

Classification Model of Food Groups in Food Exchange Table Using Decision Tree-based Machine Learning

Ji Yun Kim*, Jongwan Kim**

*Manager, Rolling Pasta of THEBORN KOREA INC., Seoul, Korea

**Professor, Software Convergence Education Center, Sahmyook University, Seoul, Korea

[Abstract]

In this paper, we propose a decision tree-based machine learning model that leads to food exchange table renewal by classifying food groups through machine learning for existing food and food data found by web crawling. The food exchange table is the standard for food exchange intake when composing a diet such as diet and diet, as well as patients who need nutritional management. The food exchange table, which is the standard for the composition of the diet, takes a lot of manpower and time in the process of revision through the National Health and Nutrition Survey, making it difficult to quickly reflect food changes according to new foods or trends. Since the proposed technique classifies newly added foods based on the existing food group, it is possible to organize a rapid food exchange table reflecting the trend of food. As a result of classifying food into the proposed model in the study, the accuracy of the food group in the food exchange table was 97.45%, so this food classification model is expected to be highly utilized for the composition of a diet that suits your taste in hospitals and nursing homes.

▶ **Key words:** Artificial Intelligence, Machine Learning, Decision Tree, Food Exchange Table, Random Search

[요 약]

본 논문에서 우리는 기존 식품과 웹 크롤링으로 찾은 식품 데이터에 대해 기계학습으로 식품군을 분류하여 식품교환표를 갱신하기 위한 의사결정트리 기반의 기계학습 모델을 제안한다. 식품교환표는 영양 관리가 필요한 환자의 식이요법이나 다이어트 식단을 편성할 때 식품 교환 섭취에 사용된다. 식단의 기준이 되는 식품교환표는 국민건강영양조사를 통한 개정과정에서 많은 인력과 시간이 소요되어 새로운 식품이나 트렌드에 따른 식품 변화를 신속하게 반영하기 어렵다. 제안 기법은 기존의 식품군을 바탕으로 새롭게 추가되는 식품을 분류하기 때문에 식품의 트렌드를 반영한 식품교환표 구성이 가능하다. 연구에서 제안 모델로 식품을 분류한 결과, 식품교환표의 식품군에 대한 정확도가 97.45%로 나타났으며, 본 식품 분류 모델은 병원, 요양원 등에서 식단 구성에 활용도가 높을 것으로 전망된다.

▶ **주제어:** 인공지능, 기계학습, 의사결정트리, 식품교환표, 랜덤서치

-
- First Author: Ji Yun Kim, Corresponding Author: Jongwan Kim
 - *Ji Yun Kim (wldbs3592@naver.com), Rolling Pasta of THEBORN KOREA INC.
 - **Jongwan Kim (kimj@syu.ac.kr), Software Convergence Education Center, Sahmyook University
 - Received: 2022. 11. 25, Revised: 2022. 12. 08, Accepted: 2022. 12. 14.

I. Introduction

과학기술과 통신의 발전으로 외국 문화를 쉽게 접할 수 있어서 한국인의 식생활은 점차 서구화·세계화되고 있으며 식품 산업의 발달로 새로운 식품들이 식탁에 오르고 있다. 식품 소비에서는 간편화, 고급화 및 다양화 등 트렌드가 변화하고 있으며 식품 섭취에 대한 질적 요구도 나타나고 있다[1]. 식품에 대한 기호 변화는 산업적 측면에서 식품의 다양화와 문화적 측면에서 외국 식품의 유입에 기인하며 식품이 다양해지면서 건강관리의 필요성도 높아지고 있다 [2-3].

한국보건산업진흥원의 국민영양통계에 따르면 한국인의 식물성 식품 섭취 비율은 2015년 79.01%에서 2020년 76.07%로 낮아졌으며 동물성 식품 섭취 비율은 2015년 20.99%에서 2020년 23.93%로 높아졌다. 식물성 식품과 동물성 식품 섭취량은 각각 2015년 1,273.3g, 2020년 1,127.01g, 2015년 336.95g에서 348.72g으로 증가하였다. 이는 한국인의 식생활에서 외식과 간편식이 늘어나면서 육류를 많이 섭취하는 것을 나타낸다[4]. 동물성 식품 섭취의 증가는 향후, 영양불균형으로 나타날 수 있으며 비만과 당뇨의 발생이 우려되는 상황이다[5].

