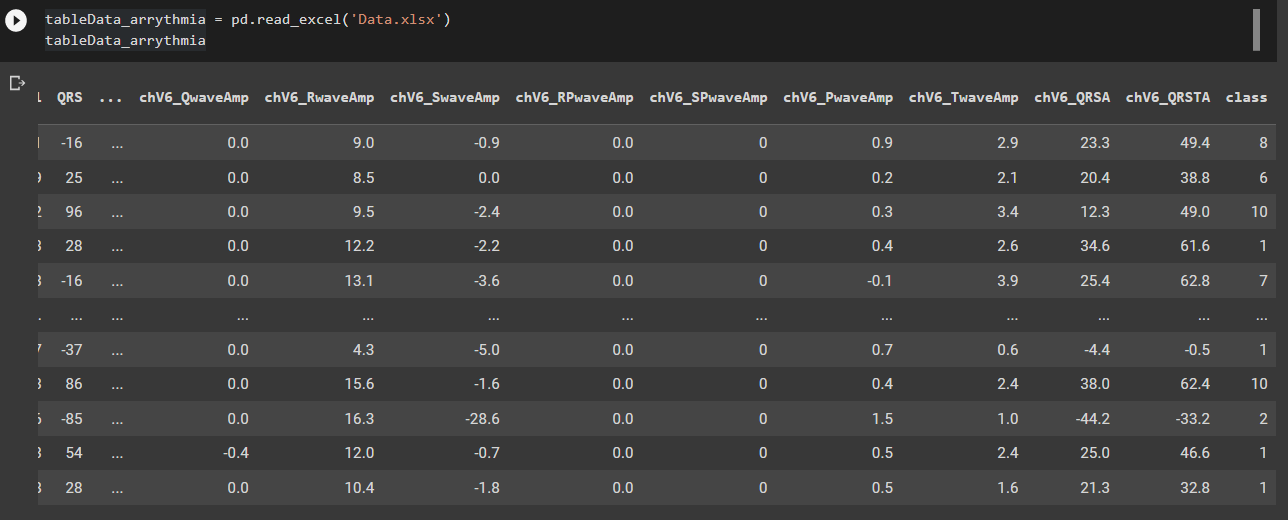
Tujuan dari tugas ini yaitu Anda diharapkan mampu menjelaskan, mengimplementasikan, menganalisis, dan mendesain teknik pembelajaran mesin supervised learning yaitu AAN/MLP/RNN/LSTM/CNN.

Pertama, selidiki masalah kualitas data yang telah diberikan di atas. Jelaskan keputusan Anda mengenai pendekatan pra-pemrosesan data. Jelajahi kumpulan data dengan meringkas data menggunakan statistik dan mengidentifikasi masalah kualitas data apa pun. Tidak ada batasan jumlah ringkasan yang akan dilaporkan tetapi Anda diharapkan hanya melaporkan yang paling relevan. Kedua, pilih salah satu dari metode unsupervised learning yang telah dipelajari yaitu AAN/MLP/RNN/LSTM/CNN. Anda hanya perlu memilih satu metode untuk diterapkan. Gunakan algoritma tersebut untuk memberikan beberapa output/outcome dengan menggunakan variasi hyperparameter, kemudian menganalisis hasilnya.

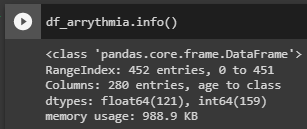
**BAB 2**

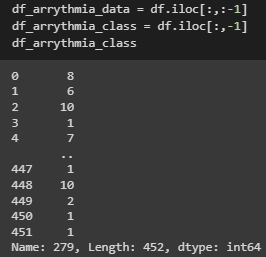
**PEMBAHASAN**

1. **Ikhtisar Data yang dipilih**   
    Case-based 1 yang saya dapat ialah data dengan NIM akhir genap, yakni “6”. Maka, saya memakai data Arrhythmia.data dan Arrhythmia.names sebagai acuan pada pengerjaan Case-Based berikut. data yang digunakan merupakan data yang sifatnya *supervised learning* punya nilai target yang ada pada kolom *class*, data tersebut merupakan data penyakit yang membedakan ada dan tidaknya penyakit *arrhythmia* jantung dan mengklasifikasikannya dalam salah satu dari 16 kelompok.
2. **Bentuk Data Table**

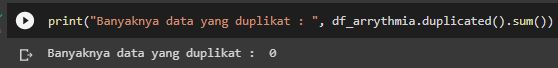
Berikut adalah dataset dari *arrythmia.data* yang bisa didapat dari format .csv :

Kemudian, pada tabel *arrhythmia.data* memiliki baris sebanyak 452 record dan juga kolom sebanyak 280 atribut serta pada dataset ini kita mendapat informasi terdiri sebanyak 280 kolom dengan tipe data float sebanyak 121 kolom dan integer 159 kolom.

