File - H:\PycharmProjects\Hons\Scripts\process.py

```
 1 import numpy as np
 2 from sklearn import preprocessing
 3
 4
 5 def load_close(file):
 6     stockFile = open(file, 'rb').readlines()[1:]
 7     data = []
 8     dates = []
 9     for line in stockFile:
10         try:
11             closePrice = float(line.split(b',')[4])
12             data.append(closePrice)
13             closeDate = (line.split(b',')[0])
14             dates.append(closeDate.decode())
15
16         except:
17             continue
18
19     return data, dates
20
21
22 def split(data, train, predict, step, scale):
23     X = []
24     Y = []
25     for i in range(0,len(data), step):
26         try:
27             xi = data[i:i+train]
28             yi = data[i+train+predict]
29
30             dataset = np.array(data[i:i + train + predict]
   )
31
32             if scale:
33                 dataset = preprocessing.scale(dataset)
34
35             xi = dataset[:-1]
36             yi = dataset[-1]
37
38         except:
39             break
40
41         X.append(xi)
42         Y.append(yi)
43
44     return X, Y
```

```
45
46
47 def shuffle(a, b):
48     assert len(a) == len(b)
49     shuffledA = np.empty(a.shape, dtype = a.dtype)
50     shuffledB = np.empty(b.shape, dtype = b.dtype)
51
52     permutation = np.random.permutation(len(a))
53     for oldIndex, newIndex in enumerate(permutation):
54         shuffledA[newIndex] = a[oldIndex]
55         shuffledB[newIndex] = b[oldIndex]
56
57     return shuffledA, shuffledB
58
59
60 def training_data(x, y, percentage):
61     xTrain = x[0:int(len(x) * percentage)]
62     yTrain = y[0:int(len(y) * percentage)]
63
64     xTrain, yTrain = shuffle(xTrain, yTrain)
65
66     xTest = x[int(len(x) * percentage):]
67     yTest = y[int(len(y) * percentage):]
68
69     return xTrain, xTest, yTrain, yTest
70
71
```