식품교환표는 비만, 당뇨와 같은 건강 관련 위험을 해소하기 위해 식품을 교환하여 섭취하는 기준을 제공한다[6]. 영양소를 기준으로 6가지 식품군으로 구분하여 식단계획을 통해 식품을 교환하면서 당뇨병 등 대사증후군 환자나 각종 질환자의 식단을 조절하고 관리할 수 있다. 따라서 사용자는 식품교환표에서 영양소가 비슷한 식품을 교환 및 섭취하여 식이요법, 다이어트 등을 위한 능동적인 영양 관리에 도움을 얻을 수 있다.

식품교환표는 대한당뇨병학회, 대한영양사협회, 한국영양학회가 공동 개발한 이후 지속해서 개정하고 있다. 현재 식품교환표는 섭취 빈도수가 높은 식품과 환자가 섭취한 음식들을 선정하여 식단을 추가하는 방식으로 식품 목록을 수정한다. 예를 들어 2010년 식품교환표는 국민건강영양조사에서 빈도수가 높은 식품, 당뇨병 환자 섭취 식품과 2009년 3월부터 5월까지 53개 병원에서 진행된 설문 조사, 2008년 1월부터 12월까지 당뇨병 환자가 작성한 식사 일기의 4가지 자료를 사용하였다[7].

식품교환표를 위한 국민건강영양 조사는 결과 도출까지 약 1년이 소요되며 인력, 시간 등 물질 비용을 수반한다. 특히, 식품 목록 추가를 위한 설문 조사에 여러 달을 소요하기 때문에 식품 섭취 변화의 신속한 반영이 어렵다는 단점을 가지고 있다. 또한, 새롭게 개발된 식품의 경우에는 기존 식품 목록에 포함되지 않아 식단을 계획하기 위한 식

품 교환 단위 수를 알기 어렵고 명확한 분류 기준이 나타나 있지 않다. 위와 같은 문제를 해결하기 위해서는 식품을 신속하게 추가하여 식품의 다양성을 높이고, 개인 선호도를 반영하는 새로운 식품교환표 구성 방법이 필요하다.

본 논문은 식품교환표 개정에 드는 시간과 인력을 절감하고 트렌드에 부합하는 새로운 식품교환표 구성을 위한 인공지능의 기계학습 모델을 제안한다. 본 모델은 의사결정트리 기반의 지도학습을 이용하여 식품이 속하는 식품군을 분류한다. 기계학습 모델은 기존의 식품군 데이터를 기반으로 구축하며 새로운 식품은 웹 크롤링(web crawling)으로 확보한 텍스트(text)에서 형태소 분석[8]으로 추출하여 모델을 통해 식품군을 분류한다.

본 연구는 기계학습을 통해 개별 식품이 속하는 식품군을 분류함으로써 인공지능을 이용한 신속한 식품교환표 구성의 가능성을 제시한다. 이를 실무에 적용할 경우, 시간과 인력 절감으로 경제적 효과를 유발하며 식품의 트렌드가 반영된 식품교환표를 신속하게 구축할 수 있으므로 식품교환으로 지속적인 건강관리의 가능성을 높일 것이다.

본 연구가 식품군 분류를 통한 식품교환표 갱신에 공헌하는 내용은 다음과 같다.

- 식품교환표 개정 시 수작업으로 수행해온 식품 분류 과정의 시간과 인력을 절약한다.
- 기계학습을 통해 다양한 식품 생산과 식생활의 변화를 반영한 신속한 식품교환표 구성을 지원한다.
- 새로운 식품을 신속히 추가하여 식품 트렌드를 반영할 수 있으므로 현실적인 식품교환표에 따른 식사 조절을 통한 영양 관리나 사용자의 선호도에 맞는 식단 구성을 지원한다.
- 본 연구는 의사결정트리 기반의 식품군 학습 모델에 하이퍼파라미터 튜닝(hyper-parameter tuning)을 적용하여 선행 연구[9]를 확장함으로써 향상된 식품군 분류 결과를 나타낸다.

본 논문의 구성은 다음과 같다.

2장에서는 식품교환표와 의사결정트리 모델을 살펴보고, 3장에서는 식품 데이터 전처리와 분류 모델을 구성한다. 4장은 시뮬레이션을 통해 식품 분류 모델 학습 및 평가를 수행한다. 5장은 기존 연구를 확장한 본 연구에 관한 통찰 내용을 설명하면서 마친다.