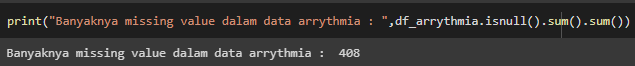


Lalu, saya mencoba memisahkan data dengan data aslinya yang terdiri data dari awal hingga data sebelum data terakhir yang dinamai dengan df\_arrythmia\_data serta df\_arrythmia\_class yang berisi kumpulan data paling akhir. Berikut adalah isi df\_arrythmia\_class : 

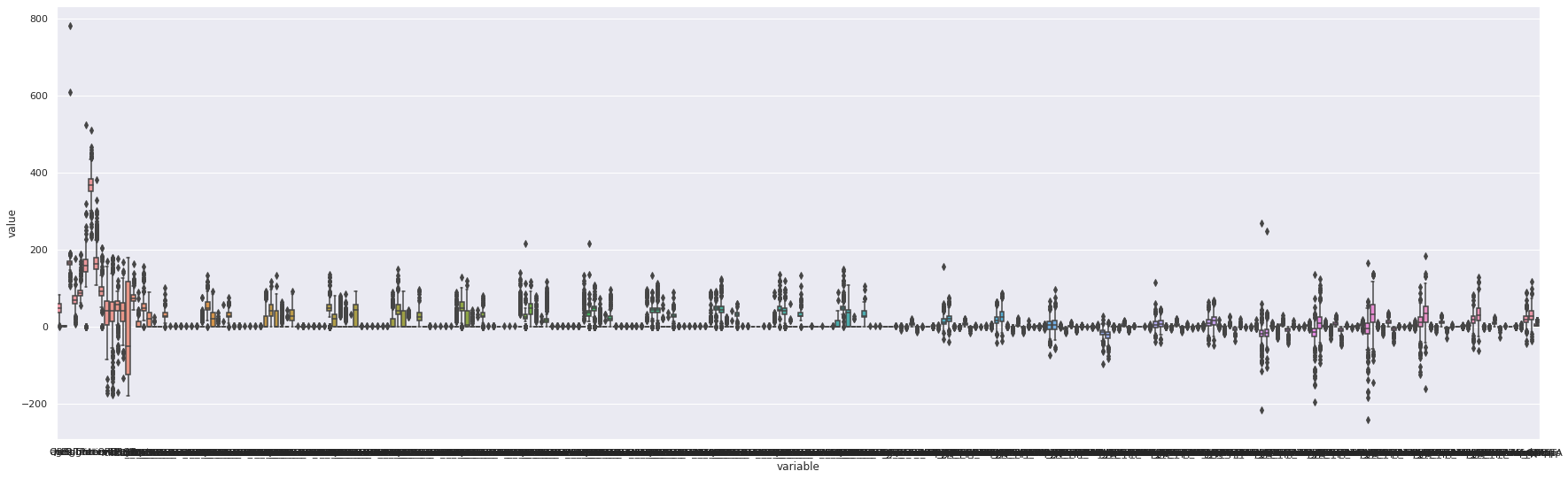
1. **Kualitas Data**
   1. **Mengecek Duplikasi Data**

Pada dataset *arrhythmia.data*, tidak ada data yang terduplikasi. Ketika dilakukan pengecekan data yang duplikasi, keluar *outputnya* ialah “0” yang menandakan tidak adanya data yang terduplikat pada *arrhythmia.data* tersebut.

