

# 사용자 맞춤 건강 식재료 변환 플랫폼

데이터마이닝팀 | 오주원 이동기 김형석 이경미 최종혁

# 목차 소개

Table of Contents

01

주제 선정 배경

02

데이터 수집

03

데이터 전처리

04

추천 시스템

05

시스템 구현

06

의의 및 한계





---

## 01 주제 선정 배경

---

# 01 주제 선정 배경

## 주제 설명

### < 사용자 맞춤 건강 식재료 변환 플랫폼 >

1

사용자가 먹고자 하는 음식을 입력

2

음식의 식재료들 중 개인의 특성에 따라  
' 좋지 않은 영양성분 ' 이 있는 식재료 파악  
or  
전체 권장 섭취량에 걸맞게 재조합

3

추천 시스템을 활용해  
해당 음식을 건강하게 먹을 수 있는  
건강 식재료 추천

# 01 주제 선정 배경

## 2030 세대의 식습관



### 화려하고 자극적인 음식의 홍수

학업이나 취업, 직장, 사회생활 등 다양한 스트레스 요인에 직면한 젊은 세대는 **스트레스**를 빠르게 해소하려고  
**자극적인 음식**을 더 많이 먹고, 짠맛, 단맛, 매운맛 등 **강렬한 맛과 향**을 선호하는 경향을 보임  
SNS 상에서 퍼지는 잘못된 푸드 트렌드 혹은 콘텐츠는 유행에 민감한 20~30대에게 잘못된 식습관을 유발할 수 있음  
ex. 균형과 거리가 먼 식단을 자주 노출하거나 과도한 양을 섭취하는 모습을 보여주는 콘텐츠

### 혼자 밥 먹는 문화도 식문화에 영향

1인 가구의 비율 중 20, 30대가 차지하는 비중이 비교적 높는데, 1인 가구는 가공식품 섭취율과 외식 비율이 높고, 혼자 식사하는 경우가 많음  
이런 경우 **채소 섭취량은 적고 음료와 주류 섭취는 높은** 것으로 나타나, 영양소가 골고루 포함된 식품을 섭취할 가능성이 낮고  
**균형 잡힌 영양 섭취로 이어지기 어려움**

# 01 주제 선정 배경

## 2030 세대의 식습관



편리하고 빠르게 식사하려는 경향

바쁜 현대사회에서 무엇이든 급히 하려는 라이프스타일 속에서 여유 있게 균형 잡힌 식사를 누리기 어려워

**빨리 먹을 수 있는 음식과 외식**에 지나치게 의존하는 등 건강하지 못한 식습관이 흔하고 익숙해짐

게다가 외식 위주 식사와 빠른 식사 시에는 탄수화물을 많이 섭취하게 되고 단백질과 채소·과일 섭취는 부족한 경우가 많아

**영양불균형**을 초래하기 쉬움

이러한 **잘못된 식습관**을 오랫동안 유지하면 신체에 필요한 영양소를 충분히 섭취하지 못해 신체기능이 저하되고 **질병 발생 위험**이 커짐  
가공식품 위주의 식사와 지방과 콜레스테롤, 나트륨 함량이 높은 식품의 섭취는 **비만, 당뇨병, 심혈관질환, 고혈압** 등을 유발할 수 있음

# 01 주제 선정 배경

## 젊은 당뇨의 증가

“우리나라 당뇨병 인구가 600만 명을 넘어선 가운데, 2030 세대의 당뇨병 유병률이 가파르게 증가하고 있다. 건강보험심사평가원 통계에 따르면 2020년 기준 30대 당뇨병 환자는 12만 1568명으로 4년 전보다 25.5% 늘었다(‘16년 9만 6891명). 같은 기간 20대 유병률은 약 47% 늘어 심각한 증가세를 보였다 (‘16년 2만 3798명→ ‘20년 3만 5005명).”

노용석. (2023.11.08). 2030 세대 60%, 자신 혈당 수치 몰라. 성인병 뉴스.  
[http://www.cdpnews.co.kr/renewal/cdpnews\\_read.asp?menuID=2&idx=48275](http://www.cdpnews.co.kr/renewal/cdpnews_read.asp?menuID=2&idx=48275)

12일 강릉아산병원에 따르면 지난 11월 발표한 건강보험심사평가원 자료를 분석한 결과 2018년 대비 2022년 2030세대 당뇨병 환자 증가율은 24%로 전체 당뇨병 환자 증가율 21%보다 높았다. 특히 20대 당뇨병 환자는 47%나 증가했다.

전영래. (2023.12.12). 늘어나는 2030세대 '젊은 당뇨'...모른 채 지나치면 '큰 일'  
노컷 뉴스. <https://www.nocutnews.co.kr/news/6061985>

# 01 주제 선정 배경

## 비만과 고혈압 유병률 증가

대한비만학회가 발간한 '2023 비만 팩트 시트(2023 Obesity Fact Sheet)'에 따르면 20세 이상 성인의 **비만 유병률**은 38.4%(2021년 기준)로 10년 전(30.2%)보다 **8.2%포인트** 증가했다.

채소·과일 섭취율은 권장량에 미치지 못하고, 배달 서비스 활성화로 안주 섭취 증가, 음료 소비 증가 등으로 비만 인구가 크게 증가

권대익. (2024.03.19). 점점 뚱뚱해지는 20~40대... 고도 비만, 10년 새 3배 증가. 한국일보. <http://dc.koreatimes.com/article/20240318/1506668>

건강보험심사평가원 보건의료빅데이터 통계자료를 분석한 결과, 20세부터 39세까지 중 **고혈압으로 진료를 받은 환자**는 2017년 19만5767명에서 2021년 25만2938명으로 **29.2%나** 증가한 것으로 나타났다.

비만과 고혈압은 연관성이 매우 높아, 비만을 증가는 고혈압 발생 확률을 증가시킴 또한 장기적인 코로나19 엔데믹과 취업난 등으로 인한 정신적 스트레스가 젊은 고혈압을 증가시키는 것으로 보임

윤병기. (2024.03.19). "20~30대 MZ세대 '고혈압' 증가 추세". 후생신보. <https://www.whosaeng.com/140683>





## 비만과 고혈압 유병률 증가

## 식습관과 질병

대한비만학회가 발간한 '2023 비만 팩트 시트(2023 Obesity Fact Sheet)'에 따르면 20세 이상 성인의 비만 유병률은 38.4%(2021년 기준)로 10년 전(30.2%)보다 8.2%포인트 증가했다.

\*HEI (Healthy Eating Index) 와 고혈압, 당뇨와의 연관성 분석 실험에서 총 HEI 점수와 제 2형 당뇨, 고혈압 유병률과의 관련성이 통계적으로 유의하게 증명됨  
질병 군 남녀 모두 정상 군에 비해 통계적으로 유의하게 낮은 HEI 점수를 나타냄

건강보험심사평가원 보건의료빅데이터 통계자료를 분석한 결과, 20세부터 39세까지 중 고혈압으로 진료를 받은 환자는 2017년 19만5767명에서 2021년 25만2938명으로 29.2%나 증가한 것으로 나타났다.

→ 이 결과는 정상 군에 비해 당뇨, 고혈압 군이 상대적으로 식사의 질이 낮다는 것을 의미

채소·과일 섭취율은 권장량에 미치지 못하고, 배달 서비스 활성화로 안주 섭취 증가, 음료 소비 증가 등으로 비만 인구가 크게 증가

식습관과 질병 간의 상관관계가 존재함!

\*HEI: 인구 집단 내 식사의 질을 평가할 수 있는 지표  
비만과 고혈압은 연관성이 매우 높으며, 비만을 증가는 고혈압 발생 확률을 증가시킴 또한 장기적인 코로나19 엔데믹과 취업난 등으로 인한 정신적 스트레스가 젊은 고혈압을 증가시키는 것으로 보임

# 01 주제 선정 배경

## 최근 푸드 트렌드



2015년 이후 국내 건강기능식품 시장은 연평균 13.7% 성장했고,  
이에 따라 건강기능식품 수입량도 지난 6년간 85% 늘어나는 등 지속적인 증가세를 보임

사회 소득이 늘고 고령 인구가 지속 증가하면서  
삶의 질과 건강한 식생활에 대한 관심도 커짐에 따라 국내 건강기능식품 시장이 지속적인 증가세를 보임

건강기능식품 중 비타민, 무기질, 단백질 등 여러 영양소로 구성된 복합 영양제품과  
칼로리가 낮고 설탕의 단맛을 대체할 수 있는 감미료 수요도 증가함

# 01 주제 선정 배경

## 최근 푸드 트렌드

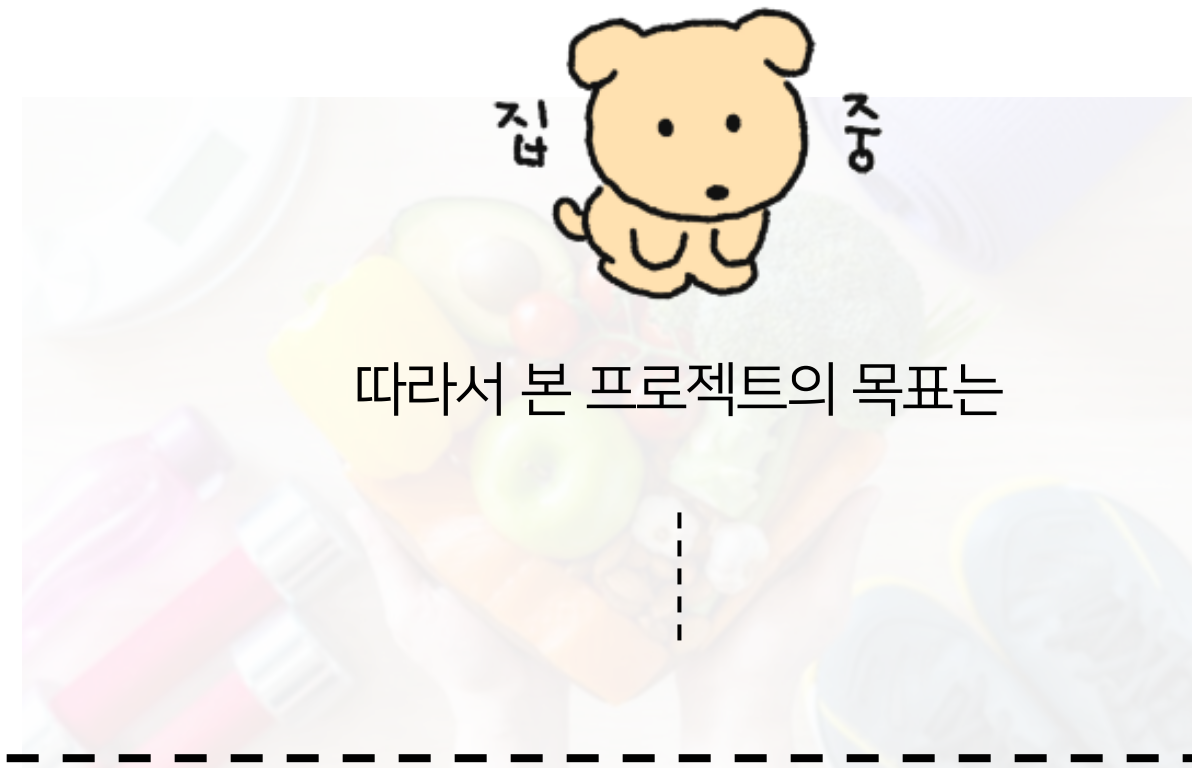


사진 출처 : pixabay

사용자가 건강한 식습관을 가질 수 있게끔 하기 위해  
개인의 특성을 반영하여 기존의 식재료들을 더 건강한 재료로 변환해 주는 것



건강기능식품 중 비타민, 무기질, 단백질 등 여러 영양소로 구성된 복합 영양제품과  
칼로리가 낮고 설탕의 단맛을 대체할 수 있는 감미료 수요도 증가함

사용자 맞춤 건강 식재료 변환 플랫폼을 구축하자!



