

# **MAKALAH PEMBELAJARAN MESIN**

**“Titanic”**



**Pengampu :**

**Dr.Juni Nurma Sari S.Kom, M.MT**

**Nama:**

Wanda Trisnahayu (2055301143)

**Kelas : 3 TI C**

**D4 - TEKNIK INFORMATIKA**

**TAHUN 2022**

## 1. Import tools/library yg dibutuhkan

```
✓ [1] import pandas as pd
0s import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import math
```

## 2. Load data

```
⦿ # load data, pastikan membaca note
data = pd.read_csv('titanic.csv')
data.head(10)
```

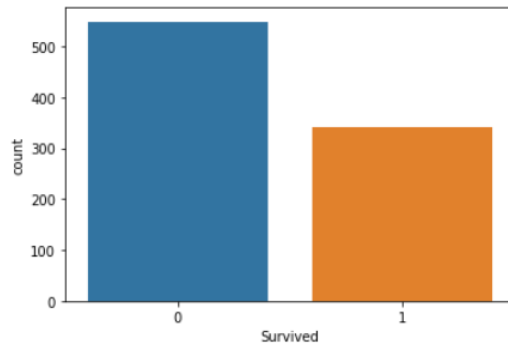
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C

### 3. Analisa data

```
[4] # Analisa data
# countplot
sns.countplot("Survived", data=data)
# 1 = selamat
```

/usr/local/lib/python3.7/dist-packages/seaborn/\_decorators.py:43: FutureWarning: Pass the following variable as FutureWarning

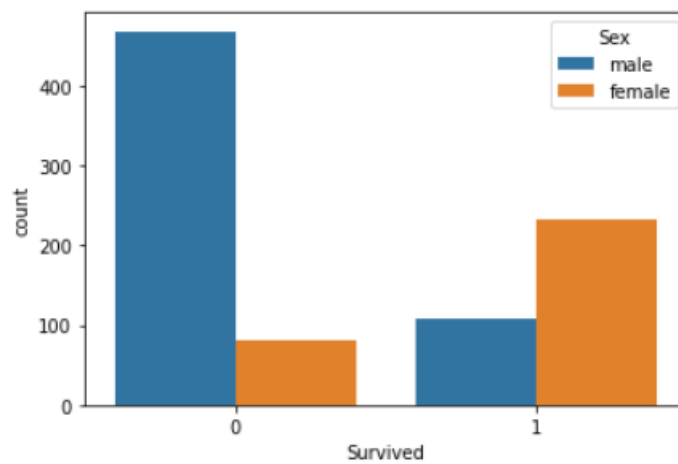
<matplotlib.axes.\_subplots.AxesSubplot at 0x7fd4140de0d0>



### 4. Countplot

```
[5] sns.countplot(x="Survived", hue="Sex", data=data)
# selamat banyak female
```

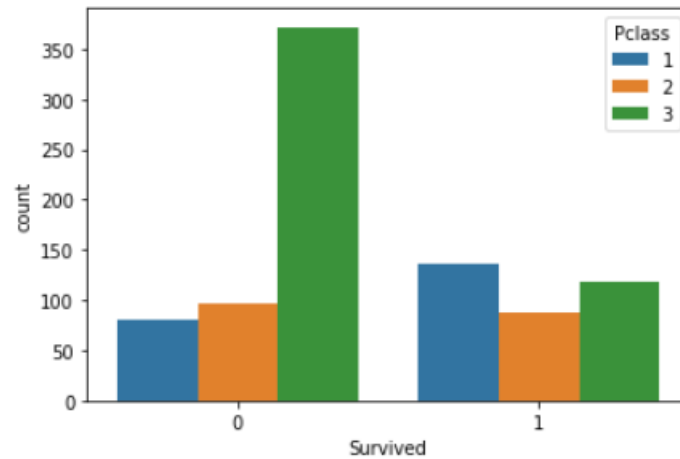
<matplotlib.axes.\_subplots.AxesSubplot at 0x7fd3fbbfa410>



