

LAPORAN
PRAKTIKUM MACHINE LEARNING



Disusun oleh:

Michael Yeremia P. Sihombing

DEPARTEMEN INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG
2019

BAB I

PEMBAHASAN

1. Data yang digunakan

Pada praktikum kali ini data yang digunakan di ambil dari web <https://archive.ics.uci.edu>.

Dalam web ini saya mengambil contoh data haberman dengan link data

<https://archive.ics.uci.edu/ml/machine-learning-databases/haberman/haberman.data>. Data

ini memiliki beberapa atribut/fitur didalamnya seperti penjelasan di bawah ini.

Number of Instances: 306

Number of Attributes: 4 (including the class attribute)

Attribute Information:

1. Age of patient at time of operation (numerical)
2. Patient's year of operation (year - 1900, numerical)
3. Number of positive axillary nodes detected (numerical)
4. Survival status (class attribute)
 - 1 = the patient survived 5 years or longer
 - 2 = the patient died within 5 year

2. Pembahasan

```
C:\WINDOWS\system32\cmd.exe - python
>>> url = "https://archive.ics.uci.edu/ml/machine-learning-databases/haberman/haberman.data"
>>> names = ['age', 'patient-operation-year=19xx', 'positive axillary nodes detected', 'class']
>>> dataset = pandas.read_csv(url, names=names)
>>> print(dataset.shape())
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: 'tuple' object is not callable
>>> print(dataset.shape)
(306, 4)
>>>
```

Gambar di atas merupakan langkah-langkah awal yang dilakukan untuk menjalankan praktikum. Pertama pilih lokasi dimana data berada, selanjutnya buat fields/atribut/fitur yang bersesuaian untuk menampilkan datanya, selanjutnya set data menggunakan fungsi `pandas.read_csv(x,y)`. `Print(dataset.shape)` digunakan untuk menampilkan jumlah rows data dan fitur/atribut/coloums data.

```
>>> print(dataset.head(20))
```

	age	patient-operation-year=19xx	positive axillary nodes detected	class
0	30	64	1	1
1	30	62	3	1
2	30	65	0	1
3	31	59	2	1
4	31	65	4	1
5	33	58	10	1
6	33	60	0	1
7	34	59	0	2
8	34	66	9	2
9	34	58	30	1
10	34	60	1	1
11	34	61	10	1
12	34	67	7	1
13	34	60	0	1
14	35	64	13	1
15	35	63	0	1
16	36	60	1	1
17	36	69	0	1
18	37	60	0	1
19	37	63	0	1

```
>>>
```

Gambar di atas merupakan fungsi yang digunakan untuk menampilkan data. Print(dataset.head(20)) akan menampilkan 20 data pertama yang ada, untuk print keseluruhan data cukup menggunakan perintah print(dataset).

```
>>> print(dataset.groupby('class').size())
```

class	size
1	225
2	81

```
dtype: int64
>>>
```

Gambar di atas merupakan fungsi untuk melihat sebaran class yang terdapat di dalam data. Fungsi yang digunakan yaitu groupby().size().

