**LAPORAN TUGAS**

**“EKSTRAKSI FITUR”**

**MACHINE LEARNING**

****

**Disusun Oleh:**

Abdullah Azzam

2141720073

TI-3E

**PROGRAM STUDI D-IV TEKNIK INFORMATIKA**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

**2023**

**Praktikum 1**

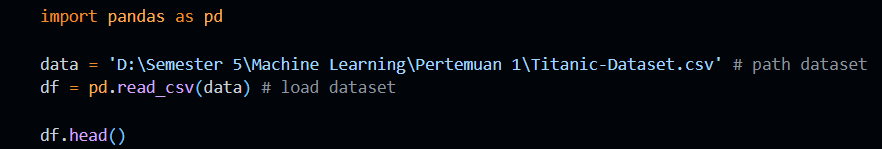
**Pra Pengolahan Data - Data Terstruktur**

## Data Imputation

Pada praktikum ini kita akan mencoba melakukan data imputation terhadap data yang hilang (missing value). Dataset yang digunakan adalah dataset Titanic.

### Langkah 1 - Load Data

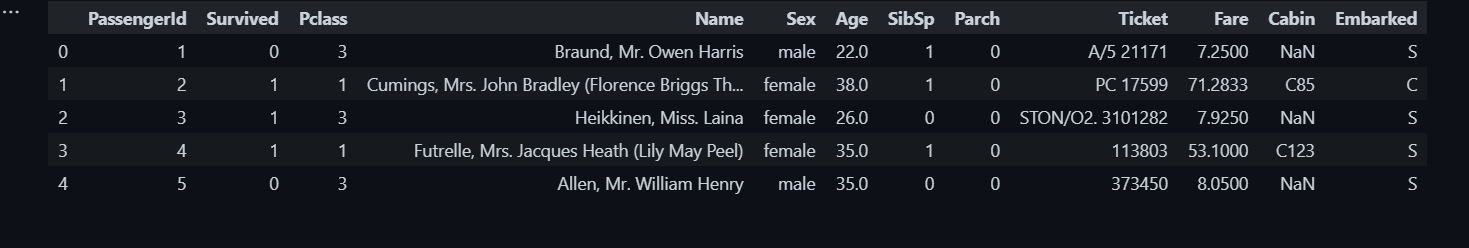
Import library yang dibutuhkan kemudian, load data kedalam DataFrame Pandas.



Tampilkan data teratas dengan perintah df.head()



Maka akan muncul tampilan seperti dibawah ini

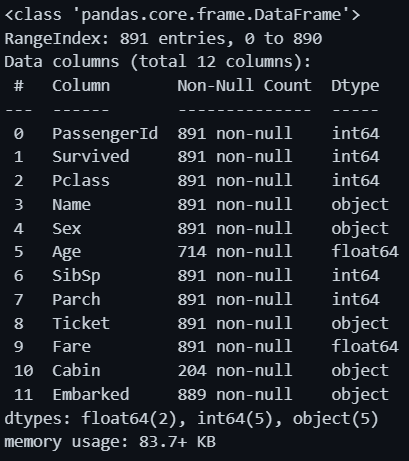


### Langkah 2 - Pengecekan Data

Lakukan pengecekan informasi mengenai data dengan menggunakan df.info()



Perhatikan hasil dari perintah yang telah dijalankan.



