



# iHerb 추천시스템(CF-KNN기반)

서비스 산업 데이터 활용 빅데이터-B반  
정은원

## 아이허브 선정이유



3만여 개의 제품



세분화 된 제품 카테고리



2천만 개의 고객후기



활발하고 신뢰도 높은  
후기 데이터



천만 이상의 고객 수



해외직구/건강식품을  
소비하는 특정 고객층  
분석가능

---

## 세분화 된 제품 카테고리

### 보충제

5-HTP  
해조류  
아미노산  
항산화  
아슈와간다  
아스타잔틴  
프로폴리스  
베타 글루칸  
바이오틴  
뼈 건강 및 관절  
칼슘  
콜라겐  
CoQ10  
커큐민  
엔자임  
섬유소  
피쉬 오일  
아마씨  
마늘  
은행  
인삼

글루코사민  
녹색 채소 & 슈퍼푸드  
햄프  
허브  
동충요법  
히알루론산  
크릴 오일  
리포소말 비타민C  
루테인  
마카  
마그네슘  
식사 대용  
말크씨슬  
미네랄  
종합비타민  
버섯  
오메가3-6-9  
식물성 단백질  
PQQ(피로로퀴놀린퀴논)  
프로바이오틱스  
케르세틴  
홍국

레스베라트롤  
SAM-e  
쏘팔메토  
스파클리나  
세인트 존스 워트  
비타민B  
비타민C  
비타민D  
비타민E  
아연

### 스포츠

BCAA  
크레아틴  
수분 보충 & 전해질  
L-아르기닌  
L-카르니틴  
L-글루타민  
MCT 오일  
스포츠 바  
스포츠용 종합비타민  
유청 단백질

### 목욕

아로마테라피 & 에센셜 오일  
목욕 & 샤워  
바디 & 마사지 오일  
방역 마스크 & 손 소독제  
헤어 케어  
립 케어  
로션  
가정상비약 & 응급처치  
남성용 그루밍  
구강 케어

### 뷰티

클렌징 & 스크럽  
페이스 마스크 팩 & 필링  
선물용 뷰티  
모이스처라이저 & 크림  
K-뷰티  
립 케어  
메이크업  
메이크업 브러시 & 도구  
트리트먼트 & 세럼

### 식료품

베이킹 재료  
버터 & 스프레드  
코코넛 오일  
커피  
꿀 & 감미료  
견과류 & 씨앗류  
허브 & 향신료  
차

### 홈 & 생활용품

### 베이비 & 키즈

목욕, 피부 & 헤어  
어린이 건강  
식품  
임마 & 임신부

### 반려동물

반려동물 간식  
반려동물 그루밍  
반려동물 보충제

## 세분화 된 이용객 카테고리

● 비건 ● 글루텐 무함유 ● 유제품 무함유 ● 베지테리언 ● 팔레오 ● 유기농 ● 코셔 ● 유전자 변형 성분 무함유 ● 크루엘티 프리 ● 케토

## 리워드 프로그램

640만 달러

2022년 1월 기준 리워드  
크레딧 적립 현황

30만 98백 명 이  
상

무료 제품을 받은 고객

1,700명 이상

100달러 이상을 현금화  
한 고객

iHerb 리워드 크레딧은 지인 및 친구들에게 iHerb를 소개하거나, 제품 구매후기를 작성하거나, 제품에 대한 질문에 답변해 준 고객들께 드리는 iHerb의 감사 표시입니다.



할인 \$5 & 적립 \$5 + 추가 혜택!

+



구매후기를 작성하고 구매후기당 최대 \$101를 적립  
하세요!

+



답변을 작성하고 답변당 최대 \$10를 적립하세요!

+

### 리워드 크레딧 사용하기



향후 구매시 리워드 크레딧을 사용하세요.