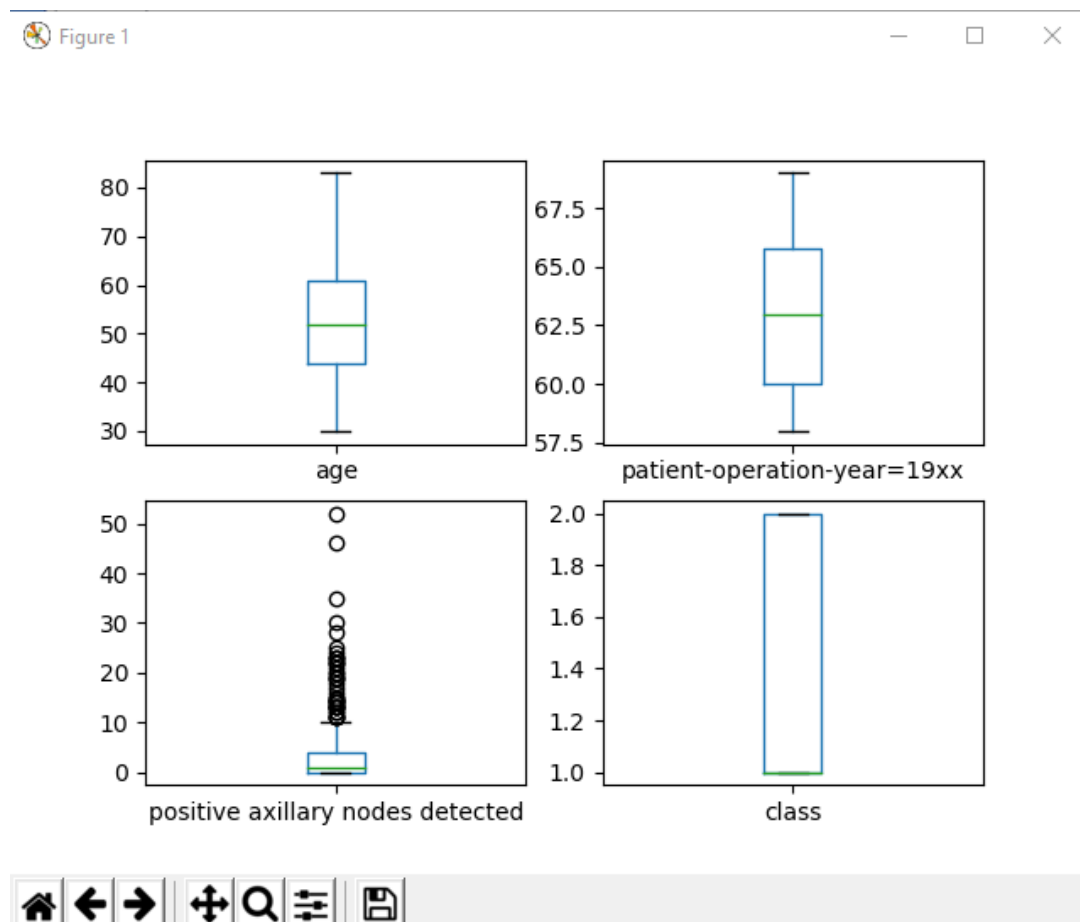
```
>>> print(dataset.describe())
```

	age	patient-operation-year=19xx	positive axillary nodes detected	class
count	306.000000	306.000000	306.000000	306.000000
mean	52.457516	62.852941	4.026144	1.264706
std	10.803452	3.249405	7.189654	0.441899
min	30.000000	58.000000	0.000000	1.000000
25%	44.000000	60.000000	0.000000	1.000000
50%	52.000000	63.000000	1.000000	1.000000
75%	60.750000	65.750000	4.000000	2.000000
max	83.000000	69.000000	52.000000	2.000000

Gambar di atas merupakan fungsi untuk menghitung statistika dari data tersebut. Fungsi tersebut dapat menghitung rata-rata, std deviasi, min, max, dan kuartil.