Dari df.info() kita dapat mengetahui jumlah data untuk setiap kolom. Selain itu, kita mengetahui bahwa,

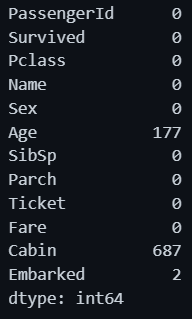
* "Age" berjenis float64
* "Cabin" berjenis object
* "Embarked" berjenis object

**Pada Pandas, object sering kali mewaliki tipe data string. Kita harus cermat dalam hal ini. Hal ini berkaitan dengan teknik imputasi yang akan kita gunakan nantinya.**

Selanjutnya, untuk lebih mengetahui jumlah data yang hilang untuk setiap kolom, gunakan perintah



Didapatkan informasi seperti berikut,



Dari informasi ini, kita dapat mengetahui data yang hilang sejumlah,

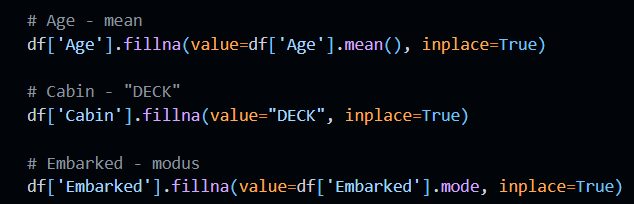
* 177 data "Age"
* 687 data "Cabin"
* 2 data "Embarked"

### Langkah 3 - Imputasi

Pada langkah ini kita akan melakukan imputasi terdapat data yang hilang pada "Age", "Cabin", dan "Embarked".

Strategi yang akan kita gunakan adalah,

* "Age" --> Dikarenakan "Age" adalah data nominal, maka kita akan menggunakan strategi mean.
* "Cabin" --> "Cabin" merupakan informasi terkait dengan nomor kabin penumpang. Disini kita akan berasumsi bahwa, seluruh penumpang yang tidak memiliki nomor kabin, merupakan penumpang yang tinggal di dek-dek kapal. Sehingga kita akan mensubtitusi data yang hilang dengan informasi "DECK".
* "Embarked" --> "Embarked" merupakan informasi lokasi embarkasi penumpang. Hanya ada 2 data yang hilang disini. Dikarenakan data merupakan data nominal, maka kita dapat menggunakan modus (mode) untuk mensubtitusi data yang hilang.



### Langkah 4 - Validasi Hasil

Lakukan pengecekan kembali dengan df.info(). Selanjutnya, kita juga dapat memastikan kembali apakah nilai yang disikan sudah benar dengan melakukan pencekan data secara langsung. Hasil dari proses imputasi adalah sebagai berikut,

****

# **Praktikum 2**

Seleksi Fitur, Encoding, dan Standarisasi

Seleksi Fitur

Pada tahap ini kita akan memilih variabel mana saja yang akan kita gunakan sebagai fitur. Pada bagian dasar teori, kita sudah mempelajari bahwa tidak selalu semua variabel dapat digunakan sebagai fitur. Oleh karena itu, kita dapat memilih fitur yang sekiranya penting untuk tujuan yang ingin kita capai dalam membuat model pembelajaran mesin. Kita akan menggunakan kasus yang sama, yaitu kita kita mengetahui ciri-ciri penumpang kapal Titanic yang selamat. Sehingga, fitur yang kita gunakan adalah,

1. “Survived”
2. “Pclass”
3. “Sex”
4. “Age”
5. “Cabin”

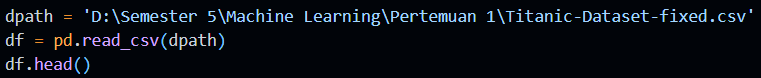
Kita akan memisahkan variabel-variabel tersebut dari variabel yang lain.

Data yang akan kita gunakan adalah data Titanic yang sudah diperbaiki dengan proses imputasi.

### Langkah 0 - Load Library



### Langkah 1 - Load Data





### Langkah 2 - Slice Data

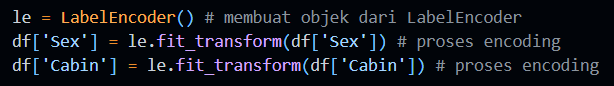


Hasilnya



### Langkah 3 - Encoding

Kita akan melakukan encoding pada data-data kategorikal, yaitu "Sex" dan "Cabin". Langkah ini perlu dilakukan karena algoritma pembelajaran mesin melakukan proses kalkulasi secara matematis. Strategi yang kita gunakan adalah **Label Encoding**. Label Encoding serupa dengan Ordinal Encoding, bedanya hanya pada proses pengurutan. Label Encoding tidak mengurutkan data terlebih dahulu.



