



# 영양제 효과 맞추기

(with. Pandas, Keras, Okt)

Intel Edge AI SW Academy

김경민 김도영 윤희승 조지현

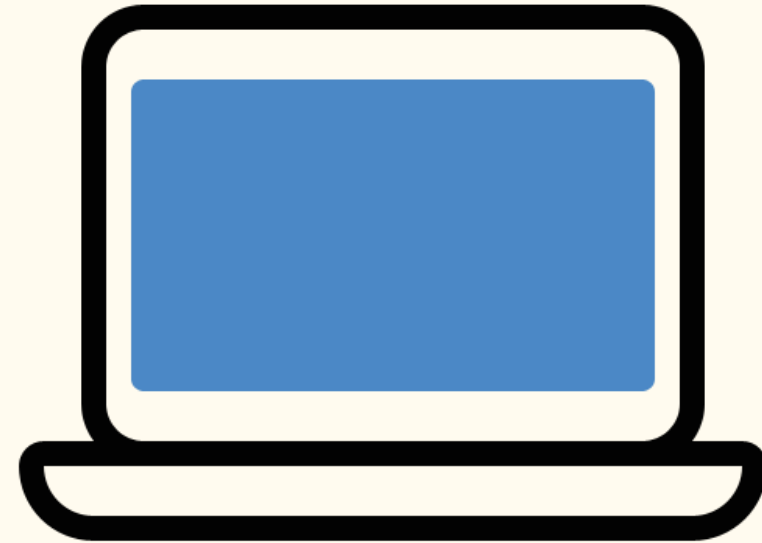


# 목차

1. 프로젝트 목표
2. 프로젝트 구조
3. 프로젝트 결과
4. 아쉬운 점

# 1. 프로젝트 목표

영양제 효과를 알려주는 프로그램 개발



## 2. 프로젝트 구조

미국 최대 건강  
제품 사이트(iHerb)  
에서 영양 제품  
설명 수집

형태소 단위로  
수집한 데이터 분리  
  
불용어 제거  
  
유사한 문장 제거

자연어 처리 모델  
(LSTM) 모델 설계  
및 학습

검증용 데이터로  
학습된 모델 예측

## 2.1. 데이터 수집

- 영양제의 기능 효과를 총 24가지로 분류

### 건강 고민별 케어

노화 & 장수

아유르베다

방광

뼈 & 관절 & 연골

두뇌 & 인지

순환기계

장건강

감기 & 독감

디톡스 & 클렌징

소화기계

귀

에너지

눈 & 시력

모발 & 피부 & 손발톱

심혈관계

면역계

통증 완화

호흡기계

계절성 알레르기

수면

체중 감량

어린이 건강

남성 건강

여성 건강

## 2.1. 데이터 수집



### 상품 정보

#### 상품 설명

- California Gold Nutrition SUPERBABOOST® 프리미엄 크릴 오일
- Aker Biomarine에서 SUPERBABOOST® 자연산 남극 크릴 오일로 제조
- 오메가3 지방산, 포스포리피드, 아스타잔틴 함유
- 심장, 뇌, 눈, 간 건강 증진\*
- 글루텐, 유전자 변형 성분, 대두 무함유
- cGMP 인증 시설에서 생산
- 100% 골드 개런티(Gold Guarantee)

해양 오일은 오메가3 지방산이 풍부하고 강력한 건강 효능을 가진 것으로 알려져 있습니다. California Gold Nutrition의 SUPERBABOOST® 프리미엄 크릴 오일은 작은 새우 같은 갑각류로 만든 일종의 해양 오일을 함유하고 있는 시스템을 건강하게 관리하는 데 효과적입니다.\*

#### 크릴 오일의 효능

크릴 오일은 오메가3의 생체활성 공급원이자 천연 카로티노이드 아스타잔틴으로 잘 알려져 있습니다. 연구에 따르면 크릴 오일의 생체이용률이 높은 이유는 포스포리피드가 오메가3 지방산인 EPA와 DHA에 결합하여 이러한 영양소를 체내 세포에 효율적으로 전달하는 데 도움을 주기 때문인 것으로 나타났습니다.\* 크릴 오일의 아스타잔틴과 오메가3 포스포리피드는 건강한 심장, 두뇌, 눈 및 간 기능을 강화합니다.\*

