########## 검색어(Search1, Search2) 테이블의 키워드(KWD\_NM) 변수에서 제품군 분류명(중분류) 분리하기 ##########

import pandas as pd

import re

from google.colab import drive

drive.mount('/content/gdrive')

search1 = pd.read\_csv('gdrive/My Drive/lpoint2.csv', engine='python', encoding='cp949')

master\_brand = pd.read\_csv('gdrive/My Drive/master\_brand.csv', engine='python', encoding='utf8')

len(set(search1["KWD\_NM"])) # search1에서 unique한 keyword의 개수 확인

# PD\_BRA\_NM 클렌징 ("[]", ".", (주) 없애기)

myBRA\_NM = []

for br\_nm in master\_brand["PD\_BRA\_NM"]:

b = br\_nm[1:-1] if re.match('\[.\*\]', br\_nm) else br\_nm

if b[-1] == ".": b = b[:-1]

if "(주)" in b: b = b.replace("(주)", "")

if "(유)" in b: b = b.replace("(유)", "")

myBRA\_NM.append(b)

master\_brand["PD\_BRA\_NM"] = myBRA\_NM

master\_brand.to\_csv("gdrive/My Drive/master\_brand.csv", encoding="cp949")

len(set(master\_brand["PD\_BRA\_NM"])) # master\_brand에서 unique한 Brand Name의 개수 확인

# 검색어, 브랜드에서 unique한 것만 추출

KWD\_unique = set(search1["KWD\_NM"])

BRD\_unique = set(master\_brand["PD\_BRA\_NM"])

# 분류명에서 unique한 것만 추출

CLAC1\_unique = set(master\_brand["CLAC1\_NM"])

CLAC2\_unique = set(master\_brand["CLAC2\_NM"])

CLAC3\_unique = set(master\_brand["CLAC3\_NM"])

# 분류명 여러 개짜리 분리 (예. 속옷/양말/홈웨어)

CLAC3\_unique2 = []

for clac3 in CLAC3\_unique:

c1 = re.compile(r'(.\*)/(.\*)').search(clac3)

if c1:

c2 = re.compile(r'(.\*)/(.\*)/(.\*)').search(clac3)

if c2:

CLAC3\_unique2.append(c2.group(1))

CLAC3\_unique2.append(c2.group(2))

CLAC3\_unique2.append(c2.group(3))

else:

CLAC3\_unique2.append(c1.group(1))

CLAC3\_unique2.append(c1.group(2))

else: CLAC3\_unique2.append(clac3)

CLAC3\_unique2 = list(set(CLAC3\_unique2))

CLAC3\_unique2

CLAC3\_unique2.remove("유아")

CLAC2\_unique2 = []

for clac2 in CLAC2\_unique:

c1 = re.compile(r'(.\*)/(.\*)').search(clac2)

if c1:

c2 = re.compile(r'(.\*)/(.\*)/(.\*)').search(clac2)

if c2:

CLAC2\_unique2.append(c2.group(1))

CLAC2\_unique2.append(c2.group(2))

CLAC2\_unique2.append(c2.group(3))

else:

CLAC2\_unique2.append(c1.group(1))

CLAC2\_unique2.append(c1.group(2))

else: CLAC2\_unique2.append(clac2)

CLAC2\_unique2 = list(set(CLAC2\_unique2))

CLAC1\_unique2 = []

for clac1 in CLAC1\_unique:

c1 = re.compile(r'(.\*)/(.\*)').search(clac1)

if c1:

c2 = re.compile(r'(.\*)/(.\*)/(.\*)').search(clac1)

if c2:

CLAC1\_unique2.append(c2.group(1))

CLAC1\_unique2.append(c2.group(2))

CLAC1\_unique2.append(c2.group(3))

else:

CLAC1\_unique2.append(c1.group(1))

CLAC1\_unique2.append(c1.group(2))

else: CLAC1\_unique2.append(clac1)

CLAC1\_unique2 = list(set(CLAC1\_unique2))

### Keyword Name에서 Brand Name 분리하기

# brand name이 분류명과 같은 것 제거

brd\_is\_clac = []

for brd in BRD\_unique:

for clac in set(CLAC1\_unique2 + CLAC2\_unique2 + CLAC3\_unique2):

if brd == clac: brd\_is\_clac.append(brd)

for brd in brd\_is\_clac:

BRD\_unique.remove(brd)

print(brd\_is\_clac)

BRD\_unique.remove('여성')

BRD\_unique.remove('크로스')

# "()" 부분 제거 (ex. 게스(구두))

myBRD\_unique = []

for brd in BRD\_unique:

myBRD\_unique.append( re.compile("[(].\*[)]").sub('', brd) )