# 1. 아이템 웹크롤링-mct oil

Item\_id/title/reviewurl/평점수집

```
Elements Console Sources Network Performance Memory Application Security >> 3
d_99739" id="pid_99739" data-istrial="false">
...
<div class="product-flag-container"> </div> == $0
▼<div class="product-inner product-inner-wide">
  ▼<div class="absolute-link-wrapper">
    <a href="https://kr.iherb.com/pr/california-gold-nutrition-sports-organic-mct-oil-3
    2-fl-oz-946-ml/99739" class="absolute-link product-link" data-ga-event="click"
    data-ga-event-category="Ecommerce" data-ga-event-action="productClick" data-ga-
    event-label="Product List_ListView" data-ga-product-id="99739" data-ga-product-
    position="1" data-product-id="99739" data-part-number="CGN-01787" data-ga-brand-
    code="CGN" data-ga-discount-price="#31,469" itemprop="url" title="California Gold N
    utrition, 유기농 MCT 오일, 946ml(32fl oz)" aria-label="California Gold Nutrition, 유
    기농 MCT 오일, 946ml(32fl oz)"></a>
    ▶<div class="product-image-wrapper">...</div>
    ▶<div itemprop="name" class="product-title">California Gold Nutrition, 유기
    농 MCT 오일, 946ml(32fl oz)>...</div>
  </div>
  ▶<div class="rating">...</div>
  ▶<div itemscope itemtype="http://schema.org/AggregateRating" itemprop="aggregateRatin
  g">...</div>
  ▶<div itemscope itemtype="http://schema.org/Offer" itemprop="offers"> </div>
```

# Item\_id  
# Title 수집

```
▶<div itemprop="name" class="product-title" content="California Gold Nutrition, 유기농 MCT 오일, 946ml(32fl oz)">...</div>
</div>
▼<div class="rating"> == $0
▶<a class="stars" data-ds-target="link" href="https://kr.iherb.com/r/california-gold-nutrition-sports-organic-mct-oil-32-fl-oz-946-ml/99739" title="4.8/5 - 5960 평가">...</a>
▶<a class="rating-count" data-ds-target="link" href="https://kr.iherb.com/r/california-gold-nutrition-sports-organic-mct-oil-32-fl-oz-946-ml/99739" title="4.8/5 - 5960 평가">...</a> flex
</div>
▶<div itemscope itemtype="http://schema.org/AggregateRating" itemprop="aggregateRating">...</div>
▶<div itemscope itemtype="http://schema.org/Offer" itemprop="offer">...</div>
▶<div class="product-price text-nowrap ">...</div>
</div>
▶<div style="display:none">...</div>
</div>
</div>
```

# 고객후기링크  
# 평점 수집

inner.col-xs-12.col-sm-12.col-md-8.col-lg-6 div#pid\_99739.product.ga-product div.product-inner.product-inner-wide div.rating

```
%%time

# 제품 url들 수집

name_list = []
num_list = []
rating_list = []
rev_url_list = []

try:
    for j in range(0,3):
        driver.get('https://kr.iherb.com/c/mct-oil?p={}'.format(j+1))
        for i in range(0,24):

            item = "div.product-inner.product-inner-wide > div.absolute-link-wrapper > a"
            item_raw = driver.find_elements_by_css_selector(item)

            rate=".rating-count"
            rating_raw = driver.find_elements_by_css_selector(rate)

            # 상품번호 크롤링
            num = item_raw[i].get_attribute('data-product-id')
            num_list.append(num)
            # 제목 크롤링
            name = item_raw[i].get_attribute('title')
            name_list.append(name)

            #평점 크롤링
            rating=rating_raw[i].get_attribute('title')
            rating_list.append(rating)
            #리뷰창 크롤링
            review=rating_raw[i].get_attribute('href')
            rev_url_list.append(review)

            print(i,name,rating)
            time.sleep(2)

except:
    driver.close()
```

