In [2]:

```
import pandas as pd
import numpy as np
#loading dataset
data=pd.read_csv("Desktop/titanic/dataset/train.csv")
#head() prints a few egs from the starting part of dataset
data.head()
#tail() prints a few egs from the ending part of dataset
data.tail()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

In [3]:

```
#creating survived as a target column
#for selecying a particular column we make use of [] and ''
column_target=['Survived']

#taking into accounts important features of dataset
column_train=['Age','Pclass','Sex','Fare']

X=data[column_train] #X_dim 891 X 4
print(np.shape(X))
Y=data[column_target] #Y_dim 891 X 1
print(np.shape(Y))
#data seperated
```

```
(891, 4)
(891, 1)
```

In [23]:

```
#checking and replacing incompatible datatypes
#basically checking Nan values
X['Sex'].isnull().sum()
X['Pclass'].isnull().sum()
X['Fare'].isnull().sum()
X['Age'].isnull().sum()
#filling null attributes with median values
X['Age']=X['Age'].fillna(X['Age'].median())
X['Age'].head()

#now, our data is not really prepared .
#in sklearn we cannot have values as categorial variables like we have in Sex viz. male, female
#so lets convert it to integer values males 0 females is 1

d={'male':0, 'female':1} # initializing dictionary
X['Sex']=X['Sex'].apply(lambda x:d[x])
X['Sex'].head()
```

```
C:\Users\kruti\anaconda3\lib\site-packages\ipykernel_launcher.py:8: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\kruti\anaconda3\lib\site-packages\ipykernel_launcher.py:16: SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  app.launch_new_instance()
```

Out[23]:

```
0    0
1    1
2    1
3    1
4    0
Name: Sex, dtype: int64
```

In [24]:

```
X.head()
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.33,random_state=42) # random_state j
ust like seed
```

In [30]:

```
from sklearn import svm
clf=svm.LinearSVC()
clf.fit(X_train,Y_train)
print(clf)
```

```
LinearSVC(C=1.0, class_weight=None, dual=True, fit_intercept=True,
          intercept_scaling=1, loss='squared_hinge', max_iter=1000,
          multi_class='ovr', penalty='l2', random_state=None, tol=0.0001,
          verbose=0)
```

```
C:\Users\kruti\anaconda3\lib\site-packages\sklearn\utils\validation.py:760: DataConversionWarning:
A column-vector y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\kruti\anaconda3\lib\site-packages\sklearn\svm\_base.py:947: ConvergenceWarning: Liblinear
failed to converge, increase the number of iterations.
  "the number of iterations.", ConvergenceWarning)
```

In [42]:

```
print(clf.predict(X_test[0:1]))
print(clf.predict(X_test[0:10]))
print(clf.score(X_test,Y_test))
```

```
[0]
[0 0 0 1 1 1 1 0 1 1]
0.823728813559322
```