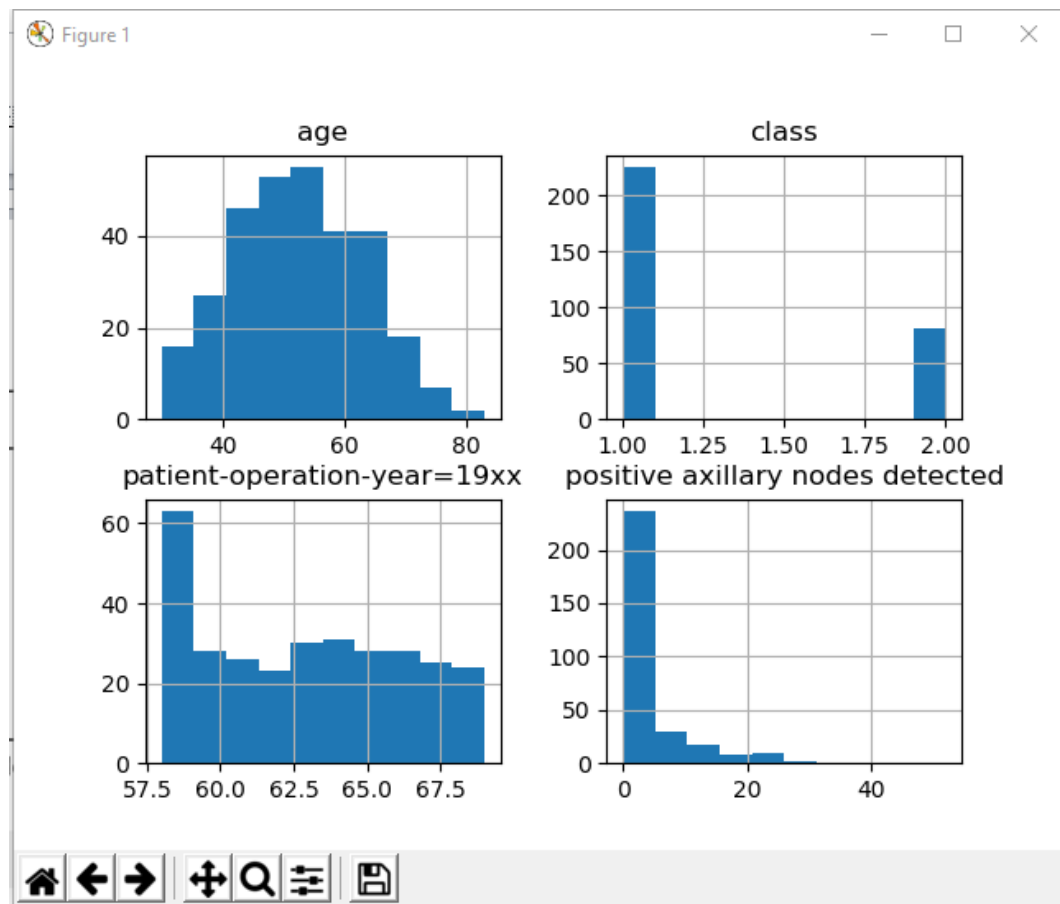
```
>>> dataset.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
age                               AxesSubplot(0.125,0.53;0.352273x0.35)
patient-operation-year=19xx       AxesSubplot(0.547727,0.53;0.352273x0.35)
positive axillary nodes detected  AxesSubplot(0.125,0.11;0.352273x0.35)
class                             AxesSubplot(0.547727,0.11;0.352273x0.35)
dtype: object
>>> plt.show()
```

Fungsi di atas merupakan fungsi yang digunakan untuk membuat plot univariant. Plot yang dibuat disini berbentuk box dikarenakan data-datanya berupa numeric. Layout yang ditampilkan haruslah sejumlah dengan atribut/fitur yang ada, misal 4 atribut layout(2,2). Di bawah ini merupakan gambar hasil plt.show() dari plot yang dibuat.



```
>>> dataset.hist()
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x0E9681D0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x014747D0>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x014968D0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x0FAF49D0>]],
      dtype=object)
>>> plt.show()
```

Gambar di atas merupakan fungsi yang digunakan untuk menampilkan sebaran data berdasarkan atribut/fitur dengan menggunakan histogram. Gambar di bawah merupakan hasil yang didapat dari menjalankan fungsi tersebut.



```

>>> import pandas
>>> from pandas.plotting import scatter_matrix
>>> scatter_matrix(dataset)
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x101FFEF0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x10226A90>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x10243B90>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x10263C90>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x10285D90>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x102A4E90>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x102A4F30>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x10616B10>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x108410D0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x10863290>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x10898DF0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x104718B0>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x10475DF0>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x1045D190>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x10490C30>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x106D0050>]],
       dtype=object)
>>> plt.show()

```

Gambar di atas merupakan fungsi untuk melihat interaksi antara variabel. Untuk membantu kita melihat interaksi antar variable kita menggunakan fungsi scatterplots seperti gambar di atas. Gambar di bawah merupakan hasil dari fungsi yang dijalankan.

