

🏠

HOME  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/)

💰

PEMBAYARAN  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/PEMBAYARAN)

👤

BIODATA DIRI  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA)

📖

PEMINJAMAN BUKU  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PEMINJAMAN)

📄

PENULISAN ILMIAH  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/PENULISAN)

📎

FOTOCOPY  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/FOTOCOPY)

📚

SUMBANG BUKU  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/SUMBANGBUKU)

📁

KASUS & NOTIFIKASI  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/NOTIFIKASI)

👥

BEBAS PERPUSTAKAAN  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN)

🕒

KUNJUNGAN  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/KUNJUNGAN)

📄

E-PAPER  
(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC  
SYSTEM/E-PAPER)

FILKOM/KA/SKRIPSI/5105/2022

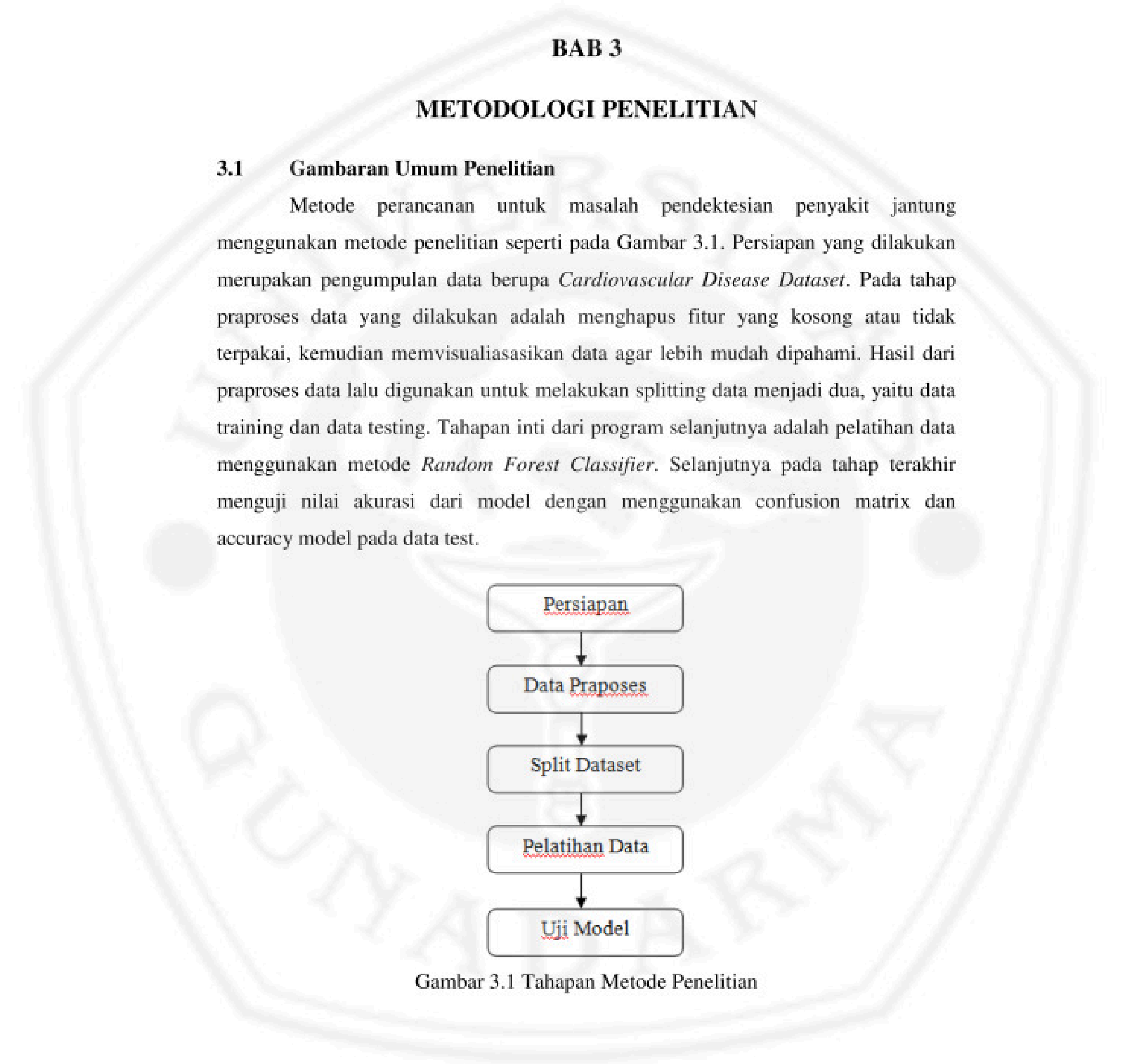
IMPLEMENTASI MACHINE LEARNING DENGAN RANDOM FOREST CLASSIFIER UNTUK DIAGNOSIS CARDIO VASCULAR