---

## 02 데이터 수집

---

# 02 데이터 수집

## 1) 전국통합식품영양성분정보(원재료성식품) 표준데이터

식품코드	식품명	식품대분류명	...	단백질(g)	지방(g)	...	칼륨(mg)	나트륨(mg)	...	제공기관명
R101-003000101-0000	기장_도정_생것	곡류		11.2	1.9		233	6		농촌진흥청 국립농업과학원
R102-006000001-0000	고구마_생것	감자 및 전분류		1.09	0.15		375	8		농촌진흥청 국립농업과학원
R108-043070001-0000	복숭아_천도_생것	과일류		0.93	0.05		231	0		농촌진흥청 국립농업과학원
...										
R211-509073901-0457	전복류_전복_전체_생것_완도_3월	어패류 및 기타 수산물		13.68	1.61					해양수산부 국립수산물과학원

해양수산부 (국립수산물과학원), 농촌진흥청 (국립농업과학원)에서 제공하는 원재료성 영양성분 공공데이터  
식품명, 식품대분류명, 각 식품에 해당하는 영양성분 등의 정보 제공



# 02 데이터 수집


## 2) 국가표준식품성분 DB

식품군	식품명	에너지_kcal	수분_g	단백질_g	지방_g	...	오메가3 지방산_g	오메가6 지방산_g	총 트랜스지방산_g
곡류 및 그 제품	귀리, 겉귀리, 도정, 생것	380	11.5	11.4	7.35		0.05	2.5	0
곡류 및 그 제품	귀리, 쌀귀리, 도정, 생것	388	11.6	11.14	8.9		0.07	2.5	0.01
채소류	당근, 뿌리, 데친것	31	91.2	0.97	0.13		0.01	0.09	0
...									
기타	인삼, 홍삼, 말린것	309	13.4	13.02	1.24		0.02	0.37	0.02

농촌진흥청 (국립농업과학원)에서 제공하는 국가표준 식품성분표  
식품군, 식품명, 각 식품에 해당하는 더욱 다양한 영양성분 등의 정보 제공

# 02 데이터 수집

## 3) 식약처 레시피 API 활용

서비스	서비스명	조리식품의 레시피 DB	서비스유형	OPEN API
	최종수정일	2021-08-03	업데이트주기	상시
	최초개방일	2017-12-19	API호출제한	1000
제공기관	기관	식품의약품안전처	분류	식품영양정보
설명	속성정보	일련번호, 메뉴명, 조리방법, 요리종류, 중량(1인분), 열량, 탄수화물, 단백질, 지방, 나트륨, 해쉬태그, 이미지경로(소), 이미지경로(대), 재료정보, 만드는법_01, 만드는법_이미지_01, 만드는법_02, 만드는법_이미지_02, 만드는법_03, 만드는법_이미지_03, 만드는법_04, 만드는법_이미지_04, 만드는법_05, 만드는법_이미지_05, 만드는법_06, 만드는법_이미지_06, 만드는법_07, 만드는법_이미지_07, 만드는법_08, 만드는법_이미지_08, 만드는법_09, 만드는법_이미지_09, 만드는법_10, 만드는법_이미지_10, 만드는법_11, 만드는법_이미지_11, 만드는법_12, 만드는법_이미지_12, 만드는법_13, 만드는법_이미지_13, 만드는법_14, 만드는법_이미지_14, 만드는법_15, 만드는법_이미지_15, 만드는법_16, 만드는법_이미지_16, 만드는법_17, 만드는법_이미지_17, 만드는법_18, 만드는법_이미지_18, 만드는법_19, 만드는법_이미지_19, 만드는법_20, 만드는법_이미지_20, 저감 조리법 TIP		
	비고			
이용허락범위		 출처표시/상업적, 비상업적 이용가능/변형 등 2차적 저작물 작성 가능		

식품의약품안전처에서 제공한 조리식품의 레시피 관련 데이터  
약 1134건이며, 메뉴명 / 조리명확인

# 02 데이터 수집

## 식약처 레시피 API 결과 데이터프레임

메뉴명	조리방법	요리종류	중량(1인분)	열량	탄수화물	...	재료정보	만드는법	저감 조리법 Tip
양파 토마토 스투	끓이기	국&찌개	288	36.2	4.9	...	[2인분] 마늘(4쪽), 토마토 (2개) ...	1. 물(8컵)에 육수 재료와 마늘(2쪽)을 넣어 ...	멸치 육수로 간을 맞추면...
황금팽이 비빔국수	기타	기타	176.2	102.7	6.9	...	[4인분] 미역 줄기(1팩 =300g)...	1. 미역 줄기는 흐르는 물에 여러 번 ...	천도복숭아의 단맛으로 ...
...	...	...	...	...	...	...	...	...	...
해산물 샐러드와 미나리 소스	볶기	반찬	192.8	216.9	8.6	...	[2인분] 새송이버섯(1개),방울토마토(200g), ...	1. 새송이버섯은 모양대로 ...	미나리를 사용하여 ...

메뉴명, 조리방법, 요리 종류, 중량(1인분), 열량, 탄수화물, 단백질, 지방, 나트륨, 해쉬태그, 재료정보, 만드는법, Tip 정보 수집

# 02 데이터 수집

## 4) 만개의 레시피

RCP_SNO	RCP_TTL	CKG_NM	...	CKG_MTRL_ACTO_NM	...	CKG_MTRL_CN	CKG_DODF_NM	...	CKG_TIME_NM	FIRST_REG_DT
128671	어묵김말이	어묵김말이		튀김		[재료] 어묵 2개 김밥용김 3장 당면 1웁큼 ...	2인분		60분 이내	20070402131403
128892	두부에 꼬리가 달렸어요!!	두부새우전		부침		[재료] 두부 1/2모 당근 1/2개 고추 2개 ...	3인분		30분 이내	20070402205937
128932	입안에서 툭툭툭	알밥		굽기		[재료] 밥 1+1/2공기 당근 1/4개 ...	2인분		30분 이내	20070402224355
	...					...				
139247	부들부들 보들보들 북어갈비♥	북어갈비		굽기		[재료] 북어포 1마리 찹쌀가루 1C [양념] 간장 2T ...	2인분		60분 이내	20070501000844

KADX 농식품 빅데이터 거래소에서 제공하는 만개의 레시피 크롤링 데이터 확보  
184,963개의 행으로 이루어져 있음



---

## 03 데이터 전처리

---



# 03 데이터 전처리

## 원재료 DB | 식품명 컬럼 값 형식 통일

식품대분류	식품명	에너지_kcal	수분_g	단백질_g	지방_g	...	오메가3 지방산_g	오메가6 지방산_g	총 트랜스지방산_g
어패류 및 기타 수산물	가자미류, 기름가자미 (21-24cm), 암컷, 육, 생것	82	79.6	17.36	0.24	...	NaN	NaN	NaN
어패류 및 기타 수산물	가자미류, 기름가자미, 생것	82	79.6	17.36	0.24	...	NaN	NaN	NaN
원재료성 식품성분표 데이터 셋 표기 형식 ...									

다른 출처의 데이터들을 병합하다 보니,  
같은 대상에 대해 다른 형식의 식품명이 기입 됨



엑셀 함수 및 파이썬 코드를 이용해,  
원재료성 식품성분표 데이터 셋의 표기 방법으로 형식을 통일

Ex)

가자미류, 기름가자미(21-24cm), 암컷, 육, 생것



가자미류, 기름가자미, 생것

# 03 데이터 전처리

## 원재료 DB | 이상값 대체

식품대분류	식품명	에너지_kcal	수분_g	단백질_g	지방_g	...	오메가3 지방산_g	오메가6 지방산_g	총 트랜스지방산_g
곡류 및 그 제품	귀리, 겉귀리, 도정, 생것	380	11.5	11.4	7.35	...	NaN	2.5	Tr 0
곡류 및 그 제품	귀리, 겉귀리, 도정, 생것	388	11.6	11.14	8.9	...	0.07	2.5	(Tr) 0
...									
기타	인삼, 홍삼, 말린것	309	NaN	13.02	1.24	...	tr 0	0.37	0.02

데이터 셋에 적힌 설명에 따르면  
'Tr, (Tr), tr' 은 극소량을 의미하는 문자열



0 으로 대체

'-' 의 경우  
결측치 값의 또 다른 표기법이라고 판단



NaN 으로 결측치 표기법 통일

# 03 데이터 전처리

## 원재료 DB | 중복값 평균 대체

식품대분류	식품명	에너지_kcal	수분_g	단백질_g	지방_g	...	오메가3 지방산_g	오메가6 지방산_g	총 트랜스지방산_g
곡류 및 그 제품	귀리, 겉귀리, 도정, 생것	380	11.5	11.4	7.35	...	NaN	2.5	0
곡류 및 그 제품	귀리, 겉귀리, 도정, 생것	388	11.6	11.14	8.9	...	0.07	2.5	0

↓ 컬럼 별 평균 처리

곡류 및 그 제품	귀리, 겉귀리, 도정, 생것	384	11.55	11.27	11.8	...	0.07	2.5	0
-----------	-----------------	-----	-------	-------	------	-----	------	-----	---

중복되는 식품명 값들의 경우, 컬럼 별로 평균값 대체를 진행해 중복을 제거

# 03 데이터 전처리

## 원재료 DB | 중복값 평균 대체

식품대분류	식품명	에너지_kcal	수분_g	단백질_g	지방_g	...	오메가3 지방산_g	오메가6 지방산_g	총 트랜스지방산_g
곡류 및 그 제품	귀리, 걸귀리, 도정, 생것	380	11.5	11.4	7.35	...	NaN	2.5	0
곡류 및 그 제품	귀리, 걸귀리, 도정, 생것	388	11.6	11.14	8.9	...	0.07	2.5	0

컬럼 별 평균 처리

곡류 및 그 제품	걸귀리	384	11.55	11.27	11.8	...	0.07	2.5	0
-----------	-----	-----	-------	-------	------	-----	------	-----	---

컬럼 내 결측치가 존재하는 경우, 결측치가 아닌 남은 값들의 평균 값으로 대체  
다만 컬럼 내 값이 모두 결측치인 경우 결측치를 그대로 작성

## 03 데이터 전처리

### 원재료 DB | 식품대분류 재정의 & 라벨링

'감자 및 전분류' + '감자류 및 전분류'  
'견과 및 종실류' + '견과류 및 종실류'  
'곡류' + '곡류 및 그 제품'  
'햄, 소시지, 육포, 베이컨'  
'과일류'  
'과일주스'  
'어패류 및 그 제품'  
'우유 및 그 제품'  
'육류 및 그 제품'



'감자류 및 전분류'  
'견과류 및 종실류'  
'곡류' + '면류'  
'가공육류'  
'과일류' + '잼류'  
'음료류'  
'어패류 및 기타 수산물'  
'유제품'  
'육류' + '가공육류'



분류의 기준이  
모호한 '식품군' 컬럼, 재정의



## 03 데이터 전처리

### 원재료 DB | 식품대분류 재정의 & 라벨링

'감자 및 전분류' + '감자류 및 전분류'  
'견과 및 종실류' + '견과류 및 종실류'  
'곡류' + '곡류 및 그 제품'  
'햄, 소시지, 육포, 베이컨'  
'과일류'  
'과일주스'  
'어패류 및 그 제품'  
'우유 및 그 제품'  
'육류 및 그 제품'



'감자류 및 전분류'  
'견과류 및 종실류'  
'곡류' + '면류'  
'가공육류'  
'과일류' + '잼류'  
'음료류'  
'어패류 및 기타 수산물'  
'유제품'  
'육류' + '가공육류'