* 1. **Mengecek *Missing Value* Data**

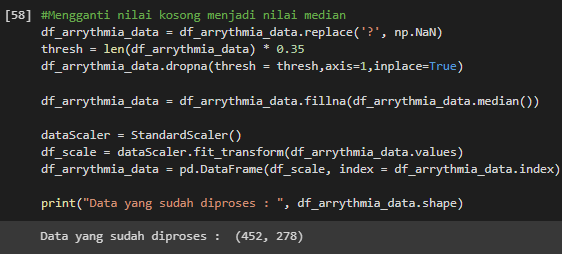
Pada dataset *arrhythmia.data*, data tersebut ternyata memiliki nilai *missing value* atau data kosong pada data tersebut. Pada data tersebut, ditemukan sebanyak 408 data kosong (*missing value*) dari 208 kolom yang terdapat pada *arrhythmia.data*. data kosong sebanyak 408 data tersebut yang nantinya akan diproses dan diatur dalam *pre-processing* data.

1. **Visualisasi Data**

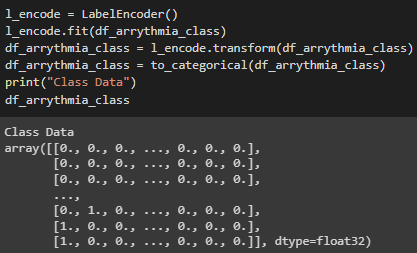
Pada bagian ini, ditampilkannya bagaimana penyebaran data pada dataset *arrythmia.data*  dengan memakai boxplot. Jika dilihat dari boxplot, tampak banyak sekali pencilan data (Outliers) yang jarak antar datanya jauh satu sama lain. Karena itu, mengatasi hal tersebut dibutuhkan normalisasi pada tahap *pre-processing* data.

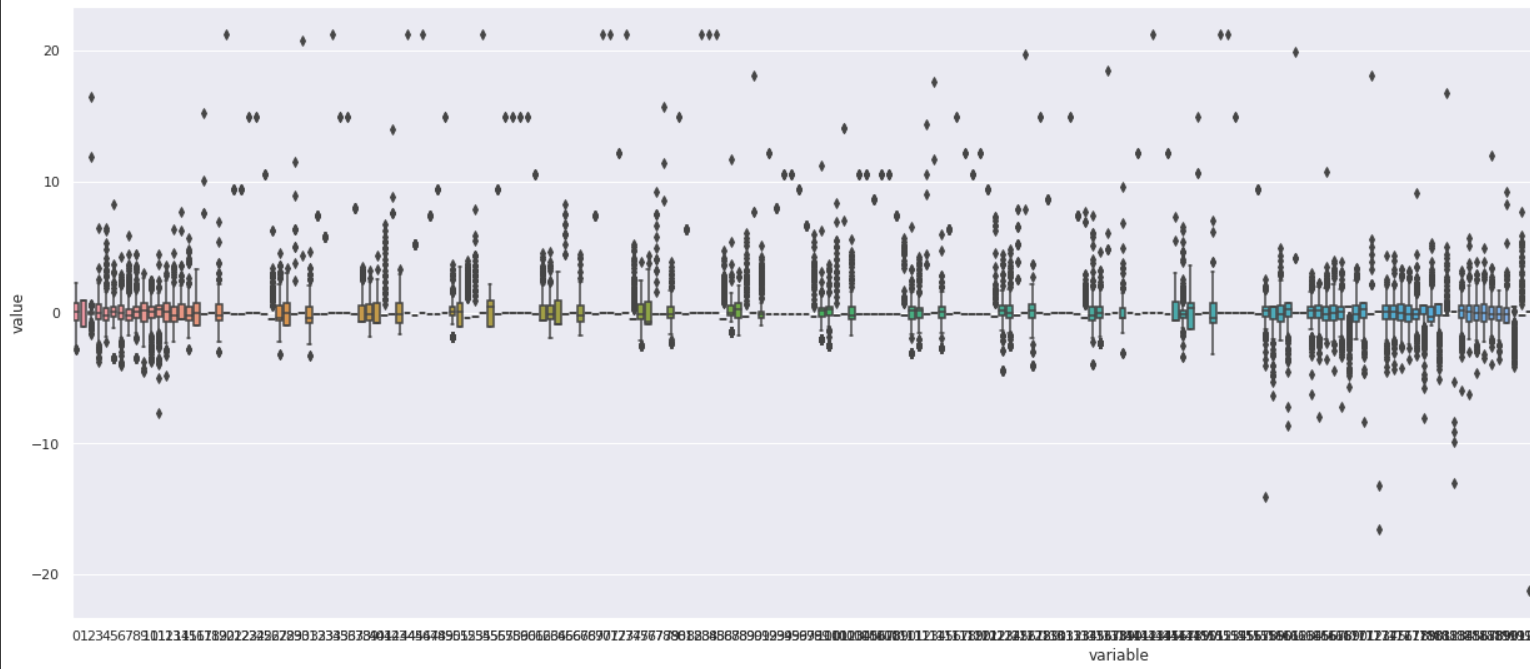
1. **Ringkasan Pra-Pemrosesan Data**
2. **Pre-Processing dan *Clean* Data**