### Langkah 4 - Verifikasi Hasil

Cek kembali dengan df.head(). Hasilnya akan seperti pada Gambar dibawah ini



### Langkah 5 - Standarisasi

Pada tahapan ini kita akan melakukan standarisasi pada "Age". **Hal ini sebetulnya tidak terlalu diperlukan karena nilai pada "Age" tidak terlalu jauh.** Tujuan dari langkah ini adalah untuk demonstrasi proses standarisasi.



### Langkah 6 - Verifikasi Hasil Standarisasi

Cek kembali dengan df.head(). Hasilnya akan seperti pada Gambar dibawah ini.



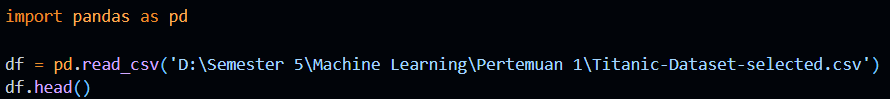
# Praktikum 3

Spliting Data

Pada pratikum ini, kita akan mencoba 3 macam strategi pembuatan data latih, validasi, dan uji. Data yang digunakan adalah data kapal Titanic yang sudah kita ekstraksi fiturnya pada pratikum sebelumnya

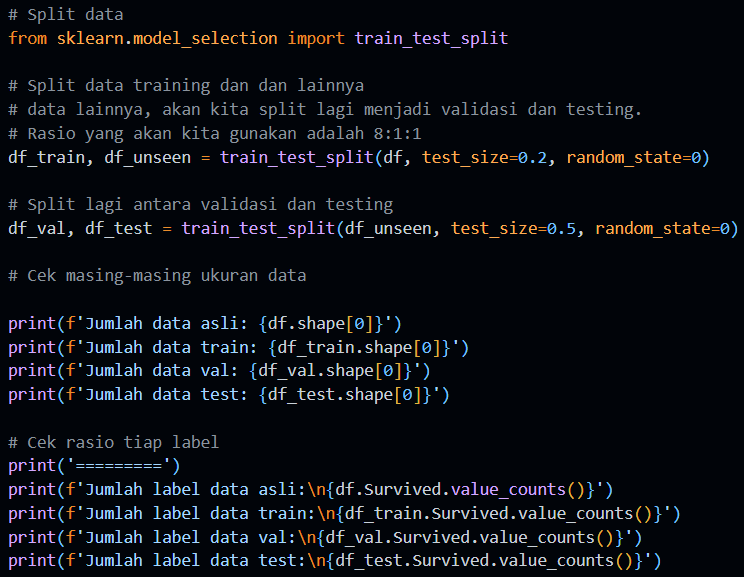
## Random Split

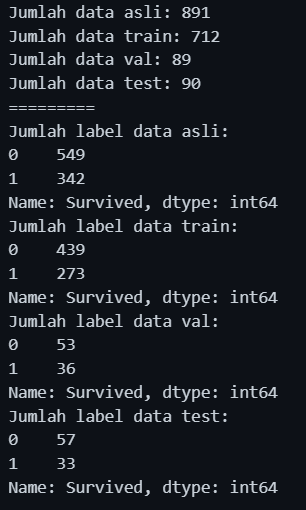
### Langkah 1 - Load Data





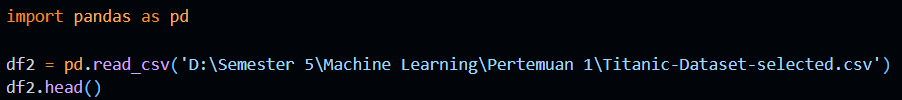
### Langkah 2 - Split Data





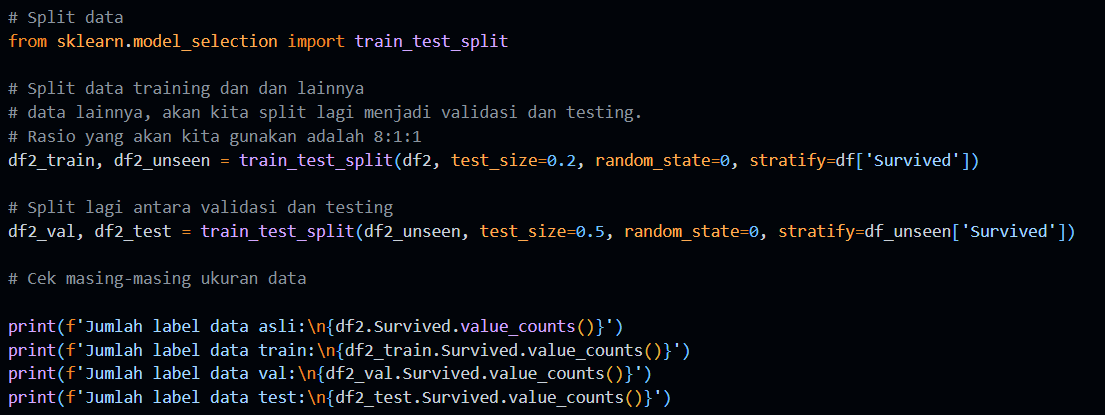
## Stratified Split

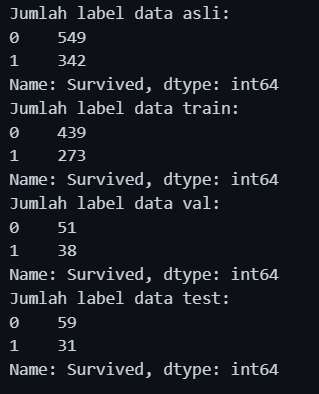
### Langkah 1 - Load Data





### Langkah 2 - Split Data

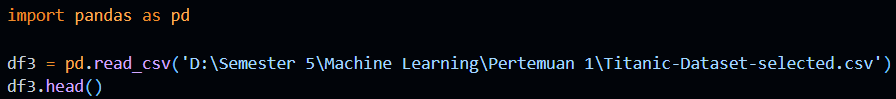




## Cross Validation 1

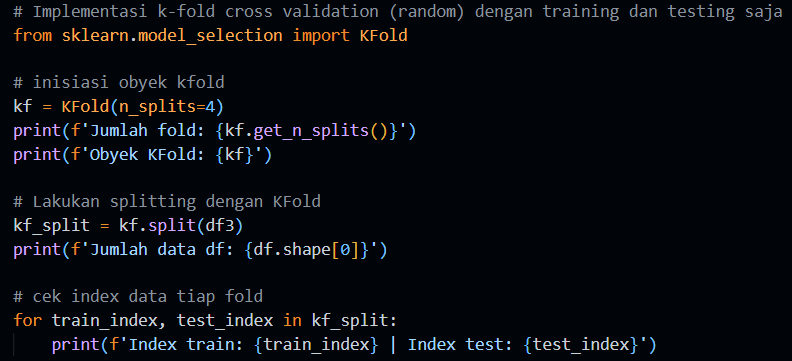
Pada praktikum ini, kita akan membuat data latih dan data uji saja dengan menggunakan metode cross validation.