1

유사한 식품군들,  
하나의 식품군명으로 통합

2

'과일류',  
과일류/잼류로 통합

3

'그 제품'이라는  
불특정한 이름 변경,  
이와 관련되어  
대분류 분할 혹은 통합

# 03 데이터 전처리

## 원재료 DB | 식품대분류 재정의 & 라벨링

식품군	식품명	에너지_kcal	...
곡류	귀리, 겉귀리, 도정, 생것	380	
곡류	귀리, 쌀귀리, 도정, 생것	388	
...			
기타	인삼, 홍삼, 말린것	309	



라벨	식품군	식품명	에너지_kcal	...
3	곡류	귀리, 겉귀리, 도정, 생것	380	
3	곡류	귀리, 쌀귀리, 도정, 생것	388	
...	...	...	...	...
5	기타	인삼, 홍삼, 말린것	309	

기계학습 모델이 이해할 수 있도록  
범주형 데이터를 수치형으로 변환

# 03 데이터 전처리

## 원재료 DB | 단위 변환

식품대분류	식품명	에너지_kcal	수분_g	단백질_g	칼슘_mg	...	비타민 A_μg	판토텐산_mg	총 불포화지방산_g
어패류 및 기타 수산물	연어, 훈제	216.17	62.32	23.19	11	...	11.67	1.25	8.74



식품대분류	식품명	에너지_kcal	수분_g	단백질_g	칼슘_g	...	비타민 A_g	판토텐산_g	총 불포화지방산_g
어패류 및 기타 수산물	연어, 훈제	216.17	62.32	23.19	0.011	...	0.00001166666667	0.00125	8.74

현재 데이터셋의 경우, 영양소량의 단위가 'g, mg, μg' 3가지로 구성됨  
-> 이는 추후 유사도 계산 과정에서 문제가 될 수 있기 때문에  
각 영양소 컬럼들에 대해 데이터의 단위를 모두 'g' 으로 통일시켜주는 작업 진행

## 03 데이터 전처리

### 원재료 DB | 결측치 대체

#### Gower's distance

수치형 및 범주형 컬럼들이 혼합 돼 존재하는 데이터 셋 내에서, 유사도 계산 시 적용 가능한 거리 계산 방법  
원-핫 인코딩 을 적용하는 거리측정 방식에 비해 **계산량이 적어 속도가 빠르다**는 장점이 존재



#### 범주형 컬럼 간의 계산인 경우

범주형 컬럼 내 로우 별로 범주 값의 동일 여부를 비교  
**값이 같다면 0, 다르다면 1** 의 값을 최종 거리 값에 반영



#### 수치형 컬럼 간의 계산인 경우

수치형 컬럼 내 로우 별로 맨허튼, 유클리디언 방식 등의  
거리 계산을 통해 나온 값을 최종 거리 값에 반영



두 값들의 가중치 합을 통해 최종 거리 값을 도출



### 원재료 DB | 결측치 대체 **왜 Gower's distance**를 사용했는가?

#### Gower's distance

수치형 및 범주형 컬럼들이 혼합된 데이터셋에서는 범주형 및 수치형 컬럼이 섞여 있고, 가능한 거리 계산 방법 원-핫 인코딩 식품 대분류별로 서로 공통된 영양소 값 특성을 가질 가능성이 존재  
(Ex. 같은 대분류 내 영양소 함유 비율 등)



범주형 컬럼 간의 계산인 경우

이에 유사도 계산을 기반으로 하여 결측치를 대체 해야겠다고 판단

범주형 컬럼 내 로우 별로 범주 값의 동일 여부를 비교  
데이터의 변수가 40 개가 넘는 관계로 차원 수를 늘리지 않는 선에서 빠르게 계산해 볼 수 있는 방법이 필요  
범주형 컬럼 내 로우 별로 맨허튼, 유클리디언 방식 등의  
거리 계산을 통해 나온 값을 최종 거리 값에 반영



수치형 컬럼 간의 계산인 경우

다만 유사도 계산 시 수치형 컬럼 뿐만 아니라 범주형 컬럼도 같이 반영해야 하고,

**Gower's distance** 를 거리 계산 방식으로 채택

두 값들의 가중치 합을 통해 최종 거리 값을 도출



# 03 데이터 전처리

## 원재료 DB | 결측치 대체

Main Idea

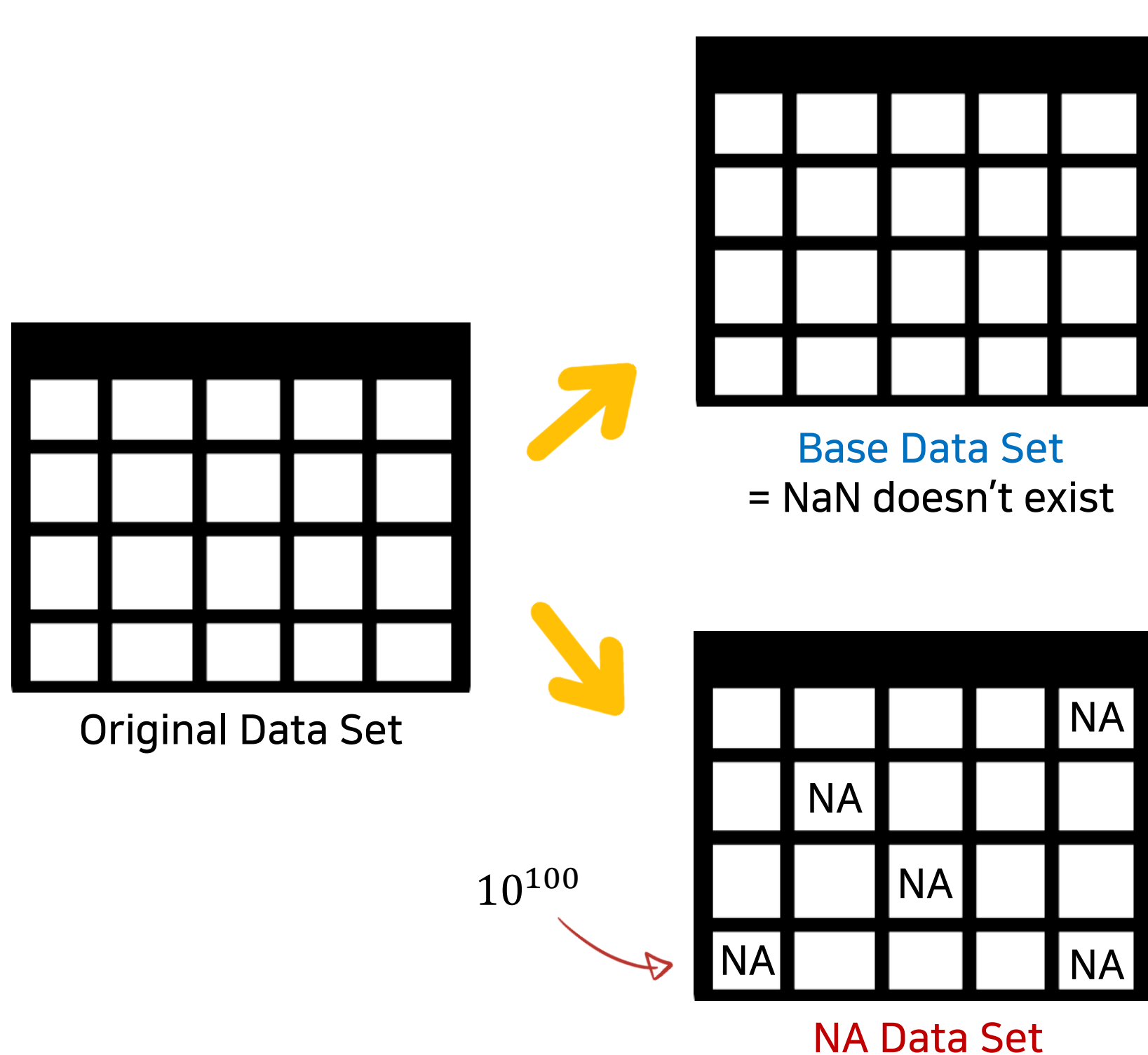
식품대분류	식품명	에너지_kcal	수분_g	...	오메가6 지방산_g
곡류 및 그 제품	귀리, 겉귀리, 도정, 생것	380	11.5	...	2.5
		a% } }	b% } }		f% } }
곡류 및 그 제품	겉귀리	384	11.55	...	2.5

유사도가 서로 높은 행들 간에는  
영양소의 함량 비율 역시 비슷할 것이라는 아이디어

## 03 데이터 전처리

### 원자료 DB | 결측치 대체

Gower's distance 기반 결측치 대체 과정



기존 데이터 셋을

결측치가 있는 컬럼이 존재하는 NA Data Set 과,  
결측치가 존재하지 않는 Base Data Set 으로 분리

Gower's distance 의 경우

결측치가 존재하면 계산 수행이 어렵기 때문에,  
유사도에 계산에 영향 주지 않을 정도의

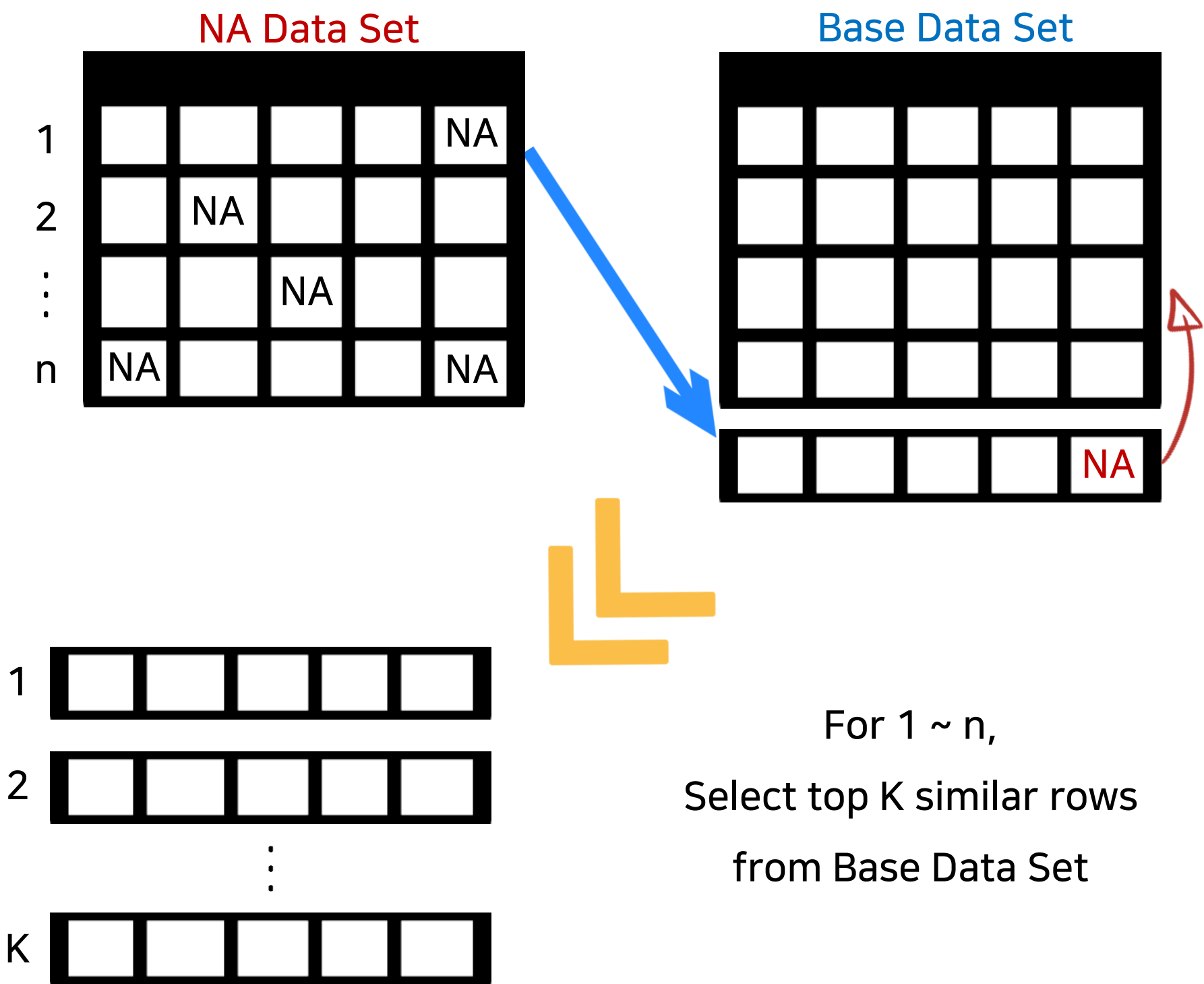
**큰 숫자로 결측치 값을 채워 줌**

(여기서는 결측치를  $10^{100}$  으로 대체)

