```
import pandas as pd


df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/titanic.csv")
df.head()
```

	PassengerId	Name	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	Braund, Mr. Owen Harris	3	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	Cumings, Mrs. John Bradley (Florence Briggs Th	1	female	38.0	1	0	PC 17599	71.2833	C85

```
df.drop(['PassengerId','Name','SibSp','Parch','Ticket','Cabin','Embarked'],axis='columns',inp
df.head()
```

	Pclass	Sex	Age	Fare	Survived
0	3	male	22.0	7.2500	0
1	1	female	38.0	71.2833	1
2	3	female	26.0	7.9250	1
3	1	female	35.0	53.1000	1
4	3	male	35.0	8.0500	0

Saved successfully!

```
olumns')
target = df.Survived

#inputs.Sex = inputs.Sex.map({'male': 1, 'female': 2})


dummies = pd.get_dummies(inputs.Sex)
dummies.head(3)
```

	female	male
0	0	1
1	1	0
2	1	0

```
inputs = pd.concat([inputs,dummies],axis='columns')
inputs.head(3)
```

	Pclass	Sex	Age	Fare	female	male
0	3	male	22.0	7.2500	0	1
1	1	female	38.0	71.2833	1	0
2	3	female	26.0	7.9250	1	0

```
inputs.drop(['Sex','male'],axis='columns',inplace=True)
inputs.head(3)
```

	Pclass	Age	Fare	female
0	3	22.0	7.2500	0
1	1	38.0	71.2833	1
2	3	26.0	7.9250	1

```
inputs.columns[inputs.isna().any()]
```

```
Index(['Age'], dtype='object')
```

```
inputs.Age[:10]
```

```
0    22.0
1    38.0
2    26.0
6    54.0
7     2.0
8    27.0
9    14.0
Name: Age, dtype: float64
```

Saved successfully!

```
inputs.Age = inputs.Age.fillna(inputs.Age.mean())
inputs.head()
```

	Pclass	Age	Fare	female
2	3	26.0	7.9250	1

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(inputs,target,test_size=0.3)
```

```
from sklearn.naive_bayes import GaussianNB
model = GaussianNB()
```

```
model.fit(X_train,y_train)
```

```
GaussianNB(priors=None, var_smoothing=1e-09)
```

```
model.score(X_test,y_test)
```

```
0.7910447761194029
```

```
X_test[0:10]
```

	Pclass	Age	Fare	female
865	2	42.000000	13.0000	1
601	3	29.699118	7.8958	0
489	3	9.000000	15.9000	0
579	3	32.000000	7.9250	0
302	3	19.000000	0.0000	0
			2	1
			8	1
646	3	19.000000	7.8958	0
76	3	29.699118	7.8958	0
576	2	34.000000	13.0000	1

Saved successfully!

```
y_test[0:10]
```

```
865    1
601    0
489    1
579    1
302    0
240    0
100    0
646    0
76     0
```

```
576    1
Name: Survived, dtype: int64
```

```
model.predict(X_test[0:10])
```

```
array([1, 0, 0, 0, 0, 0, 0, 0, 0, 1])
```

```
model.predict_proba(X_test[:10])
```

```
array([[0.34861761, 0.65138239],
       [0.97054149, 0.02945851],
       [0.95157351, 0.04842649],
       [0.97114304, 0.02885696],
       [0.96131592, 0.03868408],
       [0.53173094, 0.46826906],
       [0.52016788, 0.47983212],
       [0.96420863, 0.03579137],
       [0.97054149, 0.02945851],
       [0.35130816, 0.64869184]])
```

```
from sklearn.model_selection import cross_val_score
cross_val_score(GaussianNB(),X_train, y_train, cv=5)
```

```
array([0.688     , 0.776     , 0.76      , 0.7983871 , 0.75806452])
```

Saved successfully!

✓ 0s completed at 18:38