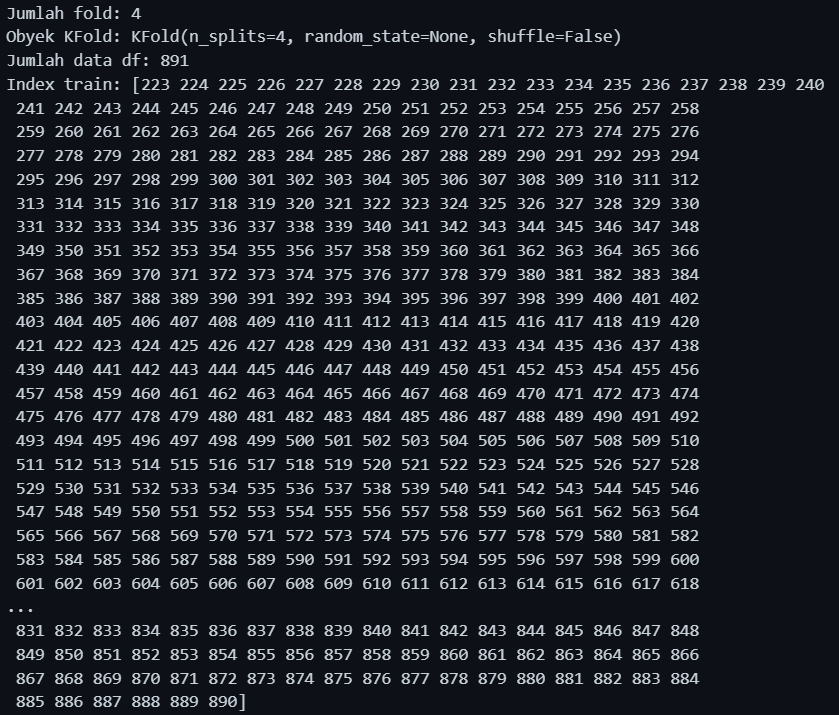
### Langkah 1 - Load Data





### Langkah 2 - Split Data

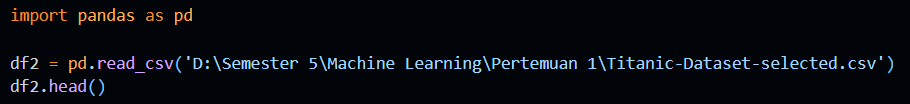




## Cross Validation 2

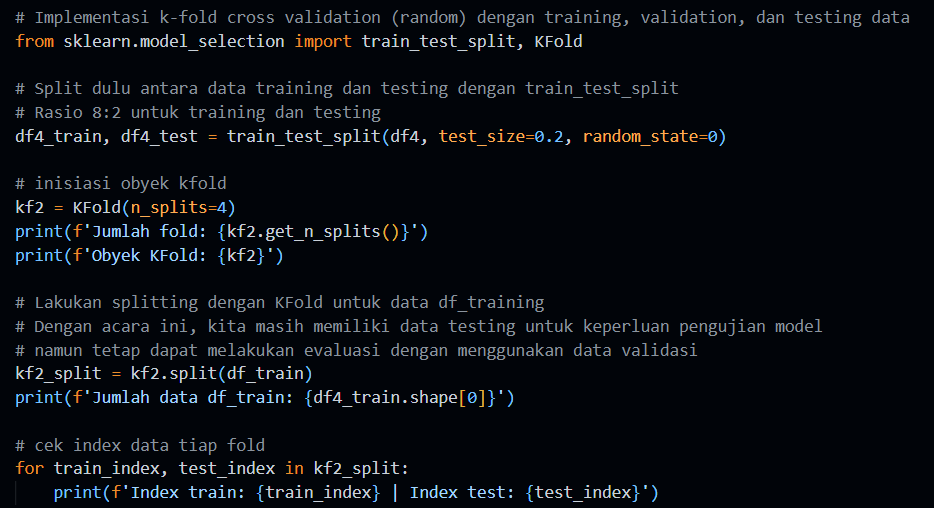
Pada praktikum ini, kita akan membuat data latih dan data uji saja dengan menggunakan metode cross validation.