## 57개 아이템 크롤링

	name	item_id	평점	리뷰url	평점수
0	California Gold Nutrition, 유기농 MCT 오일, 946ml(3...	99739	4.8	<a href="https://kr.iherb.com/r/california-gold-nutriti...">https://kr.iherb.com/r/california-gold-nutriti...</a>	5960
2	Nature's Way, 유기농 MCT 오일, 480ml(16fl oz)	66202	4.7	<a href="https://kr.iherb.com/r/nature-s-way-organic-mc...">https://kr.iherb.com/r/nature-s-way-organic-mc...</a>	6443
4	Sports Research, MCT 오일, 무맛, 32 액량 온스 (946 ml)	71107	4.7	<a href="https://kr.iherb.com/r/sports-research-mct-oil...">https://kr.iherb.com/r/sports-research-mct-oil...</a>	2213
5	NOW Foods, Sports, MCT 오일, 무맛, 473ml(16fl oz)	72853	4.6	<a href="https://kr.iherb.com/r/now-foods-sports-mct-oi...">https://kr.iherb.com/r/now-foods-sports-mct-oi...</a>	1918
6	NOW Foods, 카프릴산, 600mg, 소프트젤 100정	13376	4.8	<a href="https://kr.iherb.com/r/now-foods-caprylic-acid...">https://kr.iherb.com/r/now-foods-caprylic-acid...</a>	3598
7	Sports Research, MCT 오일, 1,000mg, 소프트젤 120정	75244	4.7	<a href="https://kr.iherb.com/r/sports-research-mct-oil...">https://kr.iherb.com/r/sports-research-mct-oil...</a>	391
8	Sports Research, 유기농 중쇄중성지방 오일 가루, 무맛, 300g(10...	84970	4.7	<a href="https://kr.iherb.com/r/sports-research-organic...">https://kr.iherb.com/r/sports-research-organic...</a>	272
9	NOW Foods, Sports, 중쇄중성지방 오일, 946ml(32fl oz)	13585	4.7	<a href="https://kr.iherb.com/r/now-foods-sports-mct-oi...">https://kr.iherb.com/r/now-foods-sports-mct-oi...</a>	708
11	Jarrow Formulas, MCT 오일, 무맛, 591ml(20fl oz)	68705	4.6	<a href="https://kr.iherb.com/r/jarrow-formulas-mct-oil...">https://kr.iherb.com/r/jarrow-formulas-mct-oil...</a>	1562
12	Nutiva, 100% 유기농 코코넛 MCT 오일, 무맛, 473ml(16fl oz)	76228	4.7	<a href="https://kr.iherb.com/r/nutiva-organic-mct-oil-...">https://kr.iherb.com/r/nutiva-organic-mct-oil-...</a>	489
13	Garden of Life, 박사가 만든 케토, 유기농 MCT 분말, 300g(10...	86686	4.7	<a href="https://kr.iherb.com/r/garden-of-life-dr-formu...">https://kr.iherb.com/r/garden-of-life-dr-formu...</a>	476
14	Garden of Life, 닥터 포몰레이티드 두뇌 건강, 100% 유기농 코코넛 ...	79831	4.7	<a href="https://kr.iherb.com/r/garden-of-life-dr-formu...">https://kr.iherb.com/r/garden-of-life-dr-formu...</a>	364
15	Nutiva, 유기농 MCT 분말, 초콜릿, 10.6 oz(300g)	89351	4.4	<a href="https://kr.iherb.com/r/nutiva-organic-mct-powd...">https://kr.iherb.com/r/nutiva-organic-mct-powd...</a>	121
16	NOW Foods, 스포츠, 유기농 MCT 오일, 32 fl oz (946mL)	86757	4.7	<a href="https://kr.iherb.com/r/now-foods-sports-organi...">https://kr.iherb.com/r/now-foods-sports-organi...</a>	62
17	Sports Research, 유기농 MCT 오일, 무맛, 946ml(32fl oz)	79969	4.6	<a href="https://kr.iherb.com/r/sports-research-organic...">https://kr.iherb.com/r/sports-research-organic...</a>	180
18	RSP Nutrition, Clean Active Nutrition, 중쇄중성지방 ...	86431	4.5	<a href="https://kr.iherb.com/r/rsp-nutrition-clean-act...">https://kr.iherb.com/r/rsp-nutrition-clean-act...</a>	249



```
: mean_df = result.mean()  
print(mean_df)
```

```
item_id    2.168254e+238  
평점        4.508696e+00  
평점수      6.281522e+02  
dtype: float64
```

```
: condition1 = result.평점 >= 4.51  
condition2 = result.평점수 >= 628  
df = result.loc[condition1, ['item_id', '평점', '평점수', '리뷰url']]  
df = result.loc[condition2, ['item_id', '평점', '평점수', '리뷰url']]  
print(df)  
print("")  
print("데이터길이: ", len(df))
```