California Gold Nutrition SUPERBABOOST® 프리미엄 크릴 오일은 고품질의 지속가능한 크릴오일 생산을 선도하는 노르웨이의 수산양식 및 생명공학 기업인 Aker Biomarine이 자연산 남극 크릴을 농축하여 만든 오일을 함유하고 있습니다. 이 크릴 오일은 돈피와 우피가 함유되지 않은 피쉬 젤라틴 소프트젤트 제공됩니다.

```
<div itemprop="description"> == $0
  <ul> ... </ul>
  <p> ... </p>
  <p> ... </p>
  <p> ... </p>
  <p> ... </p>
  <p> ... </p>
  <p> ... </p>
```

```
dt = ''
nutrient_data = driver.find_elements(By.XPATH, value: '//div[@itemprop="description"]/ul')
for data in nutrient_data:
    if not data.text == '':
        data = re.compile('[^가-힝|a-z|A-Z|0-9]').sub( repl: ' ', data.text)
        dt += data
```

```
xpath_temp = '/html/body/div[7]/div/div[3]/div/div/div[1]/div[1]/div[2]/div[2]/div[{}]/div/div[2]/div[1]/a'.format(k)
driver.find_element( by: 'xpath', xpath_temp).click() # 클릭 시 사이트로 이동
time.sleep(5) # 클릭 후 사이트로 이동할 대기 시간 부여
print('count = %d' %k)
```



## 2.2. 데이터 전처리

- 수집한 데이터(문장)을 형태소 단위로 분리
- 불용어 제거
- 유사한 문장을 제거하는 전처리 과정 수행

장관 웰빙을 위한 세척제 허브 건강보조식품 침술사이자 한의학자인 마이클 티에라에 의해서 제제된, 장건강 글루텐 무함유 비건 인증 보충제 프리바이오틱 프로바이오틱 포스트바이오틱 및 효소 함유 MICROBAC 을 통해 프로바이오틱 활성도 5배 보장 소화계 및 면역계 건강 증진 상온 보관 가능, 장건강 소화 건강 증진 음식의 영양학적 가치를 최적화 풀 스펙트럼 효소 포물라 보충제 채식주의자 비건에게 적합 프로바이오틱스 효소 소화 1968년 설립된 가족 기업 GMP 품질 보증, 장건강 규칙적인 배변 활동 측면에서 의사들이 선호하는 프로바이오틱 함유 캡슐당 300억 마리의 프로바이오틱 활성 배양균 프로바이오틱 균주 11종 소화기 균형 여성 건강 면역 건강 프로바이오틱 보충제 유전자 변형 성분 무함유 효소 전문가 가장 향상된 포물라 소화 촉진 효소 판매 1위 브랜드 가장 향상된 효소 포물라 최적화된 소화 기능 지원 미국 소화 효소 판매 1위 보충제 비건 상품 및 코셔



