import numpy as np

import pandas as pd

from sklearn.preprocessing import StandardScaler, MinMaxScaler

def process\_data(train, test, lags):

attr = 'Lane 1 Flow (Veh/5 Minutes)'

df1 = pd.read\_csv(train, encoding='utf-8').fillna(0)

df2 = pd.read\_csv(test, encoding='utf-8').fillna(0)

# scaler = StandardScaler().fit(df1[attr].values)

scaler = MinMaxScaler(feature\_range=(0, 1)).fit(df1[attr].values.reshape(-1, 1))

flow1 = scaler.transform(df1[attr].values.reshape(-1, 1)).reshape(1, -1)[0]

flow2 = scaler.transform(df2[attr].values.reshape(-1, 1)).reshape(1, -1)[0]

train, test = [], []

for i in range(lags, len(flow1)):

train.append(flow1[i - lags: i + 1])

for i in range(lags, len(flow2)):

test.append(flow2[i - lags: i + 1])

train = np.array(train)

test = np.array(test)

np.random.shuffle(train)

X\_train = train[:, :-1]

y\_train = train[:, -1]

X\_test = test[:, :-1]

y\_test = test[:, -1

return X\_train, y\_train, X\_test, y\_test, scaler