	item_id	평점	평점수	리뷰url
0	99739	4.8	5969	<a href="https://kr.iherb.com/r/california-gold-nutriti...">https://kr.iherb.com/r/california-gold-nutriti...</a>
2	66202	4.7	6447	<a href="https://kr.iherb.com/r/nature-s-way-organic-mc...">https://kr.iherb.com/r/nature-s-way-organic-mc...</a>
4	84966	4.7	719	<a href="https://kr.iherb.com/r/sports-research-organic...">https://kr.iherb.com/r/sports-research-organic...</a>
5	71107	4.7	2215	<a href="https://kr.iherb.com/r/sports-research-mct-oil...">https://kr.iherb.com/r/sports-research-mct-oil...</a>
6	72853	4.6	1921	<a href="https://kr.iherb.com/r/now-foods-sports-mct-oi...">https://kr.iherb.com/r/now-foods-sports-mct-oi...</a>
7	13376	4.8	3600	<a href="https://kr.iherb.com/r/now-foods-caprylic-acid...">https://kr.iherb.com/r/now-foods-caprylic-acid...</a>
10	13585	4.7	710	<a href="https://kr.iherb.com/r/now-foods-sports-mct-oi...">https://kr.iherb.com/r/now-foods-sports-mct-oi...</a>
12	68705	4.6	1563	<a href="https://kr.iherb.com/r/jarrow-formulas-mct-oil...">https://kr.iherb.com/r/jarrow-formulas-mct-oil...</a>

데이터길이: 8

## 아이템 8개로 필터링됨

### User\_id/사용자 개인 고객후기링크 수집

```
%%time
# 수집할 글 갯수 정하기
user_list=[]
rev_count_list=[]

# 수집한 url 돌면서 데이터 수집
for k in tqdm_notebook(range(0,8)):
    # 리뷰창 띄우기
    url=df_load.리뷰url[k]
    driver.get(url)
    time.sleep(2)

    for j in range(0,20):
        for i in range(0, 10):
            # 유저아이디링크따기
            user_id = ".customer-profile-link"
            user=driver.find_elements_by_css_selector(user_id)
            userurl=user[i].get_attribute('href')
            user_list.append(userurl)

            #유저리뷰수 따기
            rev_count=driver.find_elements_by_css_selector(".snapshot-count")
            count= rev_count[i].text
            rev_count_list.append(count)

        if i == 9:
            pg_nx= driver.find_element_by_css_selector(".paging-btn-next")
            driver.execute_script("arguments[0].click();", pg_nx)
            time.sleep(2)
    print(df_load.item_id[k], len(user_list))
```

```
df_user=pd.DataFrame({"고객링크":user_list,"리뷰수":rev_count_list})
df_user
```

	고객링크	리뷰수
0	https://kr.iherb.com/me/aminochan	275
1	https://kr.iherb.com/me/4774927486492925388	6
2	https://kr.iherb.com/me/4811466204621279669	0
3	https://kr.iherb.com/me/4680214676756300470	513
4	https://kr.iherb.com/me/4923014602873352856	8
...	...	...
1595	https://kr.iherb.com/me/C2CE0F4095	0
1596	https://kr.iherb.com/me/81A6D19053	0
1597	https://kr.iherb.com/me/4159E88F23	0
1598	https://kr.iherb.com/me/DD9B1D0957	6
1599	https://kr.iherb.com/me/998303BB49	0

1600 rows × 2 columns

## 아이템당 200명 사용자/리뷰수 크롤링

#리뷰수 10건이상 필터링

```
df_user_10 = df_user.loc[(df_user.리뷰수 >=10), ['고객링크', '리뷰수']]  
df_user_10
```

	고객링크	리뷰수
0	https://kr.iherb.com/me/aminochan	275
3	https://kr.iherb.com/me/4680214676756300470	513
8	https://kr.iherb.com/me/A22434FB52	16
10	https://kr.iherb.com/me/4883425644413566162	84
11	https://kr.iherb.com/me/5200700960892458955	32
...	...	...
1578	https://kr.iherb.com/me/5426883134385690579	10
1580	https://kr.iherb.com/me/D846920492	93
1583	https://kr.iherb.com/me/5614149106978046441	36
1588	https://kr.iherb.com/me/915515DB69	44
1594	https://kr.iherb.com/me/73E7A11C51	23