## 5. Sns.countplot

```
[6] sns.countplot(x="Survived", hue="Pclass", data=data)  
# class 3 banyak tidak selamat
```

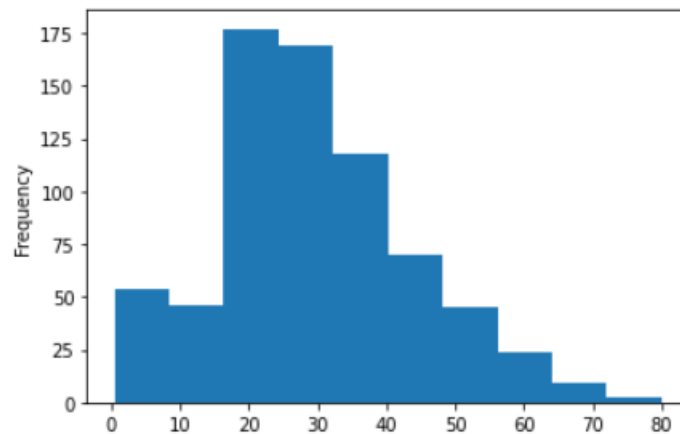
<matplotlib.axes.\_subplots.AxesSubplot at 0x7fd3fb704ad0>



## 6. Plot data yg berumur 20-30

```
data['Age'].plot.hist()  
# banyak yang berumur 20 - 30
```

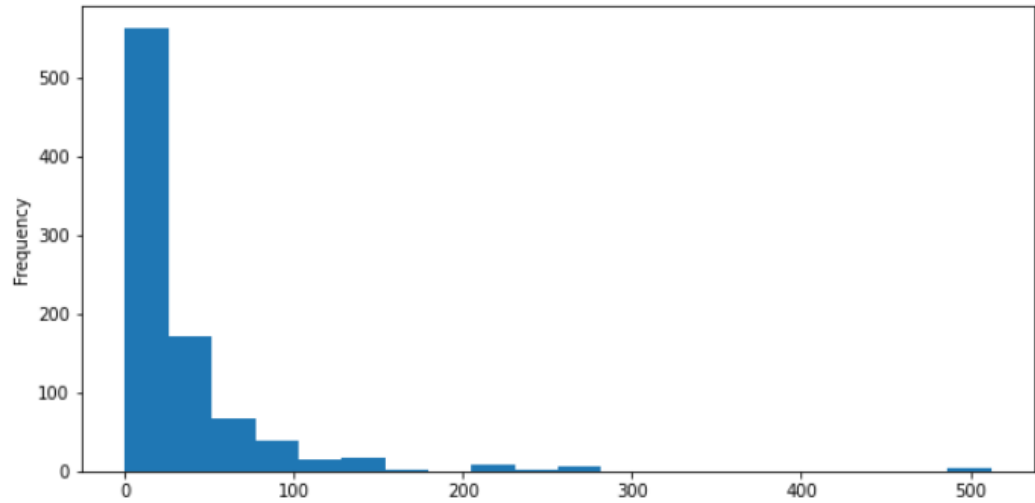
<matplotlib.axes.\_subplots.AxesSubplot at 0x7fd3fb612150>



## 7. Tarif (dollar)

```
[8] data['Fare'].plot.hist(bins=20, figsize=(10,5))  
# tarif (dolar)
```

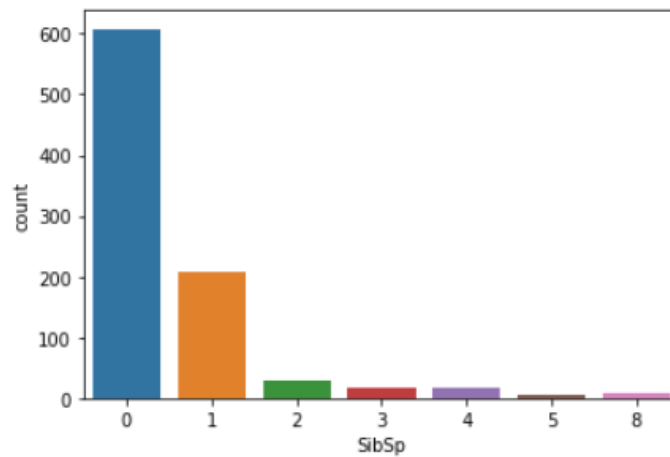
<matplotlib.axes.\_subplots.AxesSubplot at 0x7fd3fb5b1b10>