Pada subbab berikut, saya memakai beberapa cara dan metode yang saya harap bisa membuat data yang saya olah menjadi lebih baik sebelum dibuat model pada tahap selanjutnya. Pada metode ini, saya menemukan adanya data yang kosong ditandai dengan '?'. Sebelum memproses data-data yang kosong tersebut, saya melakukan replace dari yang awalnya data yang kosong ditandai dengan '?' menjadi “NaN”. Setelah diubah menjadi “Nan”, saya melakukan drop salah satu kolom dengan melakukan threshing, yakni menghapus kolom yang tidak diinginkan dengan *delete* atribut data yang mempunyai *missing value* lebih dari 35%. Setelah itu saya melakukan proses replace dengan mengganti data yang kosong dengan median dari dataset.

Setelah selesai mengganti data, saya mencoba normalisasi data tersebut dengan memakai *library “*StandardScaler” yang berfungsi sebagai normalisasi, yakni membuat data dapat normal atau jarak pencilan dan data tersebut tidak terlalu jauh.

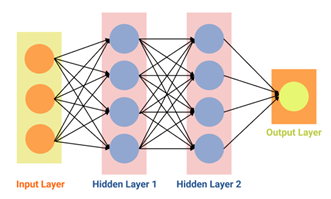
Lalu, mengubah bentuk data class yang tadinya berbentuk biner menjadi bentuk “categorical” data supaya proses pemodelan datanya bisa dilakukan. Karena data dari data class bersifat “binary”, makanya harus diubah terlebih dahulu data tersebut menjadi “categorical”. Memakai *library* “sklearn” untuk mengubah data tersebut.



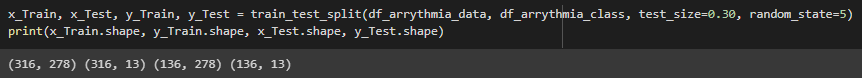
Kemudian, data tersebut jika dipresentasikan ke dalam bloxplot akan terlihat seperti gambar di bawah, bahwa persebaran data dan pencilan pada boxplot sudah tidak terlalu jauh dan datanya cukup merata.

1. **Menerapkan algoritma yang dipilih**
2. **Pembuatan Model Data**
   1. **Algoritma Artificial Neural Network (ANN)**

ANN (Artificial Neural Network) merupakan model algoritma yang mencoba meniru sebuah otak manusia yang mampu memberikan stimulasi atau rangsangan, melakukan proses, serta memberikan output untuk menemukan hubungan dari kumpulan data. Pada algoritma ANN ini, terdapat empat layer, yakni input layer, hidden layer 1, hidden layer 2, dan output layer.

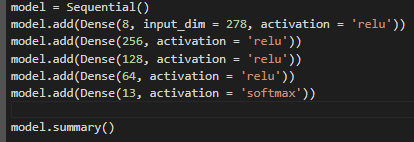


* 1. **Split Data (Training dan Set)**

Pada bagian selanjutnya adalah membuat split data dengan membagi data tersebut menjadi data training dan data set. Split data ini dilakukan dengan perbandingan 30% data test dengan 70% data training. Proses splitting data ini bertujuan untuk membantu memastikan bahwa model yang dibuat sudah akurat supaya modelnya nanti bisa dipakai pada proses selanjutnya.

* 1. **Membangun Model Data**

Pada tahap ini merupakan tahap membangun model data, model dibangun berdasarkan data yang sudah dilakukan *split* data sebelumnya dan proses pembangunan model ANN ini dibantu dengan *library* dari “keras import Sequential”. Maka, didapat cara untuk membangun model data seperti pada gambar di bawah ini. Beberapa activation function yang dipakai pada pemodelan data ini ialah relu, relu, relu, relu, dan softmax pada tiap masing-masing layer. Saya memakai 278 data (kecuali data class) yang digunakan untuk membangun dimensi dari pemodelan data yang dibuat.

