In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.naive_bayes import MultinomialNB as MB
from sklearn.naive_bayes import GaussianNB as GB
from sklearn.model_selection import train_test_split
```

executed in 58.8s, finished 15:56:29 2022-01-05

In [2]:

```
salary_train=pd.read_csv("SalaryData_Train.csv")
salary_test=pd.read_csv("SalaryData_Test.csv")
```

executed in 235ms, finished 15:56:32 2022-01-05

In [3]:

```
salary_train.head()
```

executed in 68ms, finished 23:32:51 2021-11-26

Out[3]:

	age	workclass	education	educationno	maritalstatus	occupation	relationship	race	s
0	39	State-gov	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Ma
1	50	Self-emp-not-inc	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Ma
2	38	Private	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Ma
3	53	Private	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Ma
4	28	Private	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Fema

In [4]:

```
salary_test.head()
```

executed in 54ms, finished 23:33:04 2021-11-26

Out[4]:

	age	workclass	education	educationno	maritalstatus	occupation	relationship	race	sex
0	25	Private	11th	7	Never-married	Machine-op-inspct	Own-child	Black	Male
1	38	Private	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White	Male
2	28	Local-gov	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	White	Male
3	44	Private	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	Black	Male
4	34	Private	10th	6	Never-married	Other-service	Not-in-family	White	Male

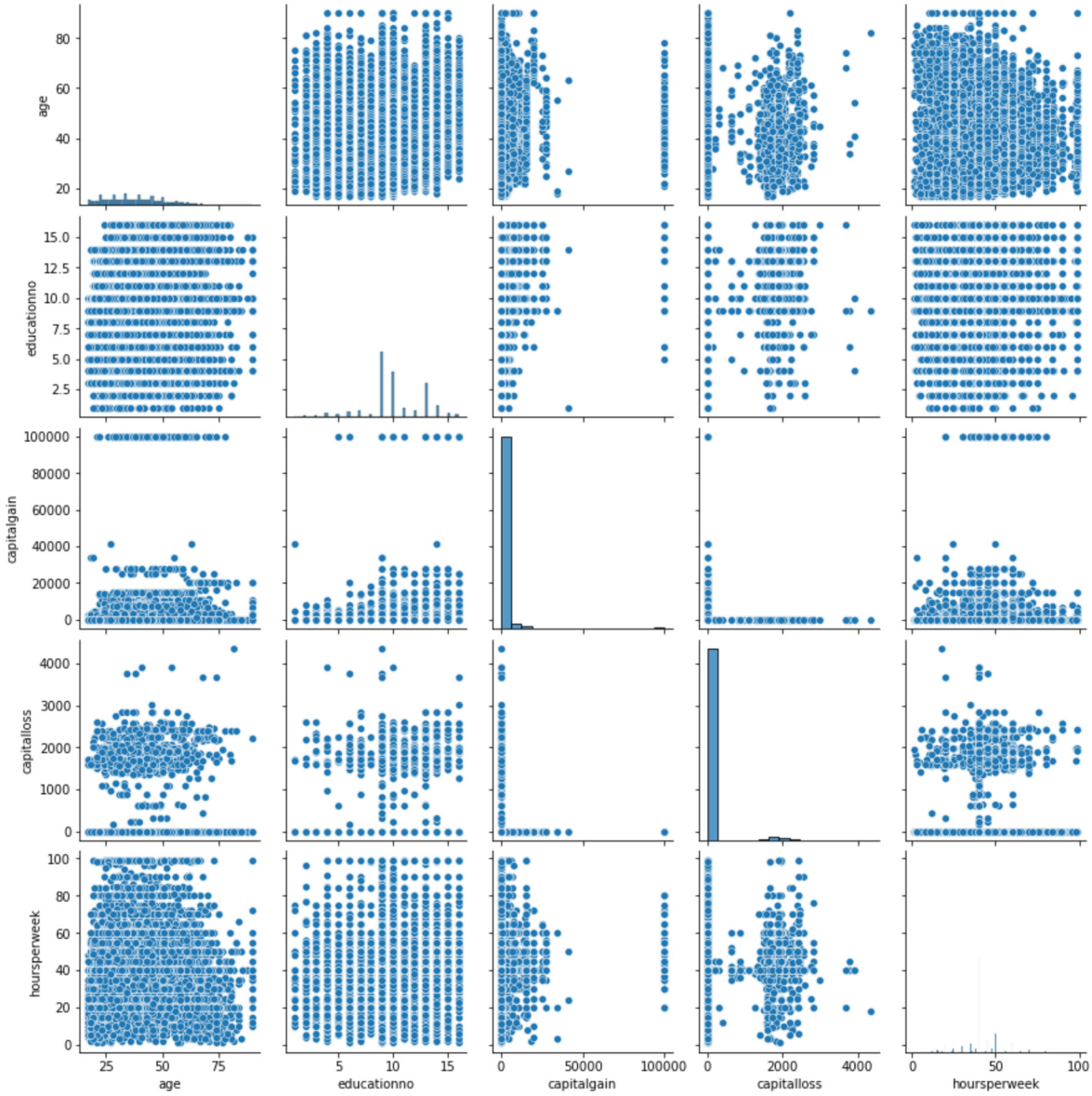
In [5]:

```
sns.pairplot(salary_train)
```

executed in 12.2s, finished 23:34:15 2021-11-26

Out[5]:

```
<seaborn.axisgrid.PairGrid at 0x1b151e72af0>
```



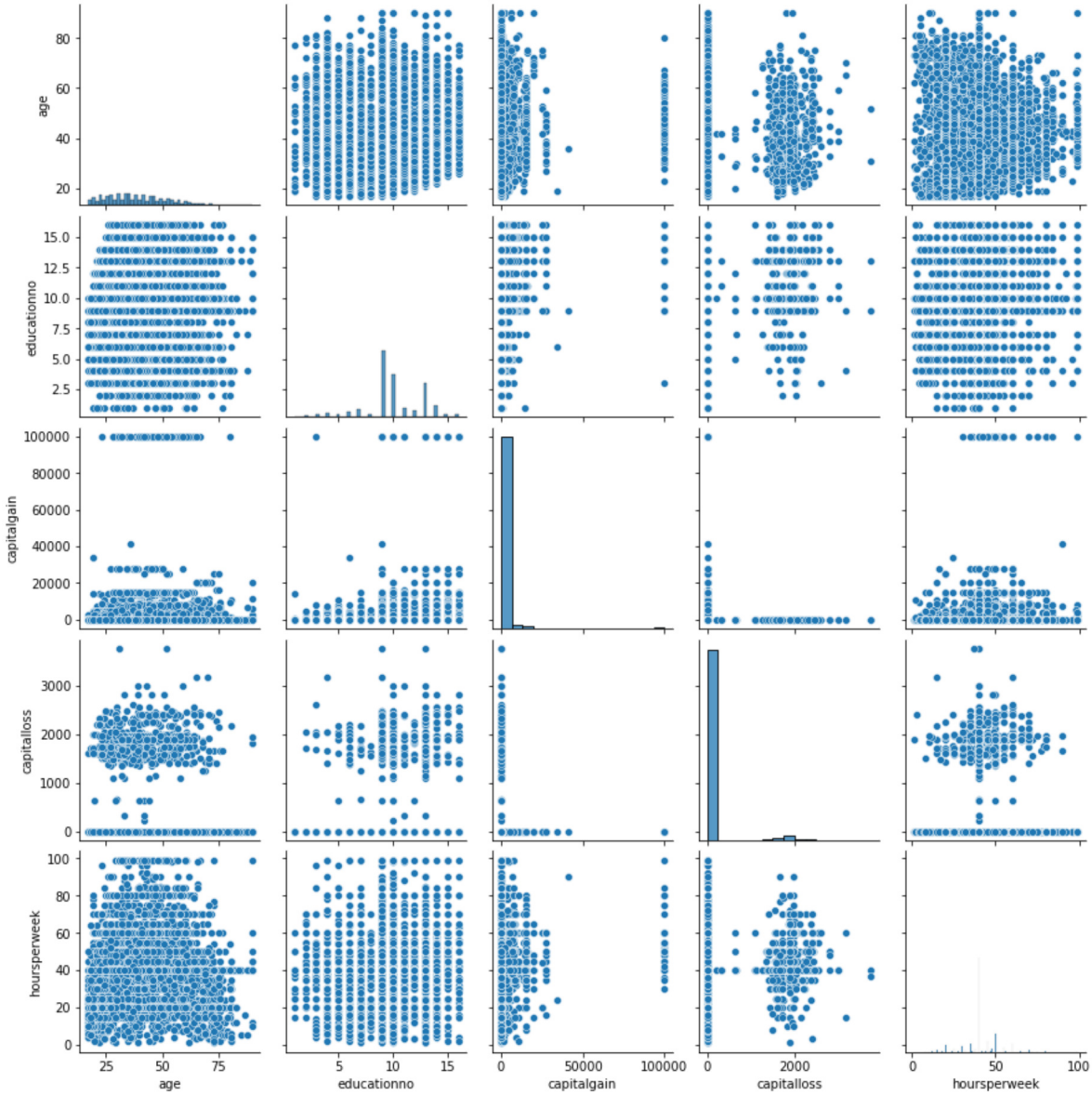
In [6]:

