```python
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler

# Source: https://medium.com/@sdoshi579/classification-of-music-into-different-genres-using-keras-82ab5339efe0

# Read in dataset
data = pd.read_csv('GTZAN Genre Classification\GTZAN Dataset\data.csv')
data.head()

# Drop unnecessary columns
data = data.drop(['filename'], axis = 1)
data.head()

# Encode genres into integers
genre_list = data.iloc[:, -1]
encoder = LabelEncoder()
y = encoder.fit_transform(genre_list)
print(y)

# Normalizing the dataset
scaler = StandardScaler()
X = scaler.fit_transform(np.array(data.iloc[:, :-1], dtype = float))
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

# Store the data
np.save('GTZAN Genre Classification\GTZAN Dataset\X_train.npy', X_train)
np.save('GTZAN Genre Classification\GTZAN Dataset\X_test.npy', X_test)
np.save('GTZAN Genre Classification\GTZAN Dataset\y_train.npy', y_train)
np.save('GTZAN Genre Classification\GTZAN Dataset\y_test.npy', y_test)


```