II. Preliminaries

2.1 Food Exchange Table and Units

건강관리나 환자를 위한 식단은 식품교환표에서 제시하

는 교환 가능한 단위로 변경하여 구성할 수 있다. Table 1의 식품교환표는 영양소가 비슷한 곡류군, 어육류군, 채소군, 우유군, 과일군, 지방군의 6개로 구성되며 어육류 군의 경우 지방함량에 기초하여 저지방, 중지방, 고지방으로 나뉜다[6].

식품교환표의 식품은 같은 군에서 바꿔 먹을 수 있으므로 선호하는 대상이나 피해야 할 음식을 교환할 수 있으며 서로 영양소 성분이 비슷하므로 영양은 유지된다.

식품교환표를 식단의 기준으로 사용하면 균형 잡힌 영양소를 반영하기 위해 식품을 효과적으로 이용할 수 있다. 에너지 측면에서는 총에너지를 조절하면서 충분한 에너지 섭취를 계획할 수 있다. 예를 들어 곡류군에서는 쌀밥을 녹두와 교환할 수 있으며 어육류 군에서는 저지방의 닭고기를 새우 등으로 대체할 수 있다.

식품교환표는 2010년에 개발된 이후 일부 개정을 위한 노력이 있었으나 현재까지 병원, 요양원 등의 실무에서 2010년 식품교환표를 기준으로 사용한다. 따라서 본 논문에서는 2010년에 개발된 식품교환표를 적용한다.

2.2 Decision Tree Model

의사결정트리(decision tree)는 의사결정 규칙을 나무(tree) 구조로 나타낸 것으로서 자료를 몇 개의 묶음으로 분류하는 분석 도구이다. 선택한 기준점으로부터 데이터의 패턴을 인식하여 “if~then” 형식으로 분기하여 최종 데이터를 검색한다.

의사결정트리는 인공지능의 기계학습에서 데이터를 분류할 때 주요 기법으로 사용하며 답이 존재하는 경우 직관적으로 답을 찾을 수 있어서 지도학습을 이용한 기계학습에 활용된다[10]. 트리의 각 노드에 분류 기준 등이 표기되어 분류과정을 화이트박스(white box) 형태로 식별할 수

있다[11].

Fig. 1은 기계학습을 통해 포도주 중 레드와인, 화이트와인을 식별하는 의사결정트리를 나타낸다. 첫 루트노드(root node)는 설탕(sugar) 농도를 기준으로 좌우 하위 노드를 생성하면서 포도주를 분류한다. 노드에서 gini는 각 노드에 서로 다른 데이터가 포함된 불순도(impurity)를 나타내며 부모노드와 자식노드 분기는 불순도의 차이가 큰 방향으로 나뉜다. 이처럼 결정트리는 노드들 사이의 분류 특징을 알 수 있어서 화이트박스라 한다.

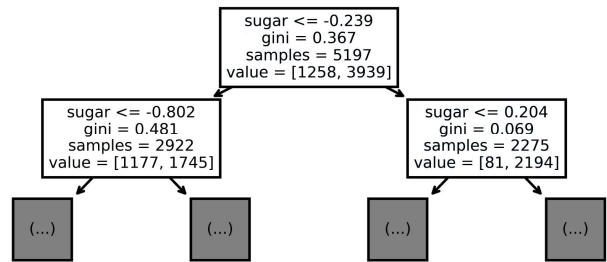


Fig. 1. Decision Tree of Machine Learning

III. Food Data Pre-processing and Machine Learning Model

식품교환표를 갱신하기 위해 설문 조사와 수작업으로 식품군을 분류하는 것은 많은 시간과 인력이 소요되므로 식단 구성을 위한 신속한 정보제공에 제약이 존재한다. 식품의 트렌드를 반영한 식품 분류 및 식품교환표 갱신을 위해서는 새로운 식품을 신속하게 기존 식품군으로 분류하는 기계학습 모델이 필요하다.

식품 데이터는 기존의 식품교환표를 구성하는 식품과

Table 1. Food Exchange Table and Exchange Units

Groups	Grain	Fishes & Meat			Vegetable	Fruit	Milk	Fat
		Low Fat	Mid Fat	High Fat				
Foods	Cooked rice (70g)	Chicken(meat) (40g)	Beef(sirloin, lean beef ribs) (40g)	Bacon (40g)	Cabbage (70g)	Banana (50g)	General milk (200g)	Sesame (8g)
	Plain bread (35g)	Dried croaker (15g)	Egg (55g)	Beef ribs (40g)	Sesame leaf (40g)	Pear (110g)	Lactose milk (200g)	Peanut butter (8g)
	Potato (140g)	Small octopus (100g)	Mackerel (50g)	Vienna sausage (40g)	Bellflower (40g)	Mandarin (120g)	Soybean milk (200g)	Margarin (5g)
	Green gram (70g)	Abalone (70g)	Saury (50g)	Canned tuna (50g)	Altitude mushroom (70g)	Pine apple (200g)	Powdered milk (25g)	Butter (5g)
	Sweet Potato (70g)	Lobstar (50g)	Black soybean (20g)	Cheese (30g)	Seeweed (70g)	Grape (80g)	Infant formula (25g)	Mayonnaisewl (5g)