559 rows × 2 columns

## User\_id/item/rating 수집

```
%%time
# 수집할 글 갯수 정하기
rev_item_list=[]
rev_item_rating_list=[]
userid_list=[]

# 수집한 url 돌면서 데이터 수집
for k in tqdm_notebook(range(0, len(df_load1))):

    # 리뷰창 띄우기
    url=df_load1.고객링크[k]
    driver.get(url)
    time.sleep(2)

    try:
        for j in range(0,100):
            for i in range(0, 10):
                # 글 띄우기

                rev_item=driver.find_elements_by_css_selector(".product-review-name")
                element1 =rev_item[i].text
                rev_item_list.append(element1)

                rev_item_rating=driver.find_elements_by_css_selector("div.col-xs-14.review-content > div > div.rating > a>svg>use")
                element2 =rev_item_rating[i].get_attribute("xlink:href")
                rev_item_rating_list.append(element2)

                userid_list.append(k)
                print(element2)

            if i == 9:
                pg_nx= driver.find_element_by_css_selector(".pagination-right")
                driver.execute_script("arguments[0].click();", pg_nx)
                time.sleep(2)

    except:
        print(k)
```

## User\_id/item/rating 수집

	userid	item	rating
0	0	Sierra Fit,프로틴 크리스프, 솔티드 토피, 12개입, 각 56g(1.98oz)	5.0
1	0	Sierra Fit,프로틴 크리스프, 땅콩버터, 바 12개입, 각 56g(1.98oz)	5.0
2	0	Sierra Fit,프로틴 크리스프, 초콜릿 칩 쿠키 도우, 12개입, 각 56g(...	3.0
3	0	Sierra Fit,유청 단백질 컴플리트, 바닐라, 2.27kg(5lb)	5.0
4	0	Bob's Red Mill,Organic Quick Cooking Rolled Oa...	5.0
...	...	...	...
31591	574	NOW Foods,비타민D-3, 50mcg(2,000IU), 소프트젤 120정	5.0
31592	574	California Gold Nutrition,LactoBif 프로바이오틱스, 50...	5.0
31593	574	California Gold Nutrition,Gold C, 비타민C, 1,000m...	5.0
31594	574	Gerber,유기농 퍼프, 무화과 열매, 42g(1.48oz)	5.0
31595	574	YumEarth,오가닉 딸기 팝스, 딸기 스매시, 14개입, 87g(3.1oz)	4.0

31596 rows × 3 columns



ratings\_matrix

[illegible]



### #nan값변환

1Kvit-C,비타 민 C, 수분, 발 포 드링크 믹 스, 천연 오 렌지 향, 1,000mg, 30 팩, 각 0.22oz(6.20g)	1Kvit-C,비타 민 C, 집중력, 발포 드링크 믹스, 천연 오 렌지 향, 1,000mg, 30 팩, 각 0.23oz(6.40g)	1Kvit-C,비타 민C, 에너지, 발포 드링크 믹스, 천연 오 렌지 맛, 1,000mg, 30 팩, 각 0.24oz(6.80g)	1Kvit-C,비타 민C, 에센셜, 발포 드링크 믹스, 천연 오 렌지 맛, 1,000mg, 30 팩, 각 0.24oz(6.90g)	1Kvit-C,비타 민C, 필수, 발 포 드링크 믹 스, 천연 오 렌지 향, 1,000mg, 3 팩, 각 0.24oz(6.90g)
item				

21st  
Century, 600+D3  
플러스 미네랄,  
120정

21st  
Century, 600+D3,  
칼슘 & 비타민D3  
보충제, 75정

21st  
Century, Arthri-  
Flex 어드밴티  
지 + 비타민D3,  
장용성 코팅 정  
제 180정

21st  
Century, Arthri-  
Flex 어드밴티  
지 + 비타민D3,  
코팅 정제 120  
정

21st  
Century, B-  
100 복합  
체, 서방정, ...  
타블렛 60  
정

userid

[illegible]



```
# 아이핀-사용자 행렬로 transpose 한다.
ratings_matrix_T = ratings_matrix.transpose() # 전치 행렬

print(ratings_matrix_T.shape)
ratings_matrix_T.head(3)
```