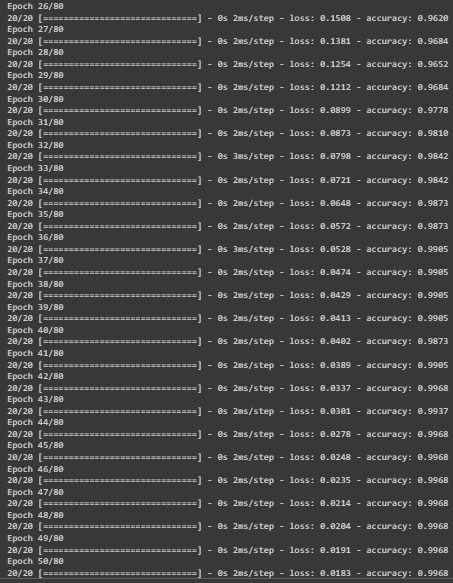
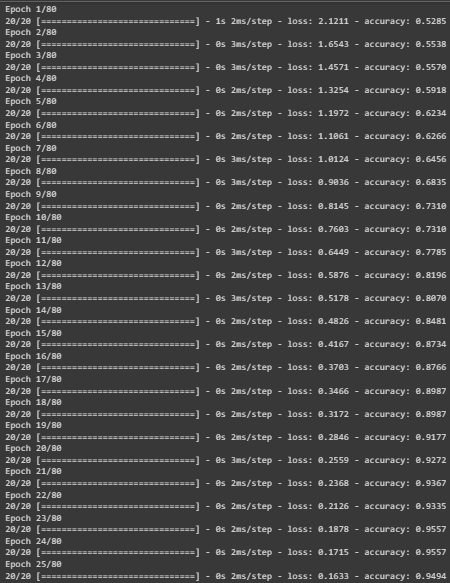


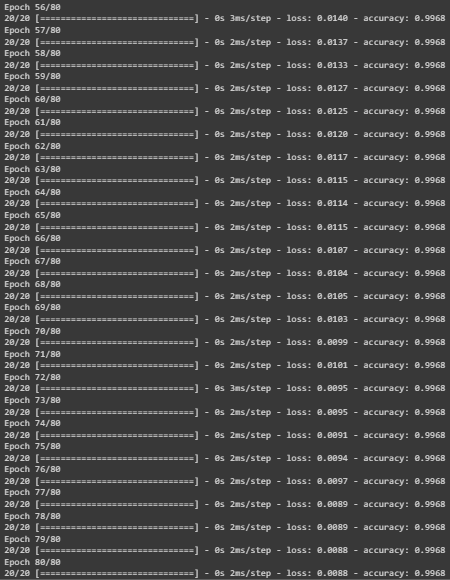
1. **Evaluasi Hasil**

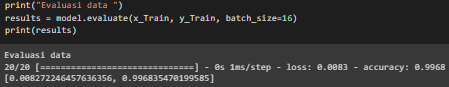
Pada bagian ini, tahap yang dilakukan ialah evaluasi hasil akurasi serta *loss* untuk mengetahui apakah model yang sudah dibuat pada tahap-tahap sebelumnya akurat dan bagus. Berikut beberapa uji coba dan hasil evaluasi dari data train dan data test.

