Load Dataset
Load iris.csv to two separate **numpy** arrays **X**, **y**
Load features into **X** [sepal.length, sepal.width, petal.length, petal.width]
Load labels into **y** [0,1]

Shuffling
Randomize X and corresponding y simultaneously

Train Test Split
X_train = first 80% of X
y_train = first 80% of y
X_test = rest X
y_test = rest y

Algorithm [Test Prediction]
k = 5
X_train = (M, N) # N columns with M rows [for this case N should be 4]
y_train = (M, 1) # 1 columns with M rows
X_test = (M', N) # N columns with M' rows [for this case N should be 4]
y_test = (M', 1) # 1 columns with M' rows

```
y_test_predicted = new numpy array of size M'
for i in range(len(X_test)):
        x_test = X_test[i]
        D = new numpy array of size M
        D = Calculate euclidean distances between x_test and X_train
        min_dist_indices  = find k indices in D where values are minimum
        y_neighbor = y_train[ min_dist_indices ]
        y_test_predicted[ i ] = the value that occurs most in y_neighbor
```

Metrics Calculation
Calculate the accuracy by comparing y_test and y_test_predicted
Print the accuracy (Test)

Notes:

- Pandas library for csv read and shuffling
- Load csv into numpy array
- shuffle two numpy arrays together
- train test split
- distance between two numpy arrays
- numpy.argmin