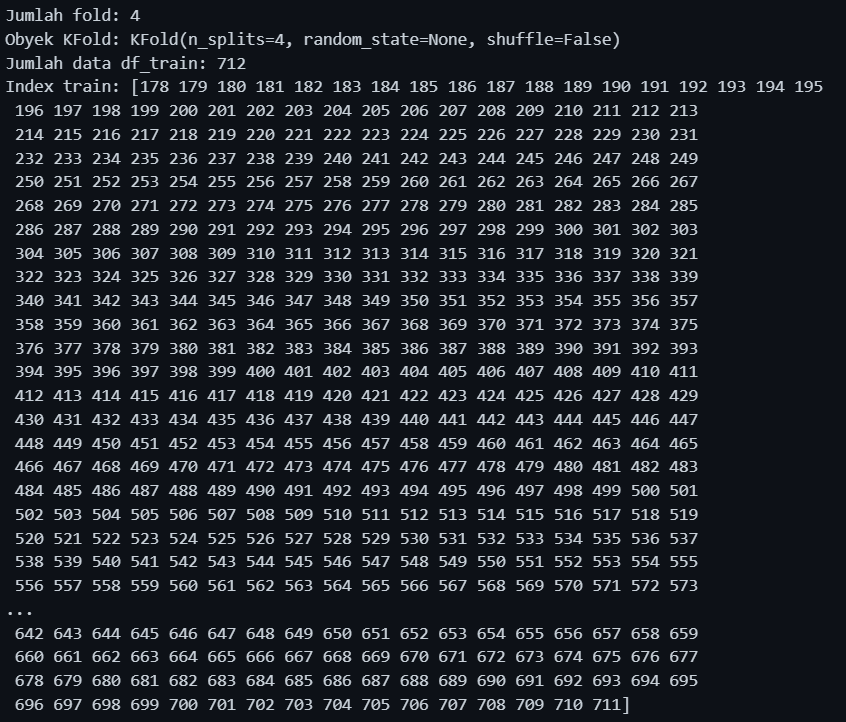
### Langkah 1 - Load Data





### Langkah 2 - Split Data





# Praktikum 4

Ekstraksi Fitur Data Tidak Terstruktur

Pada pratikum ini, kita akan mencoba untuk mengekstraksi fitur dari sebuah citra. Proses yang akan kita lakukan hanya proses sederhana. Pembahasan lengkap terkait dengan pra pengolahan data dan ekstraksi fitur pada data citra akan dibahas pada perkuliahan pengolahan citra digital.

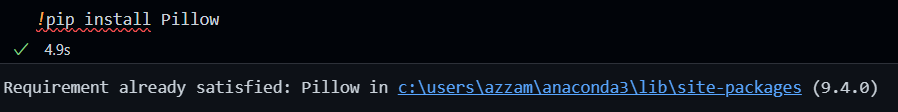
Data yang akan kita gunakan adalah citra berikut.



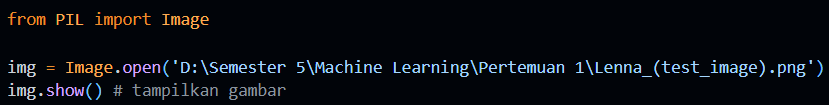
## Langkah 0 - Instal Library

Pada praktikum ini kita akan menggunakan library pillow. Tambahan pillow terlebih dahulu apabila belum terinstal didalam perangkat yang Anda gunakan.