# 03 데이터 전처리

## 원자료 DB | 결측치 대체

Gower's distance 기반 결측치 대체 과정



👉

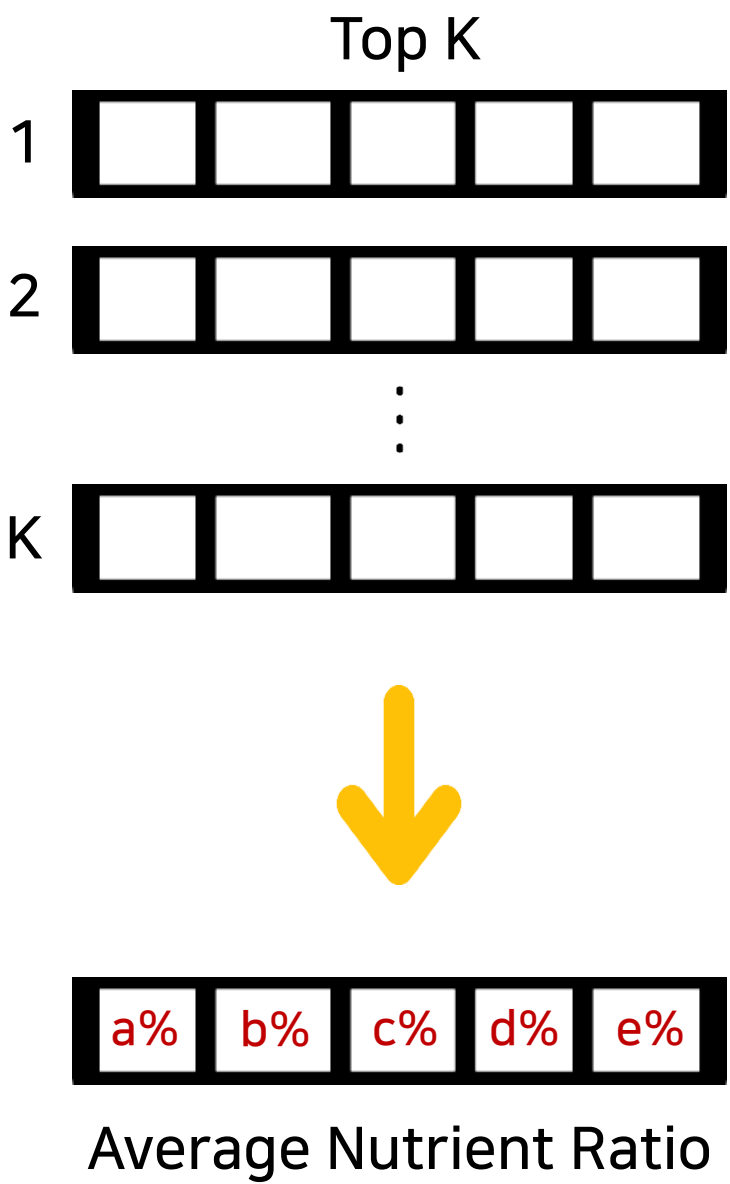
Na Data Set 의 전체 행에 대해,  
Base Data Set 에 행을 한 개씩 붙인 뒤  
Gower's distance 로 유사도 계산 작업을 진행

이후 유사도 계산 결과,  
마지막에 덧붙인 행과 유사성이 가장 높은  
K 개의 행들을 산출  
(여기서는 K 값을 Base Data Set 의 10% 정도로 지정)

# 03 데이터 전처리

## 원재료 DB | 결측치 대체

Gower's distance 기반 결측치 대체 과정

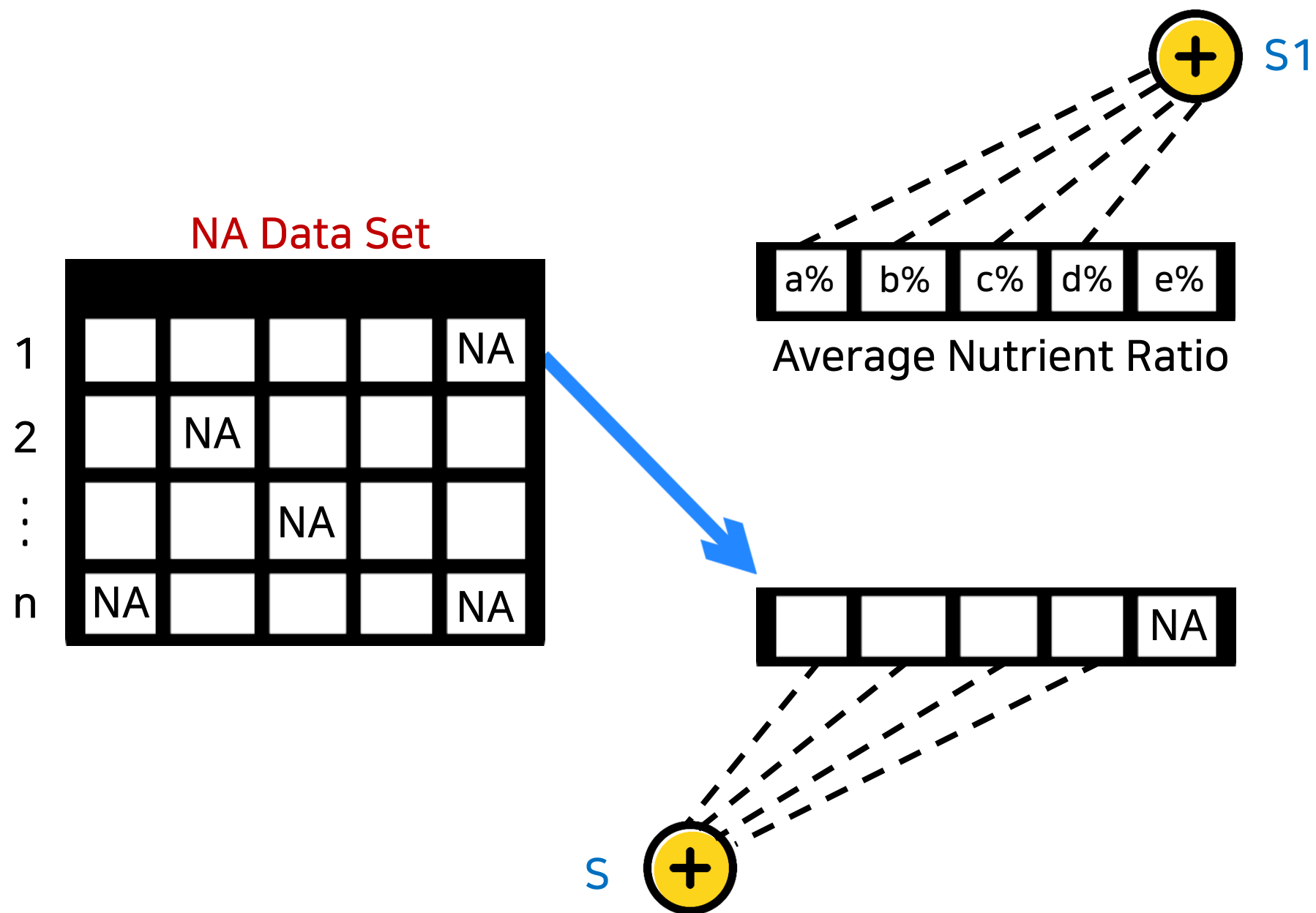


뽑아본 Top K 개의 행들에 대해  
각 영양소 컬럼 별 평균 ratio 를 계산

## 03 데이터 전처리

### 원재료 DB | 결측치 대체

Gower's distance 기반 결측치 대체 과정



Na Data Set 의 행들에 대해  
결측치가 없는 컬럼들의 인덱스를 파악하고,  
해당 인덱스 컬럼들의 영양소 값 합계인 S 를 계산

같은 방식으로 이번엔 평균 ratio 배열에 대해  
해당 인덱스 컬럼들의 평균 ratio 값 합계인 S1 을 계산

## 03 데이터 전처리

### 원재료 DB | 결측치 대체

Gower's distance 기반 결측치 대체 과정

Average Nutrient Ratio

a%	b%	c%	d%	e%
----	----	----	----	----



				NA
--	--	--	--	----

←  $(S/S1) \times e = P \times e$

Imputation



영양소 값들의 합계인 S 를  
평균 ratio 값들의 합계인 S1 으로 나눠 줌으로써,  
평균 ratio 100% 시 영양소량 값인 P 값을 도출

이후 결측치가 있는 컬럼 인덱스에  
매칭되는 ratio 를 P 에 곱한 뒤 해당 값으로  
결측치 보간 작업을 진행



## 원재료 DB | 탄단지 기반 칼로리 재계산

사람이 가장 많이 사용하는 영양소로,  
신체에 에너지를 공급하고 신체의 구조와 시스템을  
유지하기 위해 필요한 음식의 영양성분





## 03 데이터 전처리

### 원재료 DB | 탄단지 기반 칼로리 재계산

#### 미량영양소

적은 양이지만 생명유지에 필수적인 영양소로,  
다량영양소로부터 얻은 에너지를  
대사화시키는 역할을 수행



비타민



무기질



미네랄



## 03 데이터 전처리

### 원재료 DB | 탄단지 기반 칼로리 재계산

#### 다량영양소

우리 음식의 칼로리는 모두 세 가지 다량 영양소 중 하나에서 나온다는  
전문가들의 말을 토대로 탄단지의 양으로만 칼로리를 재계산하기로 결정!