NANDA OCTAVIA | 15116311 | 2022 | S1 - SKRIPSI | Sistem Informasi - Ilmu Komputer dan Teknologi Informasi

Pembimbing: Dr. Yuli Karyanti, S.kom., MMSI | Promotor:

Table Of Contents

ENGANTAR	DAFTAR ISI	DAFTAR TABEL	DAFTAR GAMBAR	DAFTAR RUMUS	DAFTAR LAMPIRAN	BAB I	BAB II	BAB III	BAB IV	BAB V	DAFTAR PUSTAKA	JURNAL
----------	------------	--------------	---------------	--------------	-----------------	-------	--------	---------	--------	-------	----------------	--------



### 3.2 Persiapan

Persiapan sangat diperlukan dalam pembuatan program terutama dalam pengumpulan data, agar saat program dijalankan dapat bekerja dengan baik.

Data penelitian yang digunakan ini bernama *Cardiovascular Disease Dataset* dibuat oleh Svetlana Ulianova yang diambil dari <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>. Pada deskripsi data terdapat 3 jenis fitur *input*:

- *Objective* : informasi faktual.
- *Examination* : hasil pemeriksaan kesehatan.
- *Subjective* : informasi yang diberikan oleh pasien.

Beberapa fitur yang termasuk dalam dataset tersebut dapat dilihat pada tabel 3.1.

Tabel 3.1 Cardiovascular Disease Dataset

Age	<i>Objective</i>	Age	Int (days)
Height	<i>Objective</i>	Height	Int (cm)
Weight	<i>Objective</i>	Weight	Float (kg)
Gender	<i>Objective</i>	Gender	Categorical code (1) Women, (2) Men
Systolic blood pressure	<i>Examination</i>	Ap_hi	Int
Diastolic blood pressure	<i>Examination</i>	Ap_lo	Int
Cholesterol	<i>Examination</i>	Cholesterol	(1) Normal, (2) Above normal, (3) Well above normal
Glucose	<i>Examination</i>	Gluc	(1) Normal, (2) Above normal, (3) Well above normal

Smoking	Subjective	Smoke	Binary
Alcohol intake	Subjective	Alco	Binary
Physical activity	Subjective	Active	Binary
Presence or absence of cardiovascular disease	Target Variable	Cardio	Binary

### 3.3 Data Praproses

#### 3.3.1 Import Library

Library adalah koleksi dari rutin-rutin program yang digunakan untuk membangun dan mengembangkan perangkat lunak. Library, umumnya mengandung kode program dan data pembantu (helper), yang menyediakan layanan-layanan kepada program-program independen.

- Import Numpy

```
import numpy as np
```

Gambar 3.3 Import Numpy

Numpy atau kependekan dari *Numerical Python* adalah *library Python* yang fokus pada *scientific computing*. Numpy memiliki kemampuan membentuk objek N-dimensional *array*, yang mirip dengan list pada *Python*. Numpy memudahkan programmer pada Aljabar Linear, terutama operasi pada *Vector* (1-d array) dan *Matrix* (2-d array). Lakukan *import* terlebih dahulu numpy as np. Penggunaan ini berguna menggantikan pemanggilan numpy dengan *prefix* np untuk proses berikutnya.

- Import Pandas

```
import pandas as pd
```

Gambar 3.4 Import Pandas

MENU NAVIGASI	
HOME	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/)
PEMBAYARAN	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/PEMBAYARAN)
BIODATA DIRI	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA)
PEMINJAMAN BUKU	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/PEMINJAMAN)
PENULISAN ILMIAH	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/PENULISAN)
FOTOCOPY	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/FOTOCOPY)
SUMBANG BUKU	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/SUMBANGBUKU)
KASUS & NOTIFIKASI	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/NOTIFIKASI)
BEBAS PERPUSTAKAAN	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN)
KUNJUNGAN	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/KUNJUNGAN)
E-PAPER	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/E-PAPER)

Pandas kependekan dari *Python for Data Analysis*, nama pandas tersebut adalah turunan dari kata *Panel Data* (Wes McKinney, 2011). Pandas merupakan sebuah *open source python package/library* untuk kebutuhan data analisis, manipulasi dan pembersihan data. pandas mendukung pembacaan dan penulisan data dengan media berupa *excel spreadsheet*, CSV, dan SQL yang kemudian akan dijadikan sebagai objek Python dengan *rows* dan *columns* yang disebut *data frame* seperti halnya pada table statistik. Disini Pandas didefinisikan sebagai *pd* sehingga pemanggilan *library* tersebut dapat dilakukan dengan memanggil *pd* lalu di ikuti perintahnya.