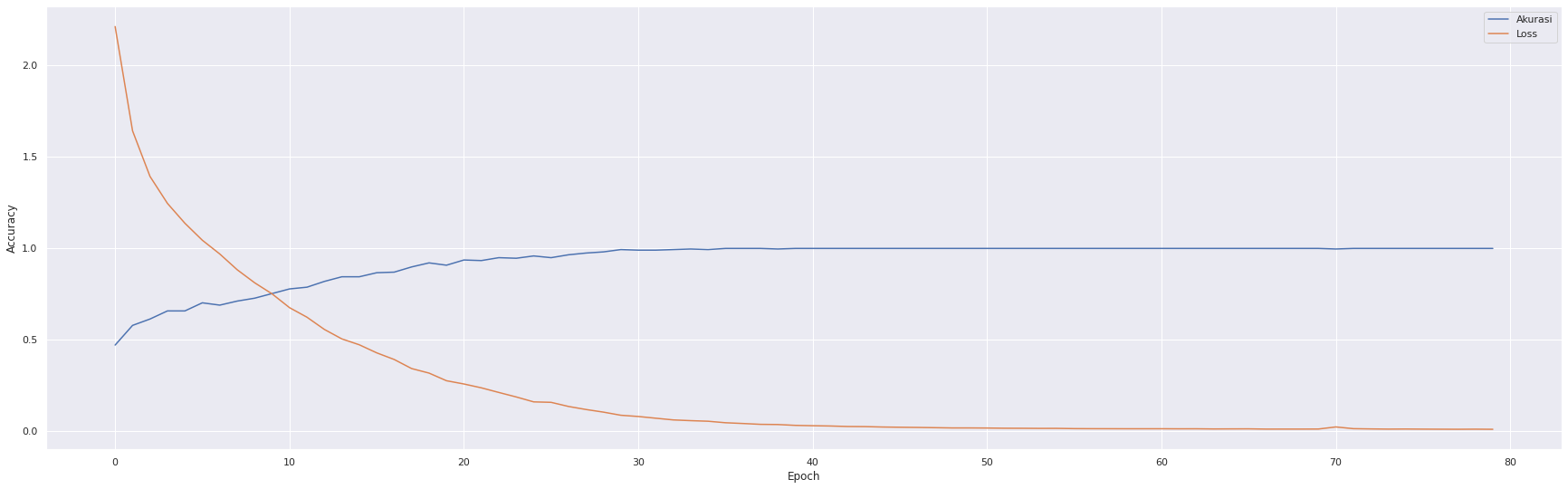
1. **Compile Model Data Train**

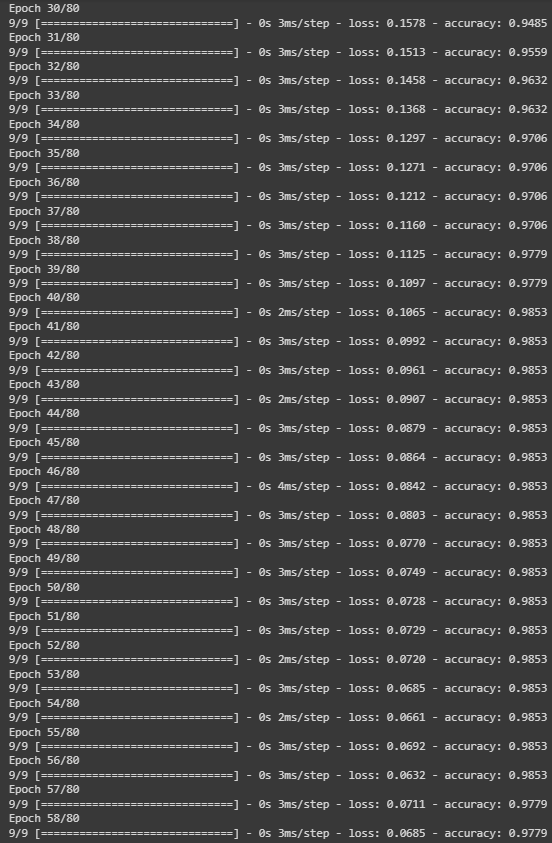
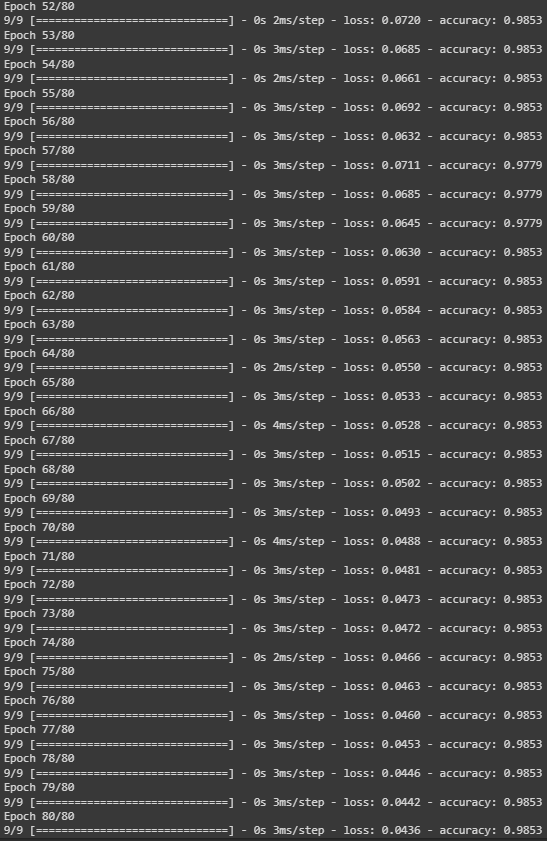
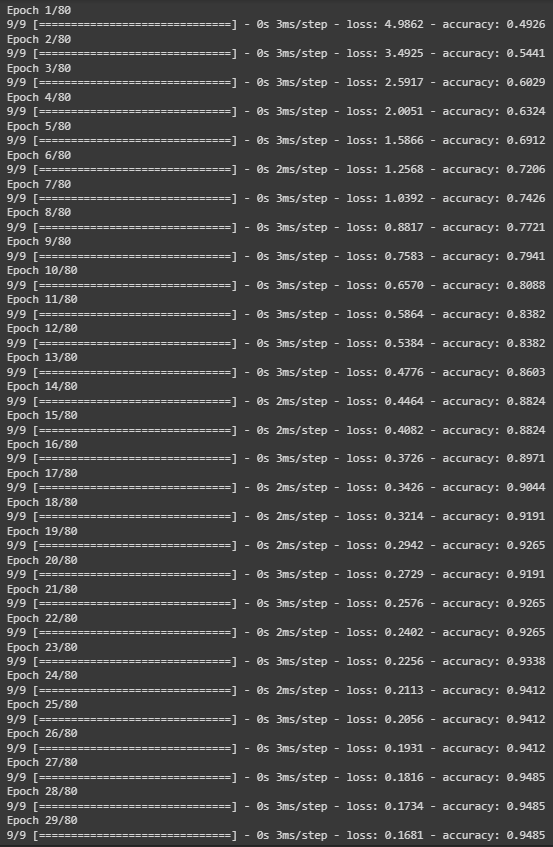
Ketika melakukan compile data pada model data train, saya melakukan iterasi dengan memakai *epoch* 80 dengan Batch\_size yakni untuk membagi beberapa data dari data set yang dipakai untuk pemodelan data secara bertahap sebanyak 16.