How calories are calculated: The science behind your food, LIVESCIENCE, October 14, 2022

1g당 4kcal



탄수화물

1g당 4kcal



단백질

1g당 9kcal



지방

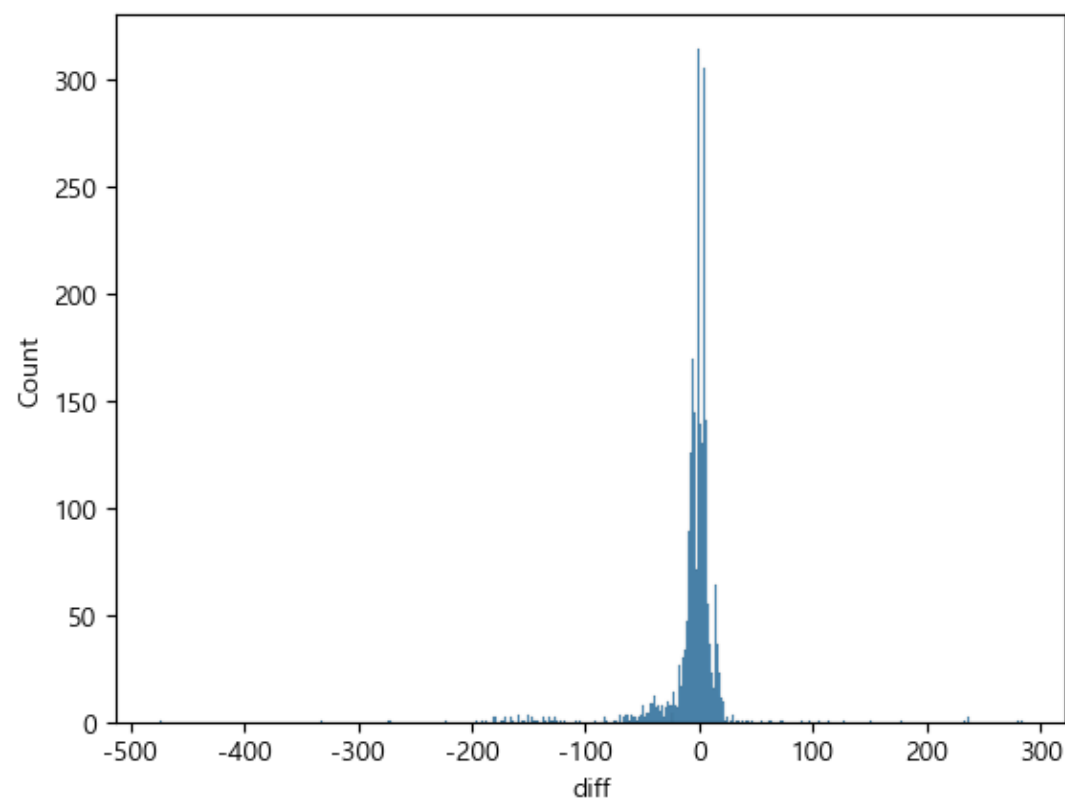
## 03 데이터 전처리

### 원재료 DB | 탄단지 기반 칼로리 재계산

탄수화물 (4kcal/g)  
단백질 (4kcal/g)  
지방 (9kcal/g)



탄단지 각각에서 1g의 오차가 발생할 수 있다는 점을 고려하여,  
원래 칼로리와 탄단지로만 계산한 칼로리의 차이가  
 $\pm(4 + 4 + 9)\text{kcal} = \pm 17\text{kcal}$  이내인 관측값만 살림



관측값 개수 확인 결과,  
2314개 -> 최종 2014개로  
총 300개의 관측값을 삭제

# 03 데이터 전처리

## 만개의 레시피 | 식재료 행 분리

RCP_SNO	CKG_NM	INQ_CNT	RCMM_CNT	SRAP_CNT	CKG_MTRL_ACTO_NM	CKG_MTRL_ACTO_NM	CKG_KND_ACTO_NM	CKG_MTRL_CN	CKG_DODF_NM
128671	어묵김말이	10072	6	66	튀김	가공식품류	디저트	어묵 2개	2인분
128671	어묵김말이	10072	6	66	튀김	가공식품류	디저트	김밥용김 3장	2인분
128671	어묵김말이	10072	6	66	튀김	가공식품류	디저트	당면 1웁큼	2인분
128671	어묵김말이	10072	6	66	튀김	가공식품류	디저트	양파 1/2개	2인분
	...							...	
128892	두부새우전	5817	3	27	부침	해물류	밑반찬	두부 1/2모	3인분
128892	두부새우전	5817	3	27	부침	해물류	밑반찬	당근 1/2개	3인분
128892	두부새우전	5817	3	27	부침	해물류	밑반찬	고추 2개	3인분
	...							...	

한 개의 레시피마다 식재료별로 분리

그 결과 약 160만 개의 row 생성!

## 03 데이터 전처리

---

만개의 레시피 | 식재료 & 단위 분리

바른 (bareun)



‘언론사를 위한 언어정보 자원 개발’ 사업의 결과물인 한국어 형태소 분석기 ‘바른(bareun.ai)’

‘바른’은 7,800만 건의 뉴스데이터와 80만 건의 언어사전을 활용해 학습시킨 결과물

# 03 데이터 전처리

## 만개의 레시피 | 식재료 & 단위 분리

바른 (bareun)

평가항목	KPF(바른)	카카오(카이)	서울대(꼬꼬마)	오픈소스(메캡)
모호성 (동음이의어) 평가	83.1%	60.1%	52.2%	46.2%
지정사(서술격 조사)	98.3%	73.7%	59.4%	1.4%
문장 전체 정확도	71.3%	14.0%	3.8%	1.9%

다른 한국어 형태소 분석기보다  
높은 성능을 보이는 '바른'



# 03 데이터 전처리

## 만개의 레시피 | 식재료 & 단위 분리

바른 (bareun)

CKG_MTRL_CN	CKG_INBUN	Year	ingredient	unit
어묵 2개	2인분	2007	어묵	2개
김밥용김 3장	2인분	2007	김밥용 김	3장
...	...	...	...	...
물 1/2컵	2인분	2023	물	1/2 컵

식재료 & 단위 분리 후 데이터프레임



# 03 데이터 전처리

## 만개의 레시피 | 단위 통일

RCP_SNO	CKG_NM	INQ_CNT	RCMM_CNT	SRAP_CNT	CKG_MTRL_ACTO_NM	Ingredient	Unit	CKG_DODF_NM	Year
331006	오징어볶음우동	25379	8	95	볶음	오징어	1마리	2인분	2007
331006	오징어볶음우동	25379	8	95	볶음	양파	1/4개	2인분	2007
331006	오징어볶음우동	25379	8	95	볶음	당근	1/3개	2인분	2007
331006	오징어볶음우동	25379	8	95	볶음	후추	약간	2인분	2007
	...					...			
331006	오징어볶음우동	25379	8	95	볶음	고추장	1T	3인분	2007
331006	오징어볶음우동	25379	8	95	볶음	다진 마늘	1T	3인분	2007
331006	오징어볶음우동	25379	8	95	볶음	설탕	1t	3인분	2007
	...					...			

전국통합식품영양성분정보(원재료성식품) 표준데이터에서 식재료를 매칭하기 위해 ml와 g으로 단위 통일

# 03 데이터 전처리

## 만개의 레시피 | 단위 통일

RCP_SN O	CKG_NM	INQ_CNT	RCM M_CN T	SRAP _CNT	CKG_MT RL_ACT O_NM	Ingredient	Unit	CKG_DOD F_NM	Year	g_possible	g	ml
331006	오징어볶음우동	25379	8	95	볶음	오징어	1마리	2인분	2007	0		0
331006	오징어볶음우동	25379	8	95	볶음	양파	1/4개	2인분	2007	0		0
331006	오징어볶음우동	25379	8	95	볶음	당근	1/3개	2인분	2007	0		0
331006	오징어볶음우동	25379	8	95	볶음	후추	약간	2인분	2007	0		0
	...					...						
331006	오징어볶음우동	25379	8	95	볶음	고추장	1T	3인분	2007	0		7.5
331006	오징어볶음우동	25379	8	95	볶음	다진 마늘	1T	3인분	2007	0		7.5
331006	오징어볶음우동	25379	8	95	볶음	설탕	1t	3인분	2007	0		2.5
	...					...						

g\_possible, g, ml 열 추가

g\_possible : 단위가 g으로 작성된 식재료면 1, ml로 작성된 식재료면 0

# 03 데이터 전처리

## 만개의 레시피 | 단위 통일

RCP_SN O	CKG_NM	...	Ingredient	Unit	CKG_DOD F_NM	...	g_possible	g	ml
331006	현미호두죽		현미	4컵	2인분		0		400
331006	현미호두죽		찹쌀	2컵	2인분		0		200
331006	현미호두죽		호두	50g	2인분		1	25	
...									
331006	현미호두죽		소금	약간	2인분		0		1.2
153040	갈릭치킨		닭	1마리	2인분		0		
153040	갈릭치킨		마요네즈	2T	2인분		0		15
...									
221101	스팸샌드위치		스팸 싱글	1개	1인분		0		

전국통합식품영양성분정보(원재료성식품) 표준데이터는 100g 기준으로 작성된 것이므로 g으로 통일해야 함

# 03 데이터 전처리

## 만개의 레시피 | 단위 통일

RCP_SN O	CKG_NM	...	Ingredient	Unit	CKG_DOD F_NM	...	g_possible	g	ml
331006	현미호두죽		현미	4컵	2인분		0		400
331006	현미호두죽		찹쌀	2컵	2인분		0		200
331006	현미호두죽		호두	50g	2인분		1	25	
...									
331006	현미호두죽		소금	약간	2인분		0		1.2
153040	갈릭치킨		닭	1마리	2인분		0		
153040	갈릭치킨		마요네즈	2T	2인분		0		15
...									
221101	스팸샌드위치		스팸 싱글	1개	1인분		0		

ml 값만 아는 경우 밀도 (g / ml) 값을 곱해 g 으로 변환 시도

# 03 데이터 전처리

## 만개의 레시피 | 단위 통일



RCP_SN O	CKG_NM	...	Ingredient	...	g_DOD _NM	...	g_possible	g	ml
331006	현미호두죽		현미	4컵	2인분		0		400
331006	현미호두죽		찹쌀	2컵	2인분		0		200
331006	현미호두죽		호두	50g	2인분		1	25	
...									
331006	현미호두죽		...						
153040	갈릭치킨		닭	1마리	2인분		0		
153040	갈릭치킨		마요네즈	...	2인분		0		15
...									
121101	스팸샌드위치		스팸 싱글	1개	1인분		0		

g 값을 모르는 경우 : 전체 데이터 160만 개 행 중 약 55%에 해당  
→ 현실적으로 약 8-90만 개의 데이터의 밀도값과 그램 수를 수집하기 어려움 ...  
**따라서 !**  
중복을 제외하고 1,000개의 레시피만을 랜덤으로 추출해서 사용하기로 결정!

가 값만 아는 경우 밀도 (g / ml) 값을 곱해 g 으로 변환 시도

# 03 데이터 전처리

원재료 DB | 변수 생성



라벨	식품군	식품명	에너지_kcal	...	...	추가변수	...
3	곡류	귀리,겉귀리, 도정,생것	380				
3	곡류	보리(기장),도 정,생것	388				
...	...	...	...	...			
5	기타	a인삼, 홍삼2, 말린것	309				

유사한 식재료 내에서 추천하기 위해서 (= 엉뚱한 식재료로 변환되지 않기 위해서)  
음식의 특성을 반영한 추가 변수를 생성하기로 결정

## 03 데이터 전처리

### 원재료 DB | 변수 생성

물성

탄수화물, 단백질, 지방에 의해 식품 물성이 정해진다

식품의 물성이란 식품의 물리학적 미각, 즉 식감과 밀접한 관련이 존재한다. 탄수화물, 단백질, 지방 등 3대 영양소는 물성에 영향을 주는 요소로 이들의 조합에 따라 물성이 달라진다. ... 이들 영양소들은 수분과 상호작용으로 식품 고유의 물성을 만들어 낼 수가 있다.

출처 : 식품저널 foodnews(<http://www.foodnews.co.kr>)



# 03 데이터 전처리

## 원재료 DB | 변수 생성

유사한 음식을 정의하기 위해 '물성' 개념을 활용하여,  
탄단지와 수분의 각 비율 변수를 추가!



라벨	식품군	식품명	에너지_kcal	탄수화물 비율	...	수분 비율
3	곡류	귀리겉가리도	380			
3	곡류	보리도정	388			
...	...	...	...			
5	기타	인삼, 홍삼, 말린것	309			

추가적으로 탄단지 각 비율, 수분 비율, 4변수에 대해서  
PCA를 적용함

물성을 이루는 4가지 변수들 중에서 수분을 제외한 매크로 텀들은  
서로 다중공선성이 클 것이라 판단함!