장관 웰빙 세척제 허브 건강 보조 식품 침술 사이 한의학자 마이클 의하다 제제 글루텐 함유 비건 인증 보충 프리바이오틱 프로바이오틱 포스트 바이오 효소 함유 MICROBAC 통해 프로바이오틱 성도 보장 소화계 면역계 건강 증진 상온 보관 가능 소화 건강 증진 음식 영양학 가치 최적화 스펙트럼 효소 포물 보충 채식주의자 비건 적합 프로바이오틱스 효소 소화 1968년 설립 가족 기업 GMP 품질 보증 규칙 배변 활동 측면 의사 선호 프로바이오틱 함유 캡슐 300억 마리 프로바이오틱 배양 프로바이오틱 균주 11 소화기 균형 여성 건강 면역 건강 프로바이오틱 보충 글루텐 함유 인증 유전자 유전자 변형 성분 함유 효소 전문가 향상 포물 소화 촉진 효소 판매 브랜드 향상 효소 포물 최적화 소화 기능 지원 미국 소화 효소 판매 보충 비건 상품 코셔 인증 코셔 파르브 증량제 함유

```

for i in range(len(X)):
    try:
        X[i] = okt.morphs(X[i], stem=True)
    except:
        print('error okt: ', i, len(X))

stopwords = pd.read_csv(filepath_or_buffer='./stopwords.csv', index_col=0)
# effect 수 만큼 반복
for i in range(len(X)):
    try:
        words = []
        # i번째 줄의 길이만큼 반복
        for j in range(len(X[i])):
            if len(X[i][j]) > 1:
                if X[i][j] not in list(stopwords['stopword']):
                    words.append(X[i][j])
            # 단어
        if len(words) >= 10:
            X[i] = ' '.join(words)
        else:
            X[i] = None
            Y[i] = None
    except:
        print('error stopwords: ', i, len(X))

```

불용어 제거

```

from difflib import SequenceMatcher
X = np.array(X)
# effect 데이터를 배열로 변환
for i in range(len(X)-10):
    if X[i] is not None:
        fst = X[i]
        for j in range(i-10, i+10):
            if i != j and X[j] is not None:
                scd = X[j]
                ratio = SequenceMatcher(None, fst, scd).ratio()
                if ratio >= 0.9:
                    print('remove similar data, {} : {}'.format(*args: j, scd))
                    Y[j] = None
                    X[j] = None