```
sns.pairplot(salary_test)
```

executed in 12.2s, finished 23:34:45 2021-11-26

Out[6]:

```
<seaborn.axisgrid.PairGrid at 0x1b154b83d30>
```



In [7]:

```
string_columns=["workclass","maritalstatus","occupation","relationship","education","race",
```

executed in 10ms, finished 23:35:01 2021-11-26

In [8]:

```
le=LabelEncoder()
```

executed in 14ms, finished 23:35:15 2021-11-26

In [9]:

```
for i in string_columns:
    salary_train[i]=le.fit_transform(salary_train[i])
    salary_test[i]=le.fit_transform(salary_test[i])
```

executed in 222ms, finished 23:35:27 2021-11-26

In [10]:

```
x_train=salary_train.iloc[:,0:12]
y_train=salary_train.iloc[:,13]
x_test=salary_test.iloc[:,0:12]
y_test=salary_test.iloc[:,13]
```

executed in 29ms, finished 23:35:39 2021-11-26

In [11]:

```
classifiers_mb=MB()
classifiers_mb.fit(x_train,y_train)
train_pred_mb=classifiers_mb.predict(x_train)
train_accu_mb=np.mean(train_pred_mb==y_train)
pd.crosstab(train_pred_mb,y_train)
```

executed in 320ms, finished 23:35:51 2021-11-26

Out[11]:

Salary	<=50K	>50K
row_0		
<=50K	21717	5913
>50K	936	1595

In [12]:

```
test_pred_mb=classifiers_mb.predict(x_test)
test_accu_mb=np.mean(test_pred_mb==y_test)
pd.crosstab(test_pred_mb,y_test)
```

executed in 101ms, finished 23:36:02 2021-11-26

Out[12]:

Salary	<=50K	>50K
row_0		
<=50K	10891	2920
>50K	469	780

In [13]:

```
test_pred_mb
```

executed in 18ms, finished 23:36:12 2021-11-26

Out[13]:

```
array([' <=50K', ' <=50K', ' <=50K', ..., ' <=50K', ' >50K', ' <=50K'],
      dtype='<U6')
```

In [14]:

```
test_accu_mb
```

executed in 15ms, finished 23:36:22 2021-11-26

Out[14]:

```
0.7749667994687915
```

In [15]:

```
classifiers_gb=GB()
classifiers_gb.fit(x_train,y_train)
train_pred_gb=classifiers_gb.predict(x_train)
train_accu_gb=np.mean(train_pred_gb==y_train)
pd.crosstab(train_pred_gb,y_train)
```

executed in 187ms, finished 23:36:37 2021-11-26

Out[15]:

Salary	<=50K	>50K
row_0		
<=50K	21506	5040
>50K	1147	2468

In [16]:

```
test_pred_gb=classifiers_gb.predict(x_test)
test_accu_gb=np.mean(test_pred_gb==y_test)
```

executed in 46ms, finished 23:36:52 2021-11-26

In [17]:

```
test_pred_gb
```

executed in 31ms, finished 23:37:04 2021-11-26

Out[17]:

```
array([' <=50K', ' <=50K', ' <=50K', ..., ' <=50K', ' >50K', ' <=50K'],
      dtype='<U6')
```

In [18]:

```
test_accu_gb
```

executed in 16ms, finished 23:37:15 2021-11-26

Out[18]:

```
0.7942895086321381
```