(11682, 531)

[illegible]



## 유사도 벡터

```
from sklearn.metrics.pairwise import cosine_similarity

item_sim = cosine_similarity(ratings_matrix_T, ratings_matrix_T)

# cosine_similarity() 로 반환된 넘파이 행렬을 영화명을 매핑하여 DataFrame으로 변환
item_sim_df = pd.DataFrame(data=item_sim, index=ratings_matrix.columns,
                           columns=ratings_matrix.columns)

print(item_sim_df.shape)
item_sim_df.head(3)

(11682, 11682)
```

item	1Kvit-C,비타민 C, 수분, 발포 드링크 믹스, 천연 오렌지 향, 1,000mg, 30 팩, 각 0.22oz(6.20g)	1Kvit-C,비타민 C, 집중력, 발포 드링크 믹스, 천연 오렌지 향, 1,000mg, 30 팩, 각 0.23oz(6.40g)	1Kvit-C,비타민 C, 에너지, 발포 드링크 믹스, 천연 오렌지 향, 1,000mg, 30 팩, 각 0.24oz(6.80g)	1Kvit-C,비타민 C, 에센셜, 발포 드링크 믹스, 천연 오렌지 향, 1,000mg, 30 팩, 각 0.24oz(6.90g)	1Kvit-C,비타민 C, 필수, 발포 드링크 믹스, 천연 오렌지 향, 1,000mg, 30 팩, 각 0.24oz(6.90g)	21st Century,600+D3 플러스 미네랄, 120정	21st Century,600+D3, 칼슘 & 비타민D3 보충제, 75정	21st Century,Arthritis Flex 어드밴티지 + 비타민D3, 장용성 코팅 정제 180정	21st Century,Arthritis Flex 어드밴티지 + 비타민D3, 코팅 정제 120정	21st Century,100% 체, 서빙 타블렛
1Kvit-C,비타민 C, 수분, 발포 드링크 믹스, 천연 오렌지 향, 1,000mg, 30 팩, 각 0.22oz(6.20g)	1.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
1Kvit-C,비타민 C, 집중력, 발포 드링크 믹스, 천연 오렌지 향, 1,000mg, 30 팩, 각 0.23oz(6.40g)	0.0	1.0	1.0	0.312348	0.0	0.0	0.0	0.0	0.0	0.0

### 예측평점 = 평점벡터 dot 유사도벡터

```
# 평점 벡터(행 벡터)와 유사도 벡터(열 벡터)를 내적(dot)해서 예측 평점을 계산하는 함수 정의
def predict_rating(ratings_arr, item_sim_arr):
    ratings_pred = ratings_arr.dot(item_sim_arr) / np.array([np.abs(item_sim_arr).sum(axis=1)])
    return ratings_pred
```

```
ratings_pred = predict_rating(ratings_matrix.values , item_sim_df.values)
ratings_pred
```

```
array([[0.18184704, 3.475089 , 3.475089 , ..., 0.18184704, 0.10432812,
        0.10432812],
       [0.         , 0.         , 0.         , ..., 0.         , 0.00566674,
        0.00566674],
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       ...,
       [0.         , 0.00779714, 0.00779714, ..., 0.         , 0.00777344,
        0.00777344],
       [0.02553839, 0.00982678, 0.00982678, ..., 0.02553839, 0.01640422,
        0.01640422],
       [0.416065  , 0.05482336, 0.05482336, ..., 0.416065  , 0.37277881,
        0.37277881]])
```