BRD\_unique = set(BRD\_unique)

len(BRD\_unique) # unique한 Brand Name의 개수 확인

myBRD, myKWD = [], []

p = re.compile('(\w\*) \*.\*')

for kwd in KWD\_unique:

br = ""

k = p.match(kwd)

for br\_nm in BRD\_unique:

if br\_nm == k.group(1): br = br\_nm

elif len(br\_nm) > len(br):

if len(br\_nm) > 2 and br\_nm in kwd: br = br\_nm

elif kwd[:2] == br\_nm: br = br\_nm

myBRD.append(br)

myKWD.append(kwd.replace(br, "").lstrip())

mysearch1 = pd.DataFrame({'KWD':list(KWD\_unique), 'myBRD':myBRD, 'myKWD':myKWD})

for i in range(len(mysearch1)):

if "스트라이프" in mysearch1.loc[i, "KWD"] and mysearch1.loc[i, "myBRD"] == "트라이":

mysearch1.loc[i, "myBRD"] = ""

mysearch1.loc[i, "myKWD"] = mysearch1.loc[i, "myKWD"].replace("스프", "스트라이프")

## search1을 제품군으로 분류하기

mysearch1\_1 = mysearch1[mysearch1["myKWD"] != ""] # keyword = brand name 인 것 제외

len(mysearch1\_1)

for clac3 in CLAC3\_unique2:

if len(clac3) == 1: print(clac3, end=" ")

# keyword가 CLAC\_NM를 포함하는 것 분류

myCLAC1, myCLAC2, myCLAC3 = [], [], []

for kwd in mysearch1\_1["myKWD"]:

c1, c2, c3 = "", "", ""

if kwd in CLAC2\_unique2: c2 = kwd

elif kwd in CLAC3\_unique2: c3 = kwd

elif kwd in CLAC1\_unique2: c1 = kwd

else:

for clac2 in CLAC2\_unique2:

if clac2 in kwd: c2 = clac2

if c1 == "" and c2 == "" and c3 == "":

for clac3 in CLAC3\_unique2:

if clac3 in kwd and len(clac3) > 1: c3 = clac3

if c3 == "":

for clac1 in CLAC1\_unique2:

if clac1 in kwd: c1 = clac1

myCLAC1.append(c1)

myCLAC2.append(c2)

myCLAC3.append(c3)

mysearch1\_2 = pd.DataFrame({'KWD':mysearch1\_1["KWD"], 'myBRD':mysearch1\_1["myBRD"], 'myKWD':mysearch1\_1["myKWD"], 'myCLAC1\_NM':myCLAC1, 'myCLAC2\_NM':myCLAC2, 'myCLAC3\_NM':myCLAC3})

mysearch1\_2 = mysearch1\_2[['KWD', 'myBRD', 'myKWD', 'myCLAC1\_NM', 'myCLAC2\_NM', 'myCLAC3\_NM']]

# 분류 안 된 것에서 '여성', '남성', '유아동' 분리

mysearch1\_3 = mysearch1\_2[mysearch1\_2["myCLAC1\_NM"]==""][mysearch1\_2["myCLAC2\_NM"]==""][mysearch1\_2["myCLAC3\_NM"]==""][["KWD", "myBRD", "myKWD", 'myCLAC1\_NM', 'myCLAC2\_NM', 'myCLAC3\_NM']]

myGROUP, myKWD = [], []

for kwd in mysearch1\_3["myKWD"]:

if "여성" in kwd:

myKWD.append(kwd.replace("여성", "").strip())

myGROUP.append("여성")

elif "여자" in kwd:

myKWD.append(kwd.replace("여자", "").strip())

myGROUP.append("여성")

elif "남성" in kwd:

myKWD.append(kwd.replace("남성", "").strip())

myGROUP.append("남성")

elif "남자" in kwd:

myKWD.append(kwd.replace("남자", "").strip())

myGROUP.append("남성")

elif "아동" in kwd or "유아" in kwd or "키즈" in kwd:

myKWD.append(kwd.replace("아동", "").strip())

myGROUP.append("유아동")

elif "남아" in kwd:

myKWD.append(kwd.replace("남아", "").strip())

myGROUP.append("남아")

elif "여아" in kwd:

myKWD.append(kwd.replace("여아", "").strip())

myGROUP.append("여아")

else:

myKWD.append(kwd)

myGROUP.append("")

mysearch1\_3["myGROUP"], mysearch1\_3["myKWD"] = myGROUP, myKWD

mysearch1\_3 = mysearch1\_3[["KWD", "myBRD", "myGROUP", "myKWD", 'myCLAC1\_NM', 'myCLAC2\_NM', 'myCLAC3\_NM']]

CLAC1\_unique3, CLAC2\_unique3, CLAC3\_unique3 = [], [], []

for clac1 in CLAC1\_unique2:

if "여성" in clac1: CLAC1\_unique3.append(clac1.replace("여성", ""))

elif "남성" in clac1: CLAC1\_unique3.append(clac1.replace("남성", ""))

elif "유아동" in clac1: CLAC1\_unique3.append(clac1.replace("남성", ""))

else: CLAC1\_unique3.append(clac1)

for clac2 in CLAC2\_unique2:

if "여성" in clac2: CLAC2\_unique3.append(clac2.replace("여성", ""))

elif "남성" in clac2: CLAC2\_unique3.append(clac2.replace("남성", ""))

elif "남아" in clac2: CLAC2\_unique3.append(clac2.replace("남아", ""))

