Nama : Aldi Pradana

Fakultas : STMIK Methodist Binjai

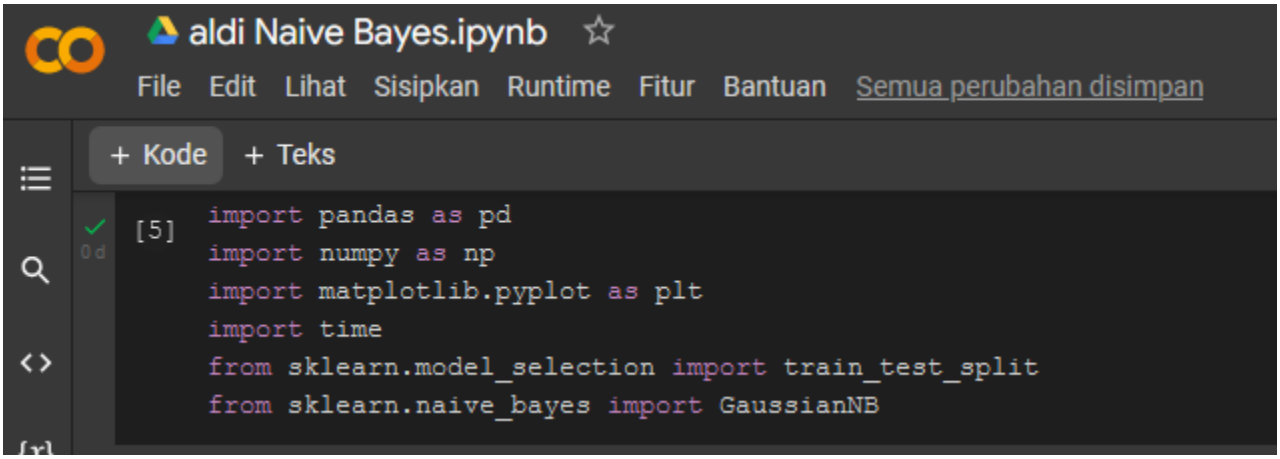
Tugas : Quiz

Date : 8 Maret 2022

Quiz 1 Minggu ke-3

Jawaban :

- Import modul terlebih dulu



- Upload file titanic ke google colab

- Jika sudah maka tuliskan variabel upload seperti berikut

```
uploaded
```

```
{'Titanic.csv': b'Name,PClass,Age,Sex,Survived\r\n"Allen, Miss Elisabeth Walton",1st,29,female,1\r\n"Allison, Miss Helen Loraine",1st,2,female,0\r\n"Allisc
```

- Import dataset untuk menampilkan tampilan dari file titanic tadi

```
# Importing dataset
data = pd.read_csv("Titanic.csv")

# Convert categorical variable to numeric
data["Sex_cleaned"]=np.where(data["Sex"]=="male",0,1)
```

- Output data titanic csv

```
data
```

	Name	PClass	Age	Sex	Survived	Sex_cleaned
0	Allen, Miss Elisabeth Walton	1st	29.00	female	1	1
1	Allison, Miss Helen Loraine	1st	2.00	female	0	1
2	Allison, Mr Hudson Joshua Creighton	1st	30.00	male	0	0
3	Allison, Mrs Hudson JC (Bessie Waldo Daniels)	1st	25.00	female	0	1
4	Allison, Master Hudson Trevor	1st	0.92	male	1	0
...	...	...	...	...	...	...
1308	Zakarian, Mr Artun	3rd	27.00	male	0	0
1309	Zakarian, Mr Maprieder	3rd	26.00	male	0	0
1310	Zenni, Mr Philip	3rd	22.00	male	0	0
1311	Lievens, Mr Rene	3rd	24.00	male	0	0
1312	Zimmerman, Leo	3rd	29.00	male	0	0

1313 rows × 6 columns

- Cleaning data set untuk menampilkan sesuai di list data

```
# Cleaning dataset of NaN
data=data[[
    "Survived",
    "Sex_cleaned",
    "Age"
]].dropna(axis=0, how='any')
```

- Nah disinii age nya masih berantakan

```
data
```

	Survived	Sex_cleaned	Age
0	1	1	0.407031
1	0	1	0.025837
2	0	0	0.421149
3	0	1	0.350558
4	1	0	0.010589
...	...	...	...
1308	0	0	0.378794
1309	0	0	0.364676
1310	0	0	0.308203
1311	0	0	0.336439
1312	0	0	0.407031

756 rows × 3 columns

- Disini saya coba pakai sklearn dan import MinMaxscaler untuk merapihkan dan menyesuaikan age nya

```
[18] from sklearn.preprocessing import MinMaxScaler
     scaler = MinMaxScaler()
     normal=['Age']
     data[normal]=scaler.fit_transform(data[normal])
```

- Maka tampilannya seperti berikut

```
data
```

	Survived	Sex_cleaned	Age
0	1	1	29.00
1	0	1	2.00
2	0	0	30.00
3	0	1	25.00
4	1	0	0.92
...	...	...	...
1308	0	0	27.00
1309	0	0	26.00
1310	0	0	22.00
1311	0	0	24.00
1312	0	0	29.00

756 rows × 3 columns

- Buat list data used features

```
[19] used_features =[
         "Sex_cleaned",
         "Age"
     ]

[20] X = data[used_features].values
     y = data["Survived"]
```

- Import train test disini saya buat train A dan train B

```
[21] from sklearn.model_selection import train_test_split

     trainA, testA, trainB, testB = train_test_split(X, y, test_size=0.5, random_state=32, stratify=y)
```

```
print(f'train a:{len(trainA)}')
print(f'train b:{len(trainB)}')
print(f'test  a:{len(testA)}')
print(f'test  b:{len(testB)}')
```

```
train a:378
train b:378
test  a:378
test  b:378
```

- Import gaussian dan matrix kemudian saya *50

```
[25] from sklearn.naive_bayes import GaussianNB
     gnb=GaussianNB
     gnb = GaussianNB()
     gnb.fit(trainA,trainB)
     y_pred_train=gnb.predict(trainA)
```

```
from sklearn.metrics import accuracy_score,classification_report,confusion_matrix
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
print("Classification Report is:\n", classification_report(trainB,y_pred_train))
print ("Confusion Matrix:\n",confusion_matrix(trainB,y_pred_train))
print("Training Score:\n", gnb.score(trainA,trainB)*50)
```

- Output nya seperti berikut

```
Classification Report is:
              precision    recall  f1-score   support

           0       0.79      0.85      0.82       221
           1       0.76      0.69      0.72       157

    accuracy                           0.78       378
   macro avg       0.78      0.77      0.77       378
weighted avg       0.78      0.78      0.78       378

Confusion Matrix:
 [[187  34]
 [ 49 108]]
Training Score:
 39.021164021164026
```