유사한 음식을 선별하는 기준 중 하나로써, 코사인 유사도 계산시 PCA의 차이가 발생함  
보다 현실적인 기준에서 유사한 음식을 정의 하기 위해, 추가 변수를 생성하기로 결정

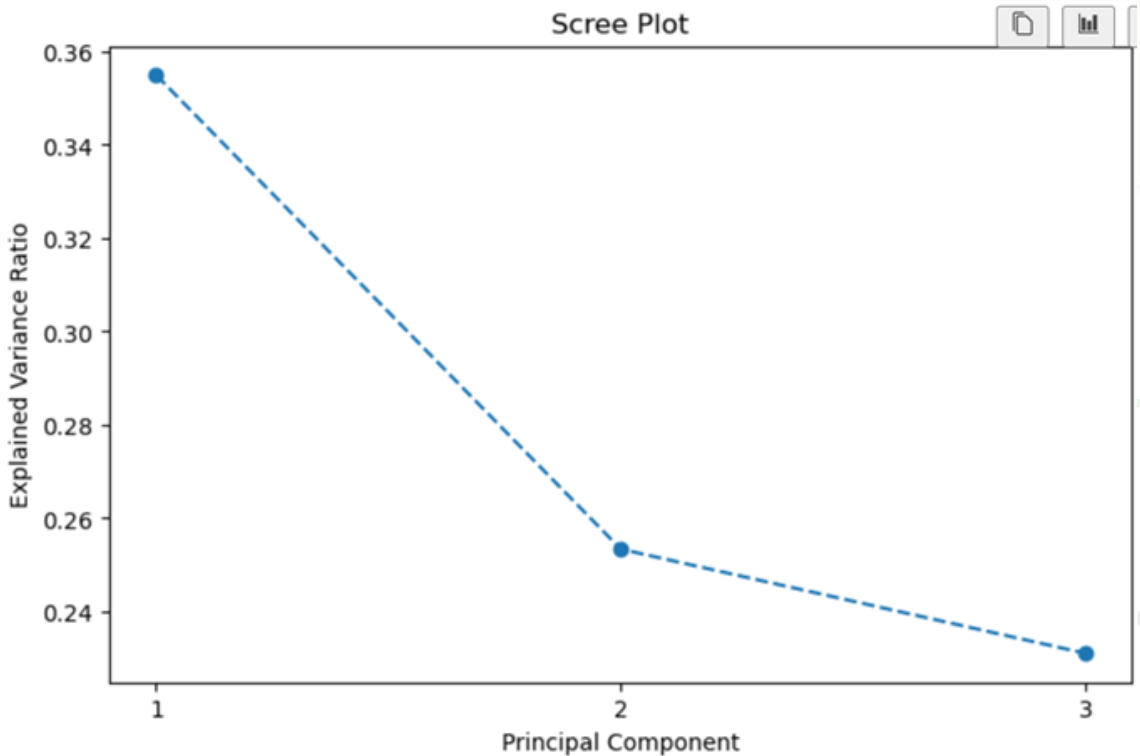
# 03 데이터 전처리

## 원재료 DB | 변수 생성



라벨	식품군	식품명	에너지_kcal	탄수화물 비율	...	수분 비율
3	곡류	귀리겉가리도	380			
3	곡류	보리도정	388			
...	...	...	...			

추가적으로 탄단지 각 비율, 수분 비율, 4변수에 대해서 PCA를 적용함



Scree plot 상, 최적의 주성분 개수는 2개로, 수분, 탄수화물, 단백질, 지방 순으로

0.154	0.401	0.664	0.609
0.954	-0.051	0.069	-0.285

수분 vs 탄단지로 잘 구분됨을 확인 할 수 있었음



물성 변수를 2차원으로  
차원 축소한 데이터 활용 !

# 03 데이터 전처리

## 식품명 임베딩

CKG_MTRL_CN	CKG_INBUN	Year	ingredient	unit
어묵 2개	2인분	2007	어묵	2
김밥용김 3장	2인분	2007	김밥용 김	3
...	...	...	...	...
물 1/2컵	2인분	2023	물	1/2

<만개의 레시피 상 원재료>



수분_g	단백질_g	지방_g	...
11.5	11.4	7.35	...
11.6	11.14	8.9	...
...	...	...	...
150	0	0	

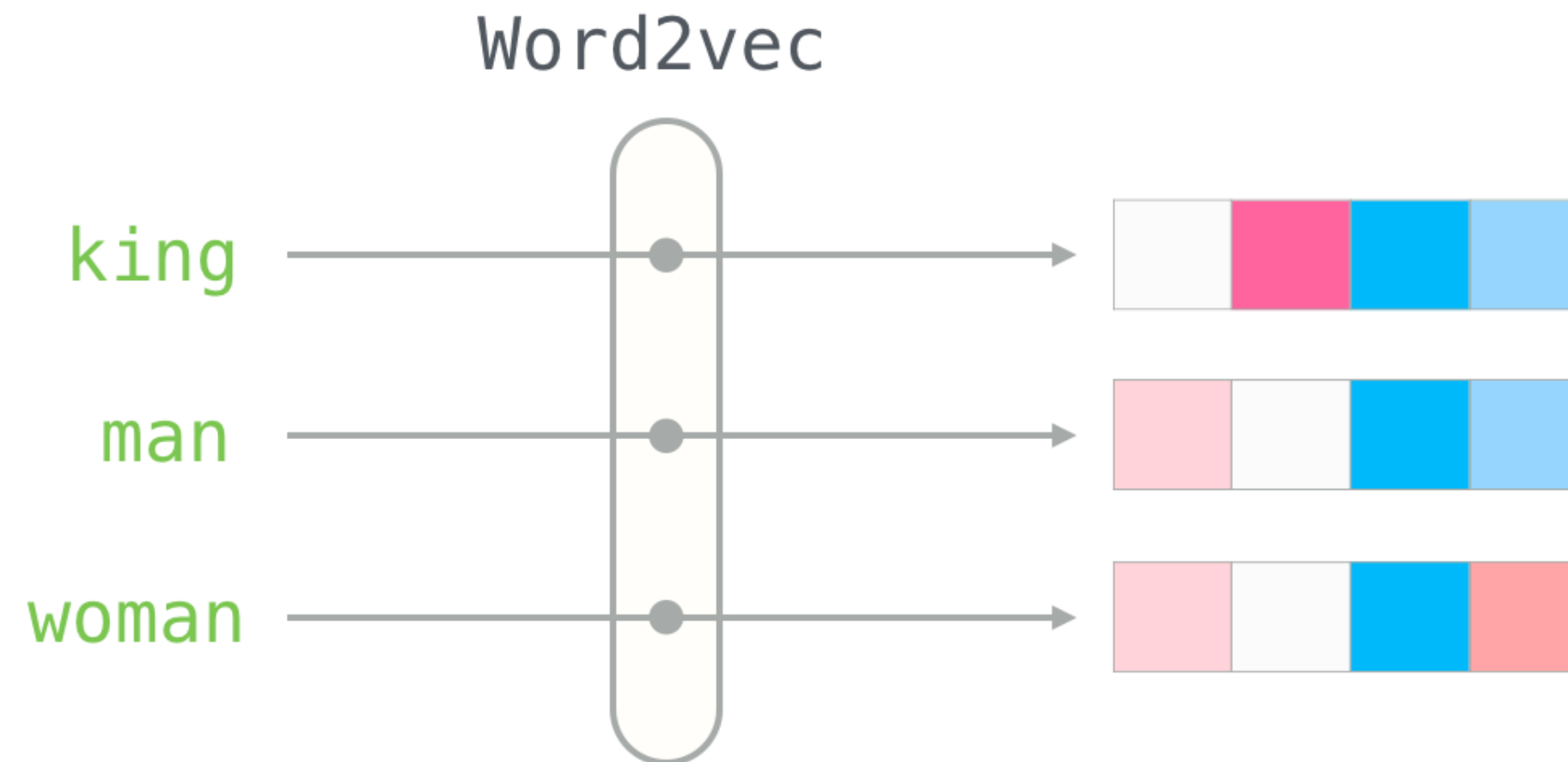
<대응되는 영양정보>

원재료 식품명 간의 유사도를 계산해서  
레시피 식재료 행에 영양성분 컬럼을 생성하기 위해 임베딩 실시

## 03 데이터 전처리

### 식품명 임베딩

Word2vec



단어를 밀집(dense) 벡터로 표현하는 방법 중 하나로,  
임베딩 벡터를 학습하기 위해 주변 단어의 문맥을 활용 하는 임베딩 방법

# 03 데이터 전처리

## 식품명 임베딩

텍스트 데이터 ' 베이컨 '



```
array([array([ 2.12392840e-03, -4.22502356e-03,  1.07915895e-02, -4.44079004e-03,
              -5.97989513e-03, -1.14261452e-02,  6.60570571e-03,  1.35580655e-02,
              2.57411011e-04,  6.41479669e-03, -6.28775125e-03, -3.13728378e-04,
              6.88820903e-04,  9.89304762e-03,  8.67232773e-03, -1.32759858e-03,
              1.01610534e-02, -1.49662709e-02, -6.38096186e-04, -5.60850045e-03,
              1.39407665e-02,  9.58110020e-03,  2.19920417e-03, -2.88818567e-03,
              8.10203608e-03,  4.52417554e-03, -4.11005138e-04, -8.74460675e-03,
              -4.35579289e-03, -8.94015003e-03,  4.76449542e-03, -7.92908669e-03,
              1.66559638e-03,  3.80423572e-03, -9.23421513e-03,  1.50978556e-02,
              -6.79752091e-03,  1.54419336e-03,  1.26507971e-03, -1.63118318e-02,
              -7.55820703e-03, -1.14210770e-02, -1.08809732e-02, -8.03406071e-03,
              6.75721513e-03, -1.19222533e-02, -1.46756601e-02,  8.10538232e-03,
              7.93689955e-03,  1.24858099e-03,  7.92294275e-03, -6.44115917e-03,
              6.59655593e-03,  3.02107516e-03,  1.75001950e-03, -4.84572072e-03,
              -5.36207762e-03, -2.24893936e-03,  1.81457493e-03, -7.28957495e-03,
              2.84954184e-03, -3.84639599e-04, -4.03054710e-03,  5.25839347e-03,
              -1.11151636e-02,  1.39582008e-02, -6.43913634e-03,  7.73468520e-03,
              -7.80889113e-03,  1.12527108e-03,  9.39650854e-05, -3.73877399e-03,
              1.06192147e-02, -9.76426713e-03, -1.87079608e-03, -4.34003258e-03,
              5.04218461e-03, -2.32792320e-03, -7.65803875e-03,  6.94661774e-03,
              -1.21707618e-02,  1.53455505e-04, -1.47578325e-02,  6.26228610e-03,
              -8.44805501e-04, -5.47548337e-03,  2.47970875e-03,  1.43824499e-02,
              6.99922210e-03, -6.38995552e-04,  3.41527443e-03, -3.66633083e-03,
              1.05038388e-02, -1.97501341e-03,  1.43065723e-02,  1.43965594e-02,
              7.15247774e-03, -1.22666853e-02,  8.19336064e-03,  1.48044247e-03],
              dtype=float32),
              dtype=object])
```