## 8. Sns.countplot


```
[10] # sibling/spouse  
sns.countplot(x='SibSp', data=data)  
# banyak yang tidak
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fd3fb449590>



## 9. Cleaning data

  `# Data Wrangling/Cleaning Data`  
`# cek data kosong`  
`data.isnull()`



	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	True	False
...	...	...	...	...	...	...	...	...	...	...	...	...
886	False	False	False	False	False	False	False	False	False	False	True	False
887	False	False	False	False	False	False	False	False	False	False	False	False
888	False	False	False	False	False	True	False	False	False	False	True	False
889	False	False	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	False	True	False

891 rows × 12 columns


## 10. Data.isnull

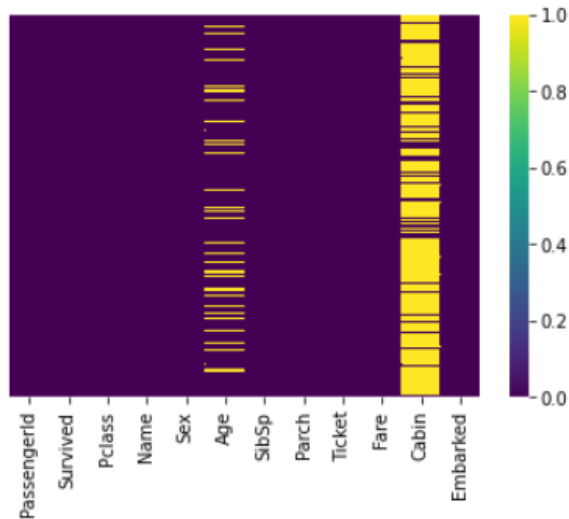
```
[12] data.isnull().sum()
```

```
PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age            177
SibSp           0
Parch           0
Ticket          0
Fare            0
Cabin          687
Embarked        2
dtype: int64
```

## 11. plot dengan heatmap

```
✓ 0s # plot dengan heatmap  
sns.heatmap(data.isnull(), yticklabels=False, cmap='viridis')
```

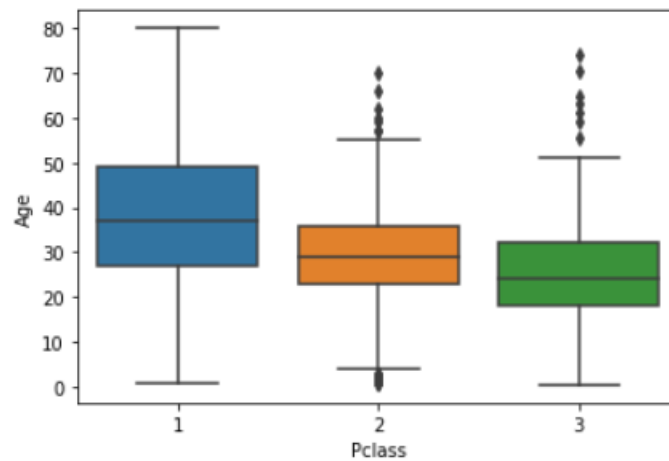
 <matplotlib.axes.\_subplots.AxesSubplot at 0x7fd3fb3f6710>



## 12. box plot

```
[14] # box plot  
sns.boxplot(x='Pclass', y='Age', data=data)
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fd3f8b07f90>



### 13. imputation

```
[15] #imputation
# cek data
data.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S



### 14. drop cabin

```
[16] # drop Cabin karena banyak yg kosong
data.drop('Cabin', axis=1, inplace=True)
```

### 15. imputasi age

```
[20] # fungsi untuk imputasi fitur Age
def impute_age(cols):
    Age=cols[0]
    Pclass=cols[1]
    if(pd.isnull(Age)):
        if(Pclass==1):
            return 37
        elif(Pclass==2):
            return 29
        else:
            return 24
```



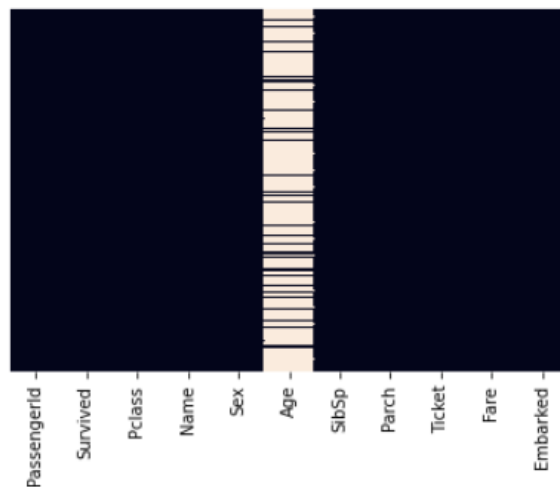
16. jalankan fungsi