- Import Seaborn

```
import seaborn as sns
```

Gambar 3.5 Import Seaborn

Seaborn berfungsi untuk melakukan *visualization*, yang memungkinkan untuk memvisualisasikan data 2D menjadi lebih bagus dan indah. Disini Seaborn didefinisikan sebagai *sns* sehingga pemanggilan *library* dapat dilakukan *sns* kemudian diikuti perintahnya.

3.3.2 Import Dataset

Dataset adalah objek yang merepresentasikann data dan relasinya di *memory*. Strukturnya mirip dengan data di *database*. Dataset berisi koleksi dari *datatable* dan *datarelation*. Pada peneltian ini, digunakan data *Cardiovascular Disease Dataset* yang telah di unduh dari [www.kaagle.com](http://www.kaagle.com).

```
from google.colab import files
```

Gambar 3.6 Import Dataset (1)

MENU NAVIGASI	
HOME	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/)
PEMBAYARAN	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/PEMBAYARAN)
BIODATA DIRI	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA)
PEMINJAMAN BUKU	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/PEMINJAMAN)
PENULISAN ILMIAH	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/PENULISAN)
FOTOCOPY	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/FOTOCOPY)
SUMBANG BUKU	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/SUMBANGBUKU)
KASUS & NOTIFIKASI	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/NOTIFIKASI)
BEBAS PERPUSTAKAAN	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/ANGGOTA/BEBASPERPUSTAKAAN)
KUNJUNGAN	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/KUNJUNGAN)
E-PAPER	(HTTPS://LIBRARY.GUNADARMA.AC.ID/DEPC SYSTEM/E-PAPER)

Fungsi diatas untuk me-load data yang akan dipakai, karena menggunakan software dari Google maka harus menggunakan *Google Library* dari komputer.

```
uploaded = files.upload()  
df = pd.read_csv("cardio.csv", sep=";")
```

Gambar 3.7 Import Dataset (2)

Fungsi upload selanjutnya akan menampilkan bar pencarian atau pilihan sehingga secara manual programmer memilih dataset yang akan dipakai nantinya. Fungsi dari df adalah data frame, lalu diikuti dengan inisial pandas untuk membaca dataset. Dataset yang di upload harus bernama ‘cardio.csv’, selain itu program tidak akan membaca dataset.

```
df.head(7)
```

Gambar 3.8 Import Dataset (3)

Setelah mengupload dataset pada *data frame*, fungsi berikutnya merupakan mencetak 7 baris pertama yang terdapat pada dataset dengan menginput *value* didalamnya berisi 7. Lalu, *run program* dengan *shortcut* CTRL+F9.

Choose Files cardio.csv

- cardio.csv(application/vnd.ms-excel) - 2941524 bytes, last modified: 10/17/2019 - 100% done

Saving cardio.csv to cardio.csv

Gambar 3.9 Import Dataset (4)

Setelah di *run*, akan muncul pilihan *choose file* yang mana akan meminta lokasi dataset yang pada komputer dengan nama data.csv lalu klik *open*.

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0
5	8	21914	1	151	67.0	120	80	2	2	0	0	0	0
6	9	22113	1	157	93.0	130	80	3	1	0	0	1	0

Gambar 3.10 Import Dataset (5)

Dataset yang telah selesai 100% loading data akan terlihat mencetak 7 baris pertama seperti perintah yang sebelumnya ditulis. Pada cardio.csv yang telah diupload terdapat atribut ID yang mendefinisikan identitas dari pasien.

```
#Get the shape of the data (the number of rows & columns)
df.shape

(70000, 13)
```

Gambar 3.11 Import Dataset (6)

Selanjutnya, menghitung jumlah baris dan kolom yang ada pada dataset yang sudah di *input* dengan df.shape lalu *run program*. Hasil yang ditampilkan adalah 70000 baris dimana setiap barisnya merepresentasikan data pasien, serta terdapat 13 baris yang merupakan *features* dari setiap pasien.

3.3.3 Remove Missing Values

Dataset yang telah di *upload* sebelumnya di cek apakah terdapat kolom yang bernilai atau nol, jika ada maka kolom ini tidak akan terpakai dan akan dihapus. Untuk mengetahui kolom yang kosong atau tidak bernilai digunakan fungsi seperti gambar 3.12 agar mencetak semua fitur yang ada pada dataset.

```
df.isna().sum()
```

Gambar 3.12 Remove Missing Value (1)

