Naive Bayes predictor

```
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import StratifiedKFold
import numpy as np
import pickle
from sklearn.metrics import accuracy_score, classification_report,
confusion_matrix
```

Load the Data Set

will be split to test and train

```
X = np.load("../X.npy")
y = np.load("../y.npy")

with open("../label_map.pkl", "rb") as f:
    label_map = pickle.load(f)
```

Split the data to n-fold

```
nfold = StratifiedKFold(n_splits=5, shuffle=True, random_state=777)
fold = 1
```

Separate them into 5 different modules

where the training data get scaled

```
import joblib
from sklearn.preprocessing import StandardScaler
for training_index, testing_index in nfold.split(X,
                                                 y):  # the loop calls next()
automatically to give the proper splitted sets for the fold
    # train_index: A NumPy array holds all indices of the training items in
this fold, e.x. [0 4 7 ... 2887 2996 2999]
    print(f"\n--- Fold {fold} ---")

    # Get the training and testing data for this fold
    X_train, X_test = X[training_index], X[testing_index]
    y_train, y_test = y[training_index], y[testing_index]


    scaler = StandardScaler()
    X_train = scaler.fit_transform(X_train)
    X_test = scaler.transform(X_test)

    # Normalize feature values
    """
    scaler = StandardScaler()
    X_train = scaler.fit_transform(X_train)
```

```
    X_test = scaler.transform(X_test)
    """

    # Train Naive Bayes model
    model = GaussianNB()
    model.fit(X_train, y_train)

    # Predict and evaluate
    y_pred = model.predict(X_test)
    joblib.dump(model, f"NB_model{fold}.pkl")
    joblib.dump(scaler, f"NB_scaler{fold}.pkl")

    acc = accuracy_score(y_test, y_pred)
    print(f"Accuracy: {acc:.4f}")
    print("Classification Report:")
    print(classification_report(y_test, y_pred,
target_names=label_map.values()))

    # Confusion matrix (optional)
    cm = confusion_matrix(y_test, y_pred)
    print("Confusion Matrix:")
    print(cm)

    fold += 1


--- Fold 1 ---
Accuracy: 0.7262
Classification Report:
              precision    recall  f1-score   support

        cats       0.67      0.69      0.68       200
       panda       0.69      0.79      0.74       199
     spiders       0.84      0.69      0.76       200

    accuracy                           0.73       599
   macro avg       0.74      0.73      0.73       599
weighted avg       0.74      0.73      0.73       599

Confusion Matrix:
[[139  40  21]
 [ 37 157   5]
 [ 31  30 139]]

--- Fold 2 ---
Accuracy: 0.7446
Classification Report:
              precision    recall  f1-score   support

        cats       0.74      0.71      0.73       200
       panda       0.70      0.86      0.77       199
     spiders       0.81      0.66      0.73       200

    accuracy                           0.74       599
   macro avg       0.75      0.74      0.74       599
```

```
weighted avg       0.75      0.74      0.74       599

Confusion Matrix:
[[143  32  25]
 [ 22 171   6]
 [ 27  41 132]]

--- Fold 3 ---
Accuracy: 0.7262
Classification Report:
              precision    recall  f1-score   support

        cats       0.73      0.65      0.69       200
       panda       0.68      0.92      0.78       199
     spiders       0.81      0.60      0.69       200

    accuracy                           0.73       599
   macro avg       0.74      0.73      0.72       599
weighted avg       0.74      0.73      0.72       599

Confusion Matrix:
[[130  45  25]
 [ 11 184   4]
 [ 37  42 121]]

--- Fold 4 ---
Accuracy: 0.7007
Classification Report:
              precision    recall  f1-score   support

        cats       0.68      0.67      0.67       200
       panda       0.66      0.84      0.74       198
     spiders       0.80      0.59      0.68       200

    accuracy                           0.70       598
   macro avg       0.71      0.70      0.70       598
weighted avg       0.71      0.70      0.70       598

Confusion Matrix:
[[133  42  25]
 [ 26 167   5]
 [ 37  44 119]]

--- Fold 5 ---
Accuracy: 0.7441
Classification Report:
              precision    recall  f1-score   support

        cats       0.68      0.69      0.68       200
       panda       0.74      0.85      0.79       198
     spiders       0.83      0.69      0.76       200

    accuracy                           0.74       598
   macro avg       0.75      0.74      0.74       598
weighted avg       0.75      0.74      0.74       598
```

```
Confusion Matrix:
[[138  37  25]
 [ 27 168   3]
 [ 38  23 139]]
```