예측평점 = 평점벡터 dot 유사도벡터

item	1Kvit-C,비타 민 C, 수분, 발 포 드링크 믹 스, 천연 오렌 지 향, 1,000mg, 30 팩, 각 0.22oz(6.20g)	1Kvit-C,비타 민 C, 집중력, 발포 드링크 믹스, 천연 오 렌지 향, 1,000mg, 30 팩, 각 0.23oz(6.40g)	1Kvit-C,비타 민C, 에너지, 발포 드링크 믹스, 천연 오 렌지 맛, 1,000mg, 30 팩, 각 6.80g(0.24oz)	1Kvit-C,비타 민C, 에센셜, 발포 드링크 믹스, 천연 오 렌지 맛, 1,000mg, 30 팩, 각 6.90g(0.24oz)	1Kvit-C,비타 민C, 필수, 발 포 드링크 믹 스, 천연 오렌 지 향, 1,000mg, 3 팩, 각 6.90g(0.24oz)	21st Century,600+D3 플러스 미네랄, 120정	21st Century,600+D3, 칼슘 & 비타민D3 보충제, 75정	21st Century,Arthri- Flex 어드밴티 지 + 비타민D3, 장용성 코팅 정 제 180정	21st Century,Arthri- Flex 어드밴티 지 + 비타민D3, 코팅 정제 120 정	21st Century,B- 100 복합 체, 서방정, 타블렛 60 정	...
userid											
0	0.181847	3.475089	3.475089	1.488804	0.059627	0.036277	0.071539	0.026464	0.034909	0.067972	...
1	0.000000	0.000000	0.000000	0.004157	0.003296	0.000000	0.000000	0.002689	0.000000	0.001249	...
2	0.000000	0.000000	0.000000	0.000000	0.001799	0.000000	0.000000	0.005769	0.000000	0.004842	...
3	0.045703	0.022471	0.022471	0.044205	0.039666	0.021131	0.053514	0.001510	0.054467	0.035294	...
4	0.033064	0.007080	0.007080	0.007412	0.013222	0.045533	0.014353	0.000000	0.012875	0.014211	...
5	0.041330	0.006743	0.006743	0.003515	0.002179	0.016566	0.006750	0.001842	0.000000	0.002629	...
6	0.033064	0.005394	0.005394	0.005834	0.005117	0.000000	0.012842	0.000000	0.016200	0.001239	...
8	0.036677	0.018866	0.018866	0.038263	0.019720	0.047213	0.021464	0.004082	0.020164	0.009913	...
9	0.036264	0.026785	0.026785	0.010701	0.003604	0.000000	0.005923	0.000000	0.000000	0.008704	...
10	0.000000	0.000000	0.000000	0.005942	0.005752	0.009325	0.018207	0.000000	0.016758	0.001218	...

### Top-n 유사도 예측평점

```
def predict_rating_topsim(ratings_arr, item_sim_arr, n=20):
    # 사용자-아이템 평점 행렬 크기만큼 0으로 채운 예측 행렬 초기화
    pred = np.zeros(ratings_arr.shape)

    # 사용자-아이템 평점 행렬의 열 크기만큼 Loop 수행.
    for col in range(ratings_arr.shape[1]):
        # 유사도 행렬에서 유사도가 큰 순으로 n개 데이터 행렬의 index 반환
        top_n_items = [np.argsort(item_sim_arr[:, col])[:-n-1:-1]]
        # 개인화된 예측 평점을 계산
        for row in range(ratings_arr.shape[0]):
            pred[row, col] = item_sim_arr[col, :][top_n_items].dot(ratings_arr[row, :][top_n_items].T)
            pred[row, col] /= np.sum(np.abs(item_sim_arr[col, :][top_n_items]))
    return pred

# 실행시간 2분 정도 걸림
ratings_pred = predict_rating_topsim(ratings_matrix.values, item_sim_df.values, n=20)
print('아이템 기반 인접 TOP-20 이웃 MSE: ', get_mse(ratings_pred, ratings_matrix.values))

# 계산된 예측 평점 데이터는 DataFrame으로 재생성
ratings_pred_matrix = pd.DataFrame(data=ratings_pred, index=ratings_matrix.index,
                                   columns=ratings_matrix.columns)
```