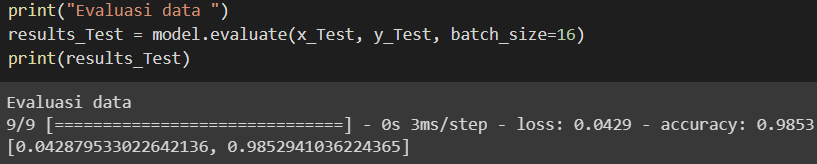
* 1. **Hasil Percobaan Data Train** 

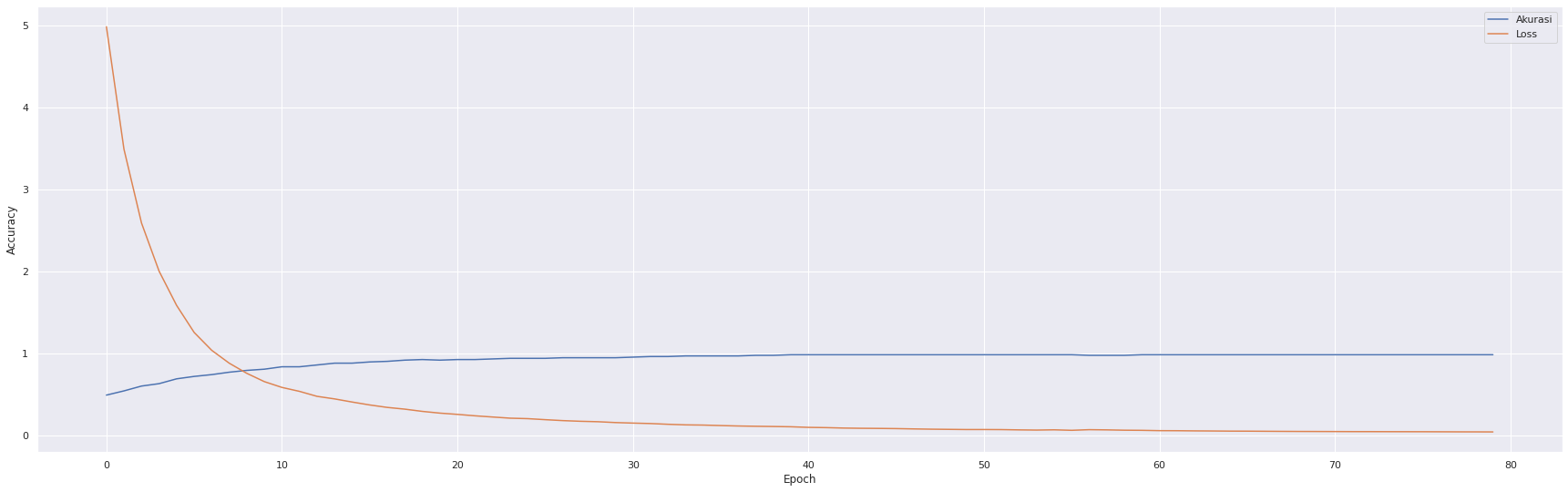




Pada hasil percobaan data train di atas, bisa dilihat bahwa akurasi serta *loss* yang diterima cukup baik, terbukti pada grafik gambar berikut yang mana tingkat akurasi mendekati 1.0 serta *loss* yang mendekati angka 0. Kemudian evaluasi model data train ini mencapai loss sebesar 0.0083 dengan akurasi sebesar 0.9968. Berarti bisa dikatakan model dari data train yang dibuat ini cukup bagus dikarenakan data loss semakin berkurang serta akurasi yang semakin meningkat.

1. **Compile Model Data Test**
   1. **Hasil Percobaan Data Test** 



****

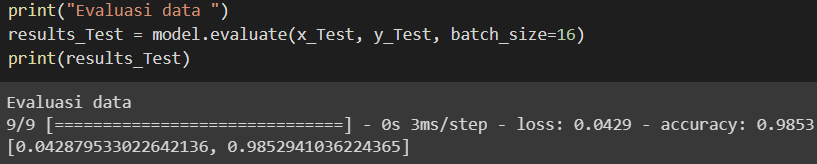
Pada hasil percobaan data test di atas, bisa dilihat bahwa akurasi serta *loss* yang diterima cukup baik, terbukti pada grafik gambar berikut yang mana tingkat akurasi mendekati 1.0 serta *loss* yang mendekati angka 0. Kemudian evaluasi model data test ini mencapai loss sebesar 0.0420 dengan akurasi sebesar 0.9853. Berarti bisa dikatakan model dari data test yang dibuat ini cukup bagus dikarenakan data loss semakin berkurang serta akurasi yang semakin meningkat.

**BAB 3**

**KESIMPULAN**

1. **Kesimpulan**

Setelah melakukan proses evaluasi pada data test dan juga data tein dengan iterasi (epochs) sebanyak 80 kali diperoleh hasil evaluasi dengan tingkat akurasi 0.9853 dan *loss* sebesar 0.0429 yang di mana hasil tersebut dapat disimpulkan bahwa model yang sudah dibuat sudah cukup akurat dan bagus.

Bagus atau tidaknya tingkat akurasi dan *loss* sendiri dipengaruhi oleh kualitas data serta metode pemodelan yang digunakan serta processing data yang dipakai dan juga banyaknya iterasi dan batch\_size yang dipakai juga mempengaruhi tingkat akurasi data dan tingkat *loss* dalam melakukan *compile* data tersebut.

**Link Codingan (Google Collabs) :**

[*https://colab.research.google.com/drive/1G\_AT3XQLlTsA0NRIiCxoThIr3Hmnr6-5?usp=sharing*](https://colab.research.google.com/drive/1G_AT3XQLlTsA0NRIiCxoThIr3Hmnr6-5?usp=sharing)

**Link PowerPoint :**

[*https://docs.google.com/presentation/d/12oLUjhlNcNARQcbcjnAFmydjqamIcjkX/edit?usp=sharing&ouid=100359807533626157288&rtpof=true&sd=true*](https://docs.google.com/presentation/d/12oLUjhlNcNARQcbcjnAFmydjqamIcjkX/edit?usp=sharing&ouid=100359807533626157288&rtpof=true&sd=true)

**Link Video (Youtube) :**

[*https://youtu.be/dfLXkky-xdM*](https://youtu.be/dfLXkky-xdM)