100 차원의 벡터를 가진 array 형태로  
잘 반환되는 것을 확인



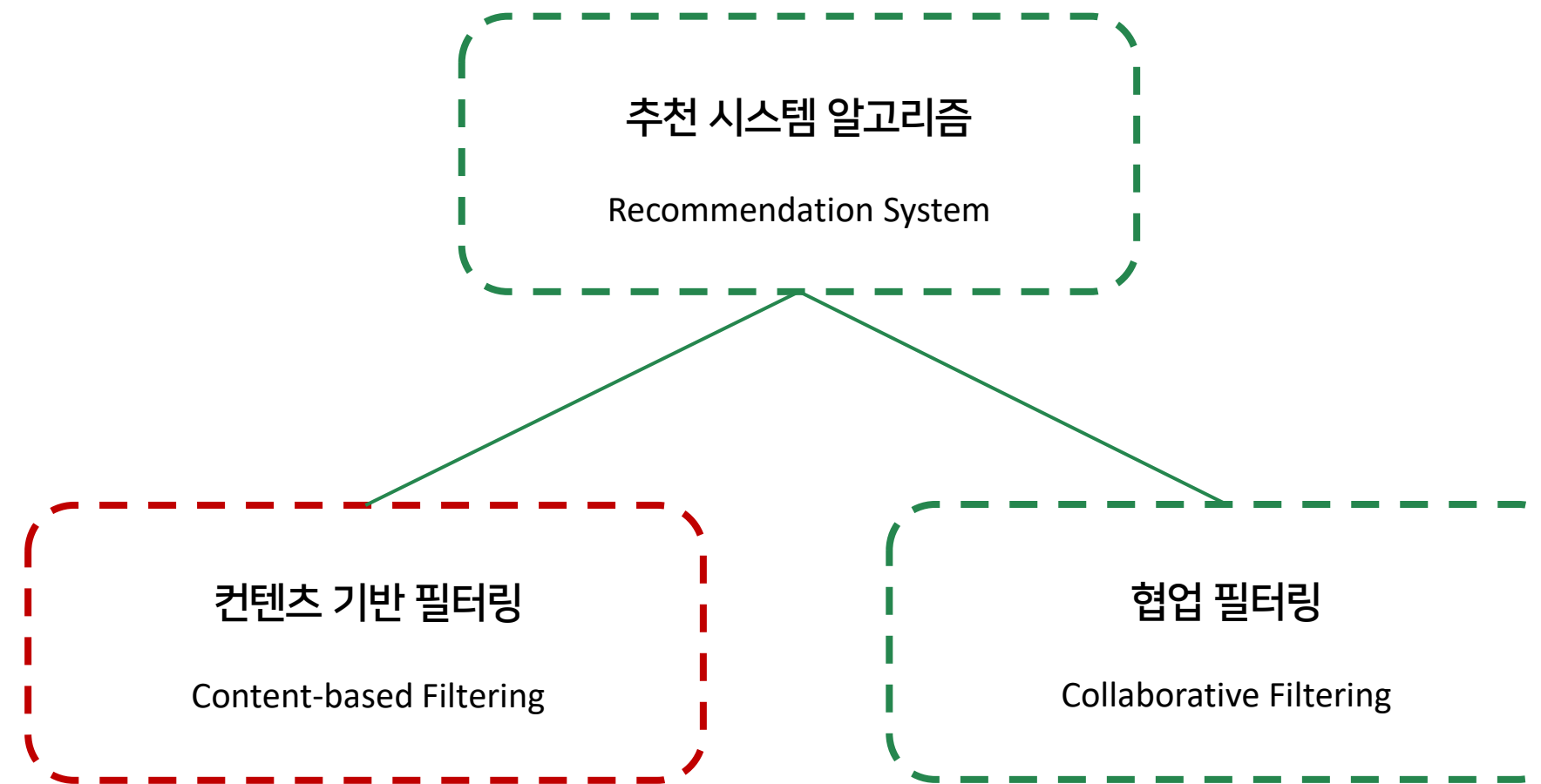
---

## 04 추천 시스템

---

## 04 추천 시스템

### 추천 시스템 소개



추천 시스템은 크게 컨텐츠 기반 필터링과 협업 필터링으로 나눌 수 있음  
→ 본 프로젝트에서는 사용자 평점 데이터가 없으므로 **컨텐츠 기반 필터링**을 사용



# 04 추천 시스템

## 1) 제약식 설정 | 질환별

고혈압 제약식

당뇨 제약식

각 질환에 좋은 영양소, 안 좋은 영양소들을 식별해서 리스트로 저장

고혈압에 좋은 영양소

식이섬유, 불포화 지방산, 마그네슘, 칼슘, 칼륨, 비타민

당뇨에 좋은 영양소

식이섬유, 불포화 지방산, 아연, 망간, 마그네슘

$$\sum a_i X_i \geq \sum good_g$$

$$\sum b_i X_i < \sum bad_g$$

## 04 추천 시스템

### 1) 제약식 설정 | 목적별

$$target\ intake - \epsilon \leq \sum a_i X_i \leq target\ intake + \epsilon$$

다이어트

벌크업

Harris - Benedict 방정식에 의거해  
일일 소비 칼로리를 계산하고, 탄수화물, 지방, 단백질 개별 양을 계산

소비 칼로리 = 활동 계수 \* 기초대사량

다이어트 / 벌크업 매크로 계산은

영양박사들의 방법론을 따름

## 04 추천 시스템

### 2) 식품대분류 생성 | Llama-3 활용

Llama-3

메타AI가 개발한 대규모 언어 모델(LLM)로,  
고급 자연어 처리 작업을 위해 설계된 인공 지능 모델



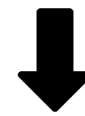
Llama-3 API를 활용하여  
입력 받은 음식 식재료의 식품대분류 생성!

## 04 추천 시스템

### 2) 식품대분류 생성 | Llama-3 활용

#### Llama-3 API 적용 결과

"['가공육류', '감자류 및 전분류', '견과류 및 종실류', '곡류', '과일류', '기타', '난류', '당류', '두류', '면류', '버섯류', '어패류 및 기타 수산물', '유제품', '유지류', '육류', '음료류', '잼류', '조리가공식품류', '조미료류', '주류', '차류', '채소류', '해조류']  
Wn{category}는 이중에 어디에 속해? 단답형으로 답해줘."



사과 입력 » '과일류'

쌀 입력 » '곡류'

소시지 입력 » '가공육류'

...

잘 분류된다는 걸 확인할 수 있음 !

### 2) 식품대분류 생성 | Llama-3 API 적용 결과

# 식품대분류 생성이 왜 필요할까 ?

Llama-3 API 적용 결과

"['가공육류', '감자류 및 전분류', '견과류 및 종실류', '곡류', '과일류', '기타', '난류', '당류', '두류', '면류', '버섯류', '어패류 및 기타 수산물', '유제품', '유지방', '육류', '음료류', '임유', '조리식품', '조미료류', '무류', '차류', '채소류', '해조류']"

입력받은 음식의 식재료들을 우리가 가지고 있던 원재료 식재료들 중  
같은 식품대분류 내에서 추천해주기 위함!



이를 통해 너무 엉뚱한 식재료를 추천해주는 것을 방지할 수 있음

사과 입력 > '과일류'

쌀 입력 > '곡류'

소시지 입력 > '가공육류'

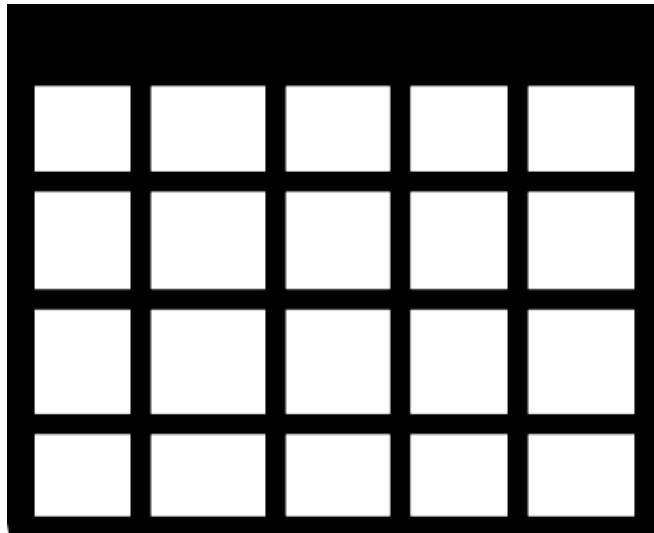
...

잘 분류된다는 걸 확인할 수 있음!

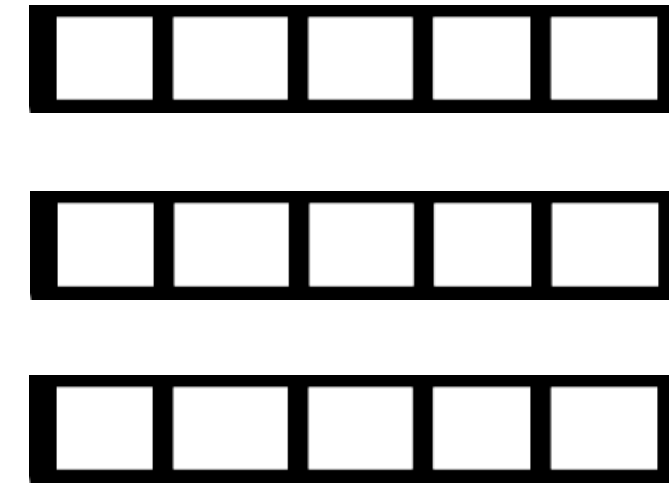
## 04 추천 시스템

### 3) 유사한 원재료 TOP K 선정

같은 대분류를 가진 원재료



Top K



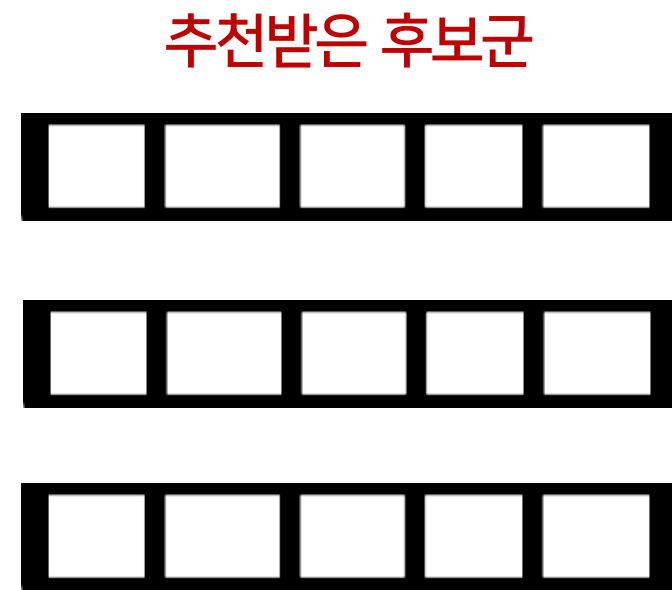
여기선 K = 3이라고 가정

대분류가 같은 원재료들 중

코사인 유사도를 이용해 해당 음식의 가장 유사한 원재료 Top K 개수를 뽑아냄

## 04 추천 시스템

### 4) 가능한 조합 찾기



'게살볶음밥' 조합 1: ~

'게살볶음밥' 조합 2: ~

⋮

추천 받은 후보군들을 이용해서  
가능한 모든 조합을 찾음

## 04 추천 시스템

### 5) 최적의 조합 찾기



안 좋은 영양소의 감소율과 좋은 영양소의 증가율이 높은 순서대로 sorting 후  
최적의 조합 FIND





---

## 05 시스템 구현

---

## 05 시스템 구현

### INPUT



INPUT

바꾸고 싶은 음식

타입(다이어트, 벌크업, X)

키 (cm단위)

제지방 (lbm)

몸무게 (kg)

평소 활동 정도 (low, med, high)

질환 유형(고혈압, 당뇨, X)

성별

## 05 시스템 구현

### INPUT



INPUT

잡채

X

178

25

76

median

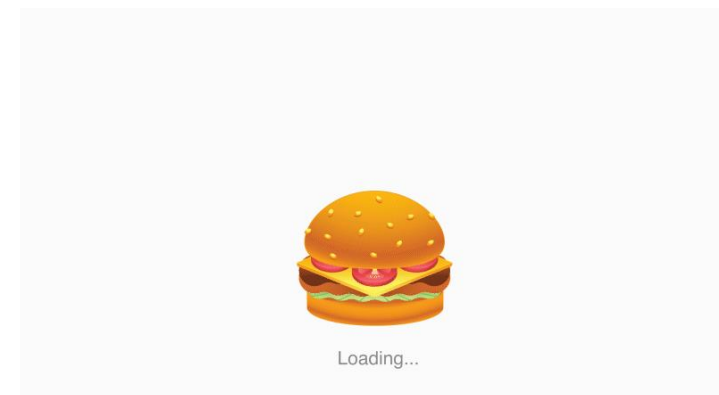
당뇨

남

## 05 시스템 구현

---

### 식재료 변환



... 건강 요정이 레시피를 재구성하고 있어요. 잠시만 기다려 주세요 ...