아이템 기반 인접 TOP-20 이웃 MSE: 7.411762916099533

### 미사용상품추천

```
# 사용자가 미사용한 상품 추천 user_rating이 0보다 크면 기존에 사용한 상품
def get_unused_item(ratings_matrix, userid):

    user_rating = ratings_matrix.loc[userid,:]
    already_used = user_rating[user_rating > 0].index.tolist()

    items_list = ratings_matrix.columns.tolist()
    unused_list = [ item for item in items_list if item not in already_used]

    return unused_list
```

### 사용자 개인화 추천시스템 userid=9

```
def recomm_item_by_userid(pred_df, userid, unused_list, top_n=10):
```

```
    recomm_items = pred_df.loc[userid, unused_list].sort_values(ascending=False)[:top_n]  
    return recomm_items
```

```
unused_list = get_unused_item(ratings_matrix, 9)
```

```
recomm_items = recomm_item_by_userid(ratings_pred_matrix, 9, unused_list, top_n=10)
```

```
recomm_items = pd.DataFrame(data=recomm_items.values, index=recomm_items.index, columns=['pred_score'])  
recomm_items
```

	pred_score
item	
NatureWise,유기농 강황 커큐민, 베지 캡슐 180정	0.318805
Annie's Naturals,유기농 케첩, 680g(24oz)	0.300946
Sir Kensington's,아보카도 오일 마요네즈, 473ml(16fl oz)	0.259400
Kitchens of India,치킨 커리 페이스트, 농축 소스, 중간, 100g(3.5oz)	0.257231
Source Naturals,Chromemate 크로뮴 GTF, 200mcg, 240정	0.256203
ChocZero,Maple Syrup, Sugar Free, 12 oz	0.249577
Quest Nutrition,토르티야 스타일 단백질 칩, 나초 치즈, 12봉, 각 32g(1.1oz)	0.246900
Solaray,칼슘 마그네슘 2:1 비율, 베지 캡슐 180정	0.243209
NOW Foods,영양 효모 플레이크, 4.5 oz(128 g)	0.242032
Rainbow Light,구미젤리 비타민C 슬라이스, 특 쏘는 오렌지 맛, 구미젤리 90개	0.236703

### 사용자 개인화 추천시스템 userid=345

```
unused_list = get_unused_item(ratings_matrix, 345)
recomm_items = recomm_item_by_userid(ratings_pred_matrix, 345, unused_list, top_n=15)
recomm_items = pd.DataFrame(data=recomm_items.values, index=recomm_items.index, columns=['pred_score'])
recomm_items
```

	pred_score
item	
Nature's Bounty,멜라토닌, 1 밀리그램, 180 정	0.332698
Nuby,Teethe Eez, 수딩 티더, 3개월 이상, 블루	0.332698
Natrol,글루코사민, 히알루론 산 & MSM , 90 캡슐	0.332698
Jarrow Formulas,PQQ, 10mg, 캡슐 30정	0.332698
NOW Foods,호박씨오일, 1,000mg, 소프트젤 200정	0.275969
California Gold Nutrition,비타민D3, 125mcg(5,000IU), 피쉬 젤라틴 소프트젤 90정	0.267296
Solgar,비타민D3가 함유된 액상 칼슘 마그네슘 시트레이트, 천연 블루베리 맛, 473ml(16fl oz)	0.266033
California Gold Nutrition,CollagenUP, 저분자 해양 콜라겐 + 히알루론산 + 비타민C, 무맛, 206g(7.26oz)	0.259916
Naturade,Total Soy, 체중 감량 셰이크, 초콜릿, 540g(1.2lbs)	0.255241
Thayers,위치하젤, 알로에 베라 포뮬러, 무알코올 토너, 장미 꽃잎, 12 fl oz(355 ml)	0.247764
California Gold Nutrition,어린이용 액상 Gold C, 비타민C, USP 등급, 천연 오렌지 향, 118ml(4fl oz)	0.245653
Nature's Plus,Source of Life, Animal Parade, 어린이용 이문 부스터, 천연 열대 베리 맛, 90개	0.243734
ChildLife,에센셜스, 에키네시아, 천연 오렌지향, 30ml(1fl oz)	0.239641
NOW Foods,엘더베리 아연-C, 사탕 정제 60정	0.238852
Stridex,Single-Step Acne Control, 맥시멈, 알코올 무함유, 소프트 터치 패드 90매	0.238243