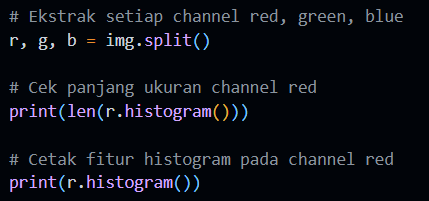
Jalan perintah berikut pada notebook



## Langkah 1 - Load Image



## Langkah 2 - Ekstrak Fitur



Anda akan mendapatkan nilai histogram dalam bentuk array



TUGAS

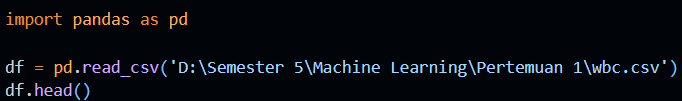
## Deskripsi Tugas

Pada tugas pratikum ini Anda akan menggunakan data "Wisconsin Breast Cancer". Data tersebut terdiri dari 569 data yang digunakan untuk mendiagnonis jenis kanker **Malignant (M) dan Benign (B)**. Tugas Anda adalah,

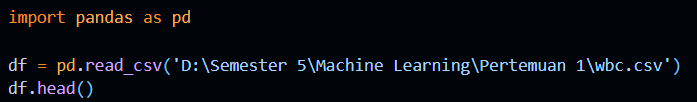
1. Pisahkan antara variabel yang dapat digunakan dan variabel yang tidak dapat digunakan.



sekarang kita load data dari wbc



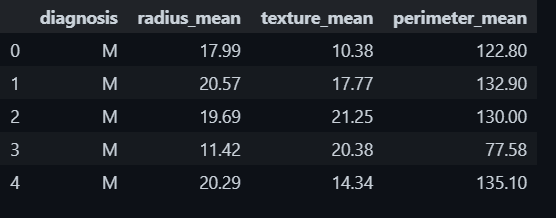
sekarang kita akan memisahkan variabel yang dibutuhkan dan tidak





sekarang kita akan memisahkan variabel yang dibutuhkan dan tidak





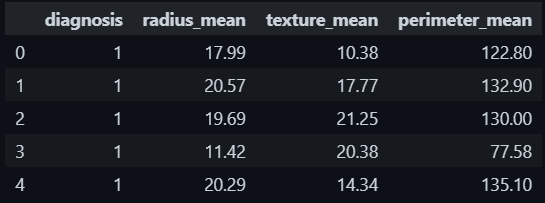
1. Lakukan proses encoding pada kolom "diagnosis".

pertama kita akan melakukan encoding pada kolom "diagnosis" dengan code berikut



selanjutnya kita akan memastikan apakah sudah ter encoding dengan baik

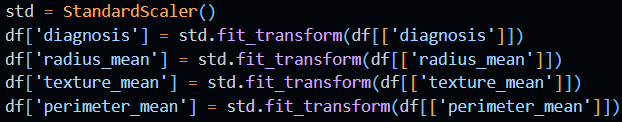




baik sudah ter encoding dengan baik

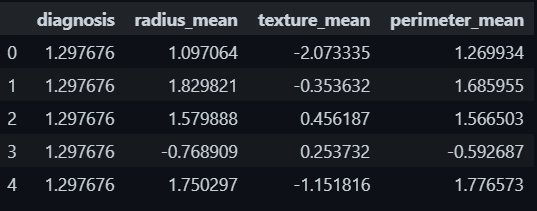
1. Lakukan proses standarisasi pada semua kolom yang memiliki nilai numerik.

pada awalan kita akan melakukan standarisasi pada sejumlah data yang dimana terdiri dari ("diagnosis", "radius\_mean", "texture\_mean", "perimeter\_mean")



standarisasi berhasil di lakukan, maka sekarang kita akan mengecek datanya



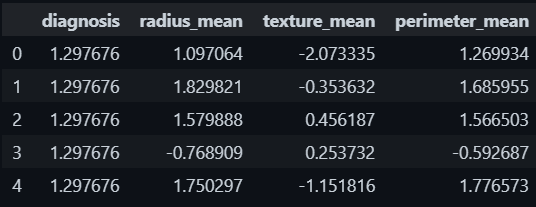


baik data sudah terstandarisasi dengan baik

1. Lakukan proses stratified split data untuk membuat data latih dan data uji dengan rasio 80:20.

pertama kita akan melihat data yang ada





okee selanjutnya kita akan split data ini menggunakan metode stratified split data

