

```
1 # Gets the feature and target vectors for each dataset
2
3 # imports
4 import pandas as pd
5 import numpy as np
6
7 # get airlines features/targets
8 def prep_airlines():
9
10     data_dir = '../data/airline_satisfaction/'
11     df = pd.read_csv(data_dir + 'train.csv', index_col=0)
12     df.drop('id', axis=1, inplace=True)
13     df.dropna(inplace=True)
14
15     to_encode = ['Gender', 'Customer Type', 'Type of Travel', 'Class',
16 'satisfaction']
17     cleanup = dict()
18     for col in to_encode:
19         if col == 'satisfaction':
20             cleanup[col] = {k:v for k,v in zip(df[col].unique(), [0,1])}
21         elif len(df[col].unique()) == 2:
22             cleanup[col] = {k:v for k,v in zip(df[col].unique(), [-1,1])}
23         elif len(df[col].unique()) == 3:
24             cleanup[col] = {k:v for k,v in zip(df[col].unique(), [-1,0,1])}
25
26     df = df.replace(cleanup)
27
28     X = np.array(df[df.columns[:-1]])
29     y = np.array(df[df.columns[-1]])
30
31     return X,y
32
33 # get incomes features/targets
34 def prep_income():
35
36     data_dir = '../data/income/'
37     df = pd.read_csv(data_dir + 'train.csv')
38
39     drop = ['native-country']
40     df.drop(drop, axis=1, inplace=True)
41     df.dropna(inplace=True)
42
43     onehot_cols = [col for col in df.columns if df[col].dtype == 'object']
44     df = pd.get_dummies(data=df, columns=onehot_cols)
45
46     X = np.array(df[df.columns[df.columns != 'income_>50K']])
47     y = np.array(df['income_>50K'])
48
49     return X,y
50
```

```
47
50 # get phishing websites features/targets
51 def prep_phishing():
52
53     data_dir = '../data/phishing_website/'
54     df = pd.read_csv(data_dir + 'phishing.csv.xls', index_col=0)
55
56     df['class'] = df['class'].map({-1:0, 1:1})
57
58     X = np.array(df[df.columns[:-1]])
59     y = np.array(df[df.columns[-1]])
60
61     return X,y
62
63 # get surgical complications features/targets
64 def prep_surgical():
65
66     data_dir = '../data/surgical_complications/'
67     df = pd.read_csv(data_dir + 'Surgical-deepnet.csv')
68
69     X = np.array(df[df.columns[:-1]])
70     y = np.array(df[df.columns[-1]])
71
72     return X,y
```