

In [1]:

```
1 !pip install catboost
```

Collecting catboost

Downloading catboost-1.2-cp39-cp39-win\_amd64.whl (101.0 MB)

----- 101.0/101.0 MB 2.4 MB/s eta

0:00:00

Requirement already satisfied: scipy in c:\users\hp\anaconda3\lib\site-packages (from catboost) (1.9.1)

Collecting graphviz

Downloading graphviz-0.20.1-py3-none-any.whl (47 kB)

----- 47.0/47.0 kB 782.9 kB/s eta

0:00:00

Requirement already satisfied: six in c:\users\hp\appdata\roaming\python\python39\site-packages (from catboost) (1.16.0)

Requirement already satisfied: pandas>=0.24 in c:\users\hp\anaconda3\lib\site-packages (from catboost) (1.4.4)

Requirement already satisfied: numpy>=1.16.0 in c:\users\hp\anaconda3\lib\site-packages (from catboost) (1.23.5)

Requirement already satisfied: matplotlib in c:\users\hp\anaconda3\lib\site-packages (from catboost) (3.5.2)

Requirement already satisfied: plotly in c:\users\hp\anaconda3\lib\site-packages (from catboost) (5.9.0)

Requirement already satisfied: pytz>=2020.1 in c:\users\hp\anaconda3\lib\site-packages (from pandas>=0.24->catboost) (2022.1)

Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\hp\appdata\roaming\python\python39\site-packages (from pandas>=0.24->catboost) (2.8.2)

Requirement already satisfied: pillow>=6.2.0 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->catboost) (9.2.0)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->catboost) (1.4.2)

Requirement already satisfied: packaging>=20.0 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->catboost) (21.3)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->catboost) (4.25.0)

Requirement already satisfied: pyparsing>=2.2.1 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->catboost) (3.0.9)

Requirement already satisfied: cyclor>=0.10 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->catboost) (0.11.0)

Requirement already satisfied: tenacity>=6.2.0 in c:\users\hp\anaconda3\lib\site-packages (from plotly->catboost) (8.0.1)

Installing collected packages: graphviz, catboost

Successfully installed catboost-1.2 graphviz-0.20.1

In [2]:

```
1 import pandas as pd
2 import numpy as np
3 from catboost.datasets import titanic
```

In [3]:

```
1 titanic_train, titanic_test = titanic()
```

In [4]:

```
1 column_sort = [ 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket',  
2 'Fare', 'Cabin', 'Embarked', 'Survived']
```

In [5]:

```
1 train = titanic_train[column_sort]
2 train.set_index('Pclass')
```

Out[5]:

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Surviv
Pclass										
3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN		S
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85		C
3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN		S
1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123		S
3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN		S
...	...	...	...	...	...	...	...	...		...
2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN		S
1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42		S
3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN		S
1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148		C
3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN		Q

891 rows × 10 columns



In [6]:

```
1 test = titanic_test
2 train.head()
```

Out[6]:

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



In [7]:

```
1 test['Survived'] = 2
2 test.sample(5)
```

Out[7]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
380	1272	3	O'Connor, Mr. Patrick	male	NaN	0	0	366713	7.7500	Na
316	1208	1	Spencer, Mr. William Augustus	male	57.0	1	0	PC 17569	146.5208	B7
15	907	2	del Carlo, Mrs. Sebastiano (Argenia Genovesi)	female	24.0	1	0	SC/PARIS 2167	27.7208	Na
276	1168	2	Parker, Mr. Clifford Richard	male	28.0	0	0	SC 14888	10.5000	Na
398	1290	3	Larsson-Rondberg, Mr. Edvard A	male	22.0	0	0	347065	7.7750	Na

In [9]:

```
1 df = pd.concat([train,test],ignore_index = False)
```

In [10]:

```
1 df = df.drop(['Name', 'Age'], axis=1)
2 df.isnull().sum(axis=0)
```

Out[10]:

```
Pclass      0
Sex          0
SibSp       0
Parch       0
Ticket      0
Fare        1
Cabin     1014
Embarked     2
Survived     0
PassengerId 891
dtype: int64
```

In [11]:

```
1 df['Embarked'] = df['Embarked'].fillna('S')
2 df['Cabin'] = df['Cabin'].fillna('Undefined')
3 df.fillna(-999, inplace=True)
```

In [12]:

```
1 train = df[df.Survived != 2]
```

In [13]:

```
1 test = df[df.Survived == 2]
2 test = test.drop(['Survived'], axis=1)
```

In [14]:

```
1 target = train.pop('Survived')
2 target.head()
```

Out[14]:

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

Name: Survived, dtype: int64

In [15]:

```
1 cat_features_index = np.where(train.dtypes != float)[0]
```

In [16]:

```
1 from sklearn.model_selection import train_test_split
```

In [17]:

```
1 X_train, X_test, y_train, y_test = train_test_split(train, target,
2 train_size=0.85, random_state=1234)
```

In [18]:

```
1 from catboost import CatBoostClassifier
2 model = CatBoostClassifier(
3     custom_loss=['Accuracy'],
4     random_seed=42)
```

In [19]:

```
1 model = CatBoostClassifier(eval_metric='Accuracy',  
2 use_best_model=True, random_seed=42)
```

In [20]:

```
1 model.fit(X_train, y_train, cat_features=cat_features_index,  
2 eval_set=(X_test, y_test))
```

```
total: 878ms remaining: 31.6s  
27: learn: 0.8203435 test: 0.8059701 best: 0.8059701 (0)  
total: 901ms remaining: 31.3s  
28: learn: 0.8203435 test: 0.8059701 best: 0.8059701 (0)  
total: 931ms remaining: 31.2s  
29: learn: 0.8216645 test: 0.8059701 best: 0.8059701 (0)  
total: 957ms remaining: 31s  
30: learn: 0.8216645 test: 0.8059701 best: 0.8059701 (0)  
total: 985ms remaining: 30.8s  
31: learn: 0.8203435 test: 0.8059701 best: 0.8059701 (0)  
total: 1.01s remaining: 30.7s  
32: learn: 0.8216645 test: 0.8059701 best: 0.8059701 (0)  
total: 1.04s remaining: 30.6s  
33: learn: 0.8216645 test: 0.8059701 best: 0.8059701 (0)  
total: 1.07s remaining: 30.3s  
34: learn: 0.8229855 test: 0.8059701 best: 0.8059701 (0)  
total: 1.09s remaining: 30.2s  
35: learn: 0.8216645 test: 0.8059701 best: 0.8059701 (0)  
total: 1.12s remaining: 30.1s  
36: learn: 0.8216645 test: 0.8059701 best: 0.8059701 (0)
```

In [21]:

```
1 from catboost import cv  
2 from sklearn.metrics import accuracy_score
```

In [22]:

```
1 print('the test accuracy is {:.6f}'.format(accuracy_score(  
2 y_test, model.predict(X_test))))
```

the test accuracy is :0.828358