elif "유아동" in clac2: CLAC2\_unique3.append(clac2.replace("유아동", ""))

elif "유아" in clac2: CLAC2\_unique3.append(clac2.replace("유아", ""))

else: CLAC2\_unique3.append(clac2)

for clac3 in CLAC3\_unique2:

if "여성" in clac3: CLAC3\_unique3.append(clac3.replace("여성", ""))

elif "남성" in clac3: CLAC3\_unique3.append(clac3.replace("남성", ""))

elif "남아" in clac3: CLAC3\_unique3.append(clac3.replace("남아", ""))

elif "유아동" in clac3: CLAC3\_unique3.append(clac3.replace("유아동", ""))

elif "유아" in clac3: CLAC3\_unique3.append(clac3.replace("유아", ""))

elif "아동용" in clac3: CLAC3\_unique3.append(clac3.replace("아동용", ""))

elif "아동" in clac3: CLAC3\_unique3.append(clac3.replace("아동", ""))

else: CLAC3\_unique3.append(clac3)

CLAC1\_unique3 = set(CLAC1\_unique3)

CLAC2\_unique3 = set(CLAC2\_unique3)

CLAC3\_unique3 = set(CLAC3\_unique3)

# 여성, 남성, 아동 빼고 2차 분류

myCLAC1, myCLAC2, myCLAC3 = [], [], []

for kwd in mysearch1\_3["myKWD"]:

c1, c2, c3 = "", "", ""

if kwd in CLAC2\_unique3: c2 = kwd

elif kwd in CLAC3\_unique3: c3 = kwd

elif kwd in CLAC1\_unique3: c1 = kwd

else:

for clac2 in CLAC2\_unique3:

if clac2 in kwd: c2 = clac2

if c1 == "" and c2 == "" and c3 == "":

for clac3 in CLAC3\_unique3:

if clac3 in kwd and len(clac3) > 1: c3 = clac3

if c3 == "":

for clac1 in CLAC1\_unique3:

if clac1 in kwd: c1 = clac1

myCLAC1.append(c1)

myCLAC2.append(c2)

myCLAC3.append(c3)

mysearch1\_4 = pd.DataFrame({'KWD':mysearch1\_3["KWD"], 'myBRD':mysearch1\_3["myBRD"], 'myGROUP':mysearch1\_3["myGROUP"],

'myKWD':mysearch1\_3["myKWD"], 'myCLAC1\_NM':myCLAC1, 'myCLAC2\_NM':myCLAC2, 'myCLAC3\_NM':myCLAC3})

mysearch1\_4 = mysearch1\_4[['KWD', 'myBRD', 'myGROUP', 'myKWD', 'myCLAC1\_NM', 'myCLAC2\_NM', 'myCLAC3\_NM']]

mysearch1\_4

mysearch1\_2.to\_csv("gdrive/My Drive/mysearch1\_2.csv", encoding="cp949")

mysearch1\_4.to\_csv("gdrive/My Drive/mysearch1\_4.csv", encoding="cp949")

########## 수요 예측 모델 (LSTM) ##########

import pandas as pd

from keras.models import Sequential

from keras.layers import Dense, Dropout

from keras.layers import LSTM

from matplotlib import pyplot as plt

from google.colab import drive

drive.mount('/content/gdrive')

count = pd.read\_csv("gdrive/My Drive/count.csv", encoding="utf8")

## 상품군별 (128개 중분류) 수요량 예측

CLAC2\_NM, predicted = [], []

tot\_length = 26

seq\_length = 8

for clac2\_nm in set(count["CLAC2\_NM"]):

n = np.array(count[count["CLAC2\_NM"]==clac2\_nm]["n"])

CLAC2\_NM.append(clac2\_nm)

# 데이터셋 구축

f = (n-np.min(n)) / (np.max(n)-np.min(n))

data, target = [], []

for i in range(0, tot\_length - seq\_length):

data.append(f[i:(i+seq\_length)])

target.append(f[i+seq\_length])

X = np.array(data).reshape(len(data), seq\_length, 1)

y = np.array(target).reshape(len(data), 1)

# 데이터 모델링

model = Sequential()

model.add(LSTM(32, input\_shape=(X.shape[1], X.shape[2])))

Dropout(0.5)

model.add(Dense(16, activation='softsign'))

Dropout(0.5)

model.add(Dense(8, activation='softsign'))

Dropout(0.5)

model.add(Dense(y.shape[1], activation='linear'))

Dropout(0.5)

model.compile(loss='mean\_squared\_error', optimizer='adam')

model.fit(X, y, epochs=1000, verbose=False)

# 수요 예측

y\_hat = model.predict(f[-8:].reshape([1, seq\_length, 1]))

y\_hat\_rescaled = y\_hat\*(np.max(n)-np.min(n))+np.min(n)

predicted.append(y\_hat\_rescaled)

predict = pd.DataFrame({'CLAC2\_NM':CLAC2\_NM, 'predicted\_count':predicted})

predict.to\_csv("gdrive/My Drive/predict\_count.csv", encoding="utf8")