```
✓ [21] # jalankan fungsi  
Is data['Age'] = data[['Age', 'Pclass']].apply(impute_age, axis=1)
```

17. cek lagi

```
✓ [22] #cek lagi  
Ds sns.heatmap(data.isnull(), yticklabels=False, cbar=False)
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fd3f8a79710>



18. konversi value ke kategorikal agar bisa diproses

```
✓ [23] # value masih aneh, dikonversi ke kategorikal supaya bisa diproses  
# logistik regression  
data.head(2)  
# menggunakan strategi one hot encoding
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	NaN	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	NaN	1	0	PC 17599	71.2833	C

## 19. Pemecahan fitur

```
# pemecahan fitur  
pd.get_dummies(data['Sex'])
```

	female	male
0	0	1
1	1	0
2	1	0
3	1	0
4	0	1
...	...	...
886	0	1
887	1	0
888	1	0
889	0	1
890	0	1

891 rows × 2 columns

## 20. Pemecahan fitur untuk ambil male saja

```
[25] # pemecahan fitur, ambil male saja  
jenkel = pd.get_dummies(data['Sex'], drop_first=True)  
jenkel.head()
```

	male
0	1
1	0
2	0
3	0
4	1

## 21. Pecah embark

```
[26] # pecah embark  
emb = pd.get_dummies(data['Embarked'])  
emb.head()
```

	c	q	s
0	0	0	1
1	1	0	0
2	0	0	1
3	0	0	1
4	0	0	1

## 22. Pecah embark menghilangkan C

```
# pecah embark, hilangkan C  
emb = pd.get_dummies(data['Embarked'], drop_first=True)  
emb.head()
```

	q	s
0	0	1
1	0	0
2	0	1
3	0	1
4	0	1

## 23. Memisahkan PCL dan hilangkan class 1

```
[28] # Pcl perlu dipisah dan dihilangkan class 1
Pcl = pd.get_dummies(data['Pclass'], drop_first=True)
Pcl.head()
```

	2	3
0	0	1
1	0	0
2	0	1
3	0	0
4	0	1

## 24. Gabungkan seluruh

```
# gabungkan semuanya
data = pd.concat([data, jenkel, emb, Pcl], axis=1)
data.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	male	Q	S	2	3
0	1	0	3	Braund, Mr. Owen Harris	male	NaN	1	0	A/5 21171	7.2500	S	1	0	1	0	1
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	NaN	1	0	PC 17599	71.2833	C	0	0	0	0	0
2	3	1	3	Heikkinen, Miss. Laina	female	NaN	0	0	STON/O2. 3101282	7.9250	S	0	0	1	0	1
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	NaN	1	0	113803	53.1000	S	0	0	1	0	0
4	5	0	3	Allen, Mr. William Henry	male	NaN	0	0	373450	8.0500	S	1	0	1	0	1

## 25. Drop pclass

```
[30] # drop Pclass, sex, dan embarked,, selain hapus PassengerId, Name, ticket
# karena tidak diolah
data.drop(['Sex','Embarked','Pclass', 'PassengerId','Name','Ticket'], axis=1,
inplace=True)
data.head()
```

	Survived	Age	SibSp	Parch	Fare	male	Q	S	2	3
0	0	NaN	1	0	7.2500	1	0	1	0	1
1	1	NaN	1	0	71.2833	0	0	0	0	0
2	1	NaN	0	0	7.9250	0	0	1	0	1
3	1	NaN	1	0	53.1000	0	0	1	0	0
4	0	NaN	0	0	8.0500	1	0	1	0	1

## 26. Train data

```
✓ 0s # Train Data
# tentukan dependen dan independen var
X= data.drop('Survived', axis=1)
y= data['Survived']
```

## 27. Library split data

```
✓ 0s [32] # library untuk split data
from sklearn.model_selection import train_test_split
```

## 28. Split ke beberapa bagian

```
✓ [33] # split data ke beberapa bagian  
0s X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,  
random_state=1)
```

## 29. Buat model

```
✓ [35] # buat model  
0s from sklearn.linear_model import LogisticRegression
```

```
✓ [36] # buat model  
0s logModel = LogisticRegression()
```