```

유사한 문장 제거



## 2.3. 데이터 학습

### • 자연어 처리 모델 설계 및 학습

Embedding

Conv1D

Max  
Pooling 1D

LSTM

LSTM

LSTM

Dropout

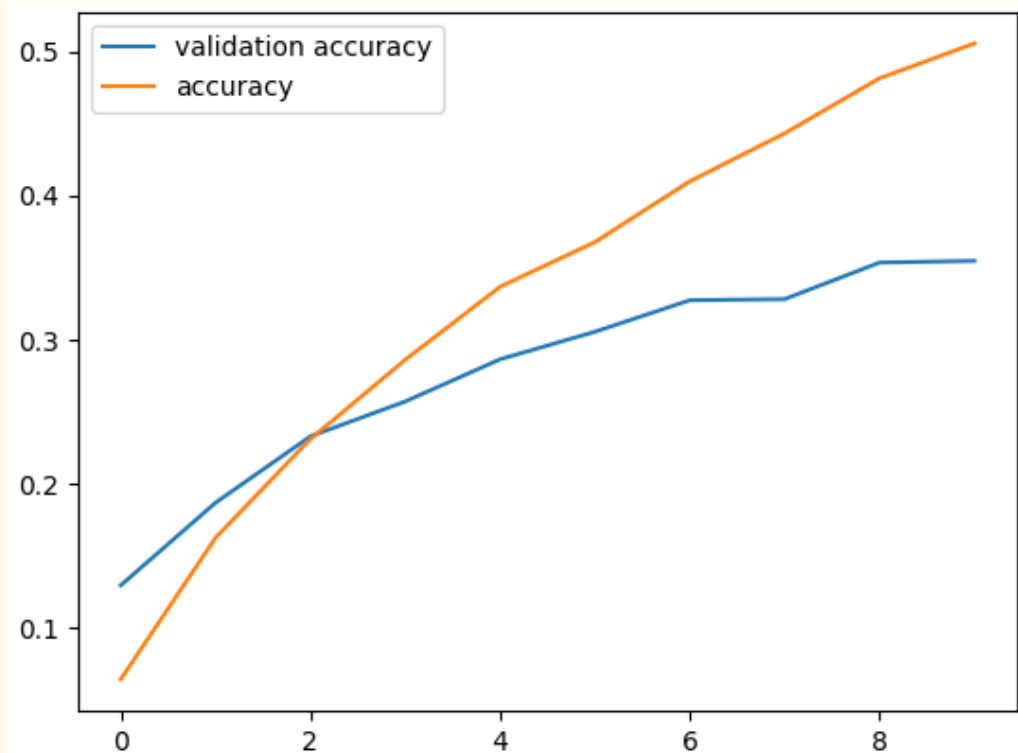
Dropout

Dropout

Flatten

Dense

Dense



```
Epoch 1/10
81/81 [=====] - 73s 882ms/step - loss: 3.0876 - accuracy: 0.0644 - val_loss: 2.8971 - val_accuracy: 0.1296
Epoch 2/10
81/81 [=====] - 71s 873ms/step - loss: 2.6466 - accuracy: 0.1627 - val_loss: 2.5564 - val_accuracy: 0.1870
Epoch 3/10
81/81 [=====] - 71s 875ms/step - loss: 2.3500 - accuracy: 0.2312 - val_loss: 2.4005 - val_accuracy: 0.2330
Epoch 4/10
81/81 [=====] - 71s 872ms/step - loss: 2.1687 - accuracy: 0.2862 - val_loss: 2.3300 - val_accuracy: 0.2572
Epoch 5/10
81/81 [=====] - 71s 873ms/step - loss: 2.0176 - accuracy: 0.3368 - val_loss: 2.2826 - val_accuracy: 0.2865
Epoch 6/10
81/81 [=====] - 69s 857ms/step - loss: 1.8991 - accuracy: 0.3679 - val_loss: 2.2270 - val_accuracy: 0.3056
Epoch 7/10
81/81 [=====] - 126s 2s/step - loss: 1.7674 - accuracy: 0.4100 - val_loss: 2.2270 - val_accuracy: 0.3275
Epoch 8/10
81/81 [=====] - 84s 1s/step - loss: 1.6809 - accuracy: 0.4433 - val_loss: 2.2046 - val_accuracy: 0.3283
Epoch 9/10
81/81 [=====] - 81s 1s/step - loss: 1.5837 - accuracy: 0.4816 - val_loss: 2.2025 - val_accuracy: 0.3536
Epoch 10/10
81/81 [=====] - 71s 874ms/step - loss: 1.4983 - accuracy: 0.5058 - val_loss: 2.1721 - val_accuracy: 0.3548

Process finished with exit code 0
```

### 3. 결과

```
Name: 39, dtype: object
effect      이온 미네랄 미네랄 보충 36회 제공 느낌 순도 확인 어린이 반려동물 안전하다 제품...
category                                         면역계
predict                                         [어린이 건강, 감기독감]
OX                                                X

Name: 41, dtype: object
effect      전립선 면역 시스템 건강 지원 미네랄 보충 강력하다 항산화제 품질 보증 시험 기관 ..
category                                         면역계
predict                                         [어린이 건강, 감기독감]
OX                                                X

Name: 43, dtype: object
effect      보충 건강 강화 순도 효능 보장
category                                         면역계
predict      [뼈 관절 연골, 불면증]
OX                                                X

Name: 44, dtype: object
effect      프로바이오틱 보충 의사 전용 글루텐 함유 인증 제품 유전자 변형 성분 함유 채식주의..
category                                         방광
predict                                         [방광, 장건강]
OX                                                0

Name: 0, dtype: object
effect      Product 90 Bladder Urinary Tract Support Helps...
category                                         방광
predict      [방광, 장건강]
OX                                                0
```

검증에 사용할 데이터를  
크롤링하여 학습된 모델 예측

```
Name: 44, dtype: object
O      25
X      20
Name: OX, dtype: int64
O      0.555556
X      0.444444
Name: OX, dtype: float64

Process finished with exit code 0
```

## 4. 아쉬웠던 점

상품마다 설명이 자세하게 적힌 섹션이 달랐으나  
한 섹션만 크롤링하여 학습시킨 점이 아쉬움

문장의 유사성을 검사할 때,  
프로그램에 적합한 방법과 수치 등 다양한 방법을 적용하지 않고  
한 가지 방법만 시도해 본 점이 아쉬움



**THANK  
YOU**