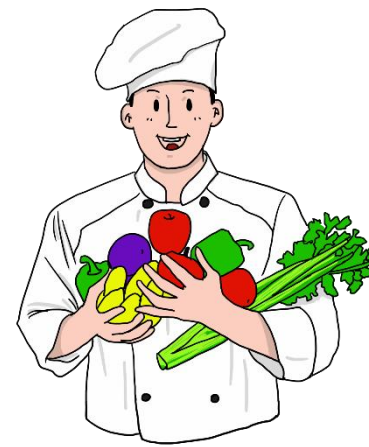
## 05 시스템 구현

### OUTPUT

당뇨가 있어도 건강하게 잡채를 먹을 수 있는 레시피를 찾았어요!

당면 60g  
부추 20g  
당근 20g  
미니파프리카 10g  
돼지고기 등심 50g  
양파 5g  
간장 5g  
참기름 10g  
설탕 5g

원래 레시피



당면 60g  
부추 20g  
당근 20g  
미니 파프리카 10g  
소고기 한우 사태 50g  
양파 5g  
산분해 간장 5g  
참기름 10g  
감 조청 5g

맞춤 레시피

## 05 시스템 구현

### OUTPUT

당뇨가 있어도 건강하게 잡채를 먹을 수 있는 레시피를 찾았어요!

727kcal,  
탄수화물 71g  
단백질 55g  
지방 24g



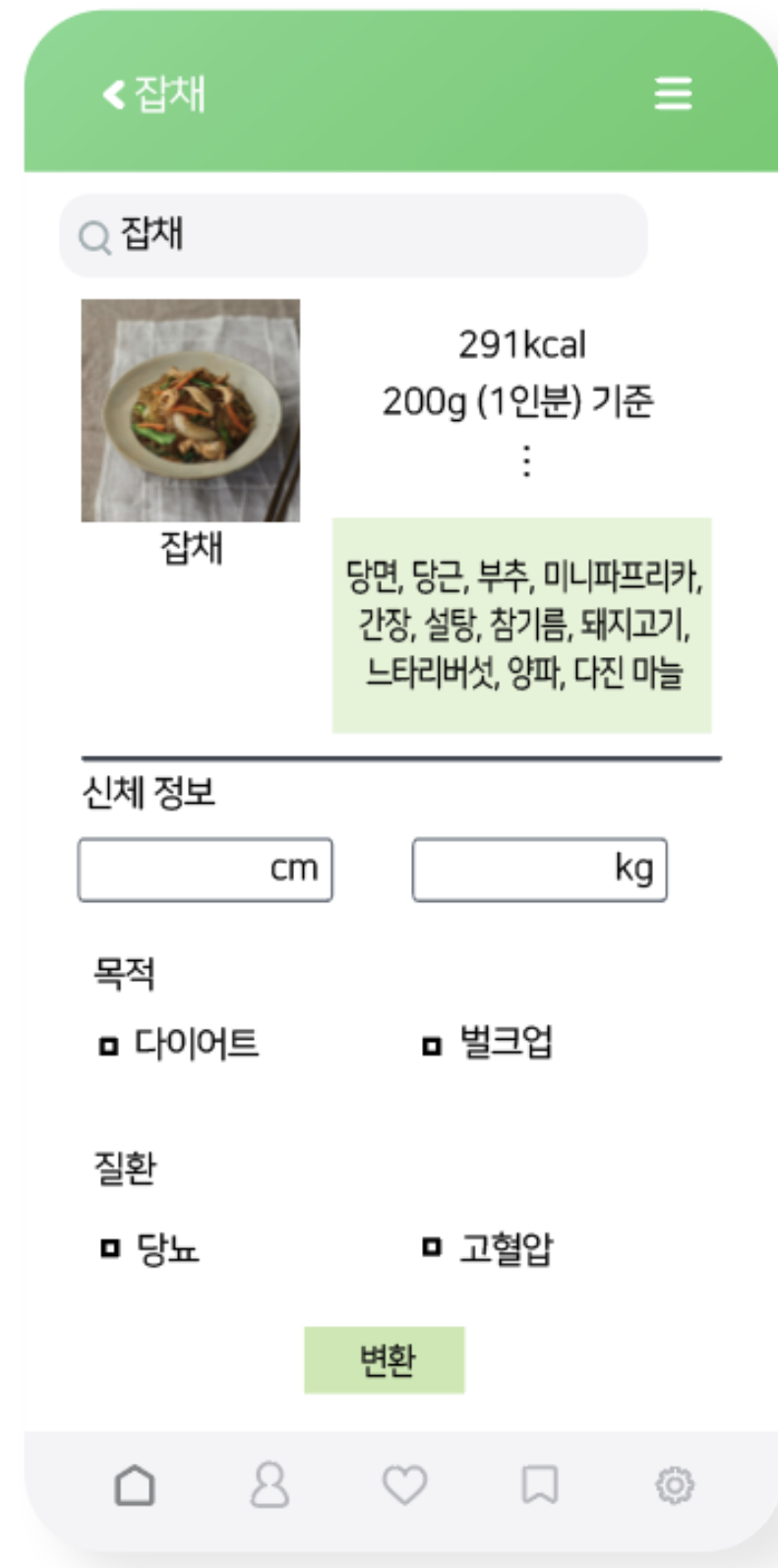
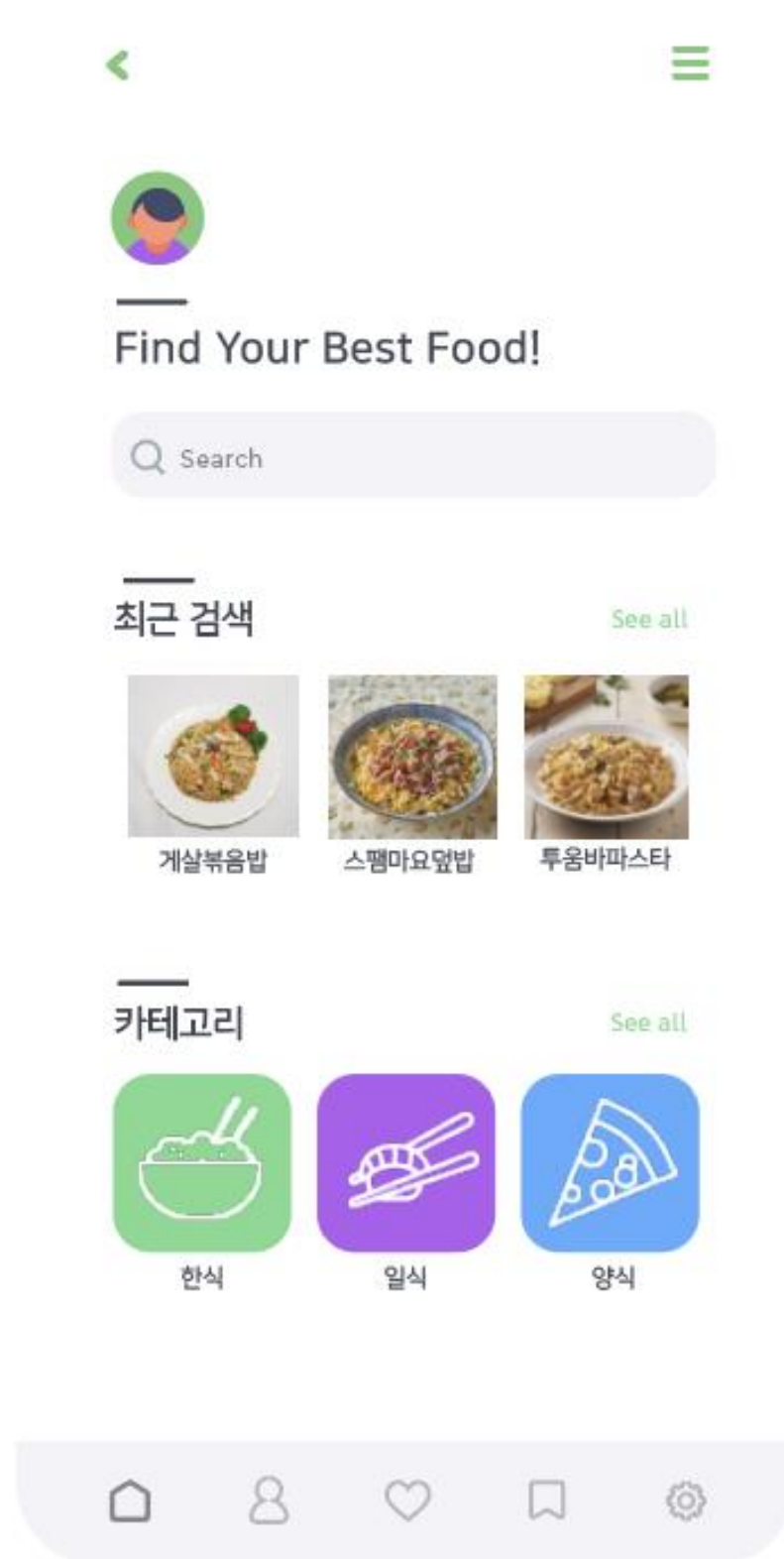
760kcal,  
탄수화물 75g  
단백질 50g  
지방 30g

한 끼 동안 먹어야 하는 kcal와 탄단지 양이에요.

기존 잡채 레시피는  
탄수화물, 지방 함량이 5g, 6g 씩 높았고,  
단백질 함량이 5g 적었어요.

## 05 시스템 구현

### 플랫폼 구현 예시







---

## 06 의의 및 한계

---

## 06 의의 및 한계



### 건강 식재료 변환 시스템의 의의



유형별로 제약식을 달리해 **사용자 맞춤 추천**



유사도 계산 시 단순히 **영양성분**뿐 아니라 식감을 결정하는 **물성**과 연관된 탄단지 비율 고려



극단적인 다이어트 혹은 벌크업이 아닌 **건강한 식단**을 할 수 있도록 **매크로 계산**



자주 추천되는 식재료의 빈도수를 보고 어떤 식재료가 건강한 식재료인지 간접적으로 평가 가능

## 06 의의 및 한계



### 건강 식재료 변환 시스템의 한계

✓ 식재료 데이터 수의 부족

✓ 모든 영양소를 고려하지 못함

✓ 더 많은 질병을 고려하지 못함

✓ 레시피 식재료 데이터에 원재료 영양성분을 덧붙여 유사도 계산을 하였기 때문에 정확도가 떨어짐

## 06 의의 및 한계

---

### 발전 가능성

본 시스템 한계의 주 원인은 **원재료 데이터의 부족**에 있음



식약처에서 원재료 데이터를 더욱 충분히 제공해준다면  
훨씬 심도 있고 정확한 시스템으로 **발전 가능성이 존재함**

---

감사합니다

---