새로 개발되거나 해외에서 유입되는 식품으로 구분된다. 새로운 식품 데이터는 웹 크롤링을 통해 현재 선호하는 식품을 탐색하여 추가한다. 웹 데이터로부터 식품 추출은 수집된 텍스트를 KoNLPy[12]로 형태소 분석 후 Table 2와 같이 명사를 기준으로 진행하며 한글은 어근, 접두사, 접미사 등과 영어는 관사, 전치사 등을 제외하여 식품명을 식별한다[13]. 형태소 분석을 통해 추출된 신규 식품은 영양 성분과 함께 식품교환표 구성에 활용된다.

Table 2. Morphological Analysis

Step	Contents
Sentence	Diabetes patients should avoid rice and eat alternative foods.
Tokens	Diabetes, Patients, Rice, Foods, Eat
Noun	Rice

식품군 분류를 위한 데이터는 현재에도 활용도가 높은 2010년도 식품교환표가 반영된 ‘당뇨병 영양 관리와 식품교환표’[14]와 농촌진흥청의 ‘국가표준식품성분표’[15]에서 제시된 분류 식품군이 존재하는 1,192개의 식품 정보를 사용한다. 기존의 식품과 새로 추가된 식품을 분류하기 위해서 식품교환표의 지방, 단백질, 탄수화물, 총 식이섬유, 총당류에 대한 함량 정보를 사용하며 세부 영양소는 농촌진흥청 국가표준식품성분표를 참고한다.

식품성분표와 식품의 카테고리에 따라 준비된 데이터는 학습 과정에서 데이터 분류율을 높이기 위해 전처리를 수행한다. 데이터 중 귀리, 토마토와 같이 당류가 없는 결측값은 Table 3과 같이 0.0(총 당류)으로 처리하였으며 이는 식품성분표에서 누락되었지만, 학습 과정에서 결측값이 학습 모델에 미치는 영향을 최소화하기 위한 것이다.

식품의 영양소인 지방, 단백질, 탄수화물, 총 식이섬유, 총당류는 각각 100 칼로리(calory)를 기준으로 정규화하여 해당 비율에 맞는 식품군으로 분류되도록 하였다. 식품군은 기존의 6개에서 총 8가지로 확장하고 데이터 수가 적은 저지방 우유와 일반 우유 군은 통합한다. 데이터는 모델의 성능을 높이기 위해 전처리 후 섞어서 사용한다.

Table 3. Pre-processing of food data

Foods	Energy (kcal)	Protein (g)	Fat (g)	Carbohydrate (g)	Total sugar (g)	Dietary fiber (g)	Category	Food Group
Oats, Outer oats, Mowed, Raw	100.0	3.000000	0.059701	22.507463	0.0	4.029851	Cereal and Product	Grain
Oats, Oatmeal	100.0	2.688312	0.103896	22.584416	0.0	2.077922	Cereal and Product	Grain
				...				
Tomato sauce, Ketchup	100.0	11.780822	0.753425	36.438356	0.0	0.000000	Flavoring matter	Grain

기계학습을 위한 의사결정 모델은 Fig. 2와 같이 구성되며 식품의 지방, 단백질, 탄수화물, 총 식이섬유, 총당류의 함량과 카테고리 정보를 데이터로 사용한다. 데이터는 훈련(train) 데이터, 검증(validation) 데이터 및 테스트(test) 데이터로 구분한다. 검증 데이터는 훈련 데이터로 훈련한 초기 모델의 성능을 확인하고자 사용하며 과적합(overfitting)이나 과소적합(underfitting)이 발생하면 검증 데이터를 다른 위치에서 선택하여 사용한다. 여러 개의 검증 데이터 중 성능이 높은 것을 훈련 데이터에 포함하여 최종 의사결정트리 모델로 학습한다. 검증 데이터로 확인한 정확도는 기본 결정트리를 사용한 것이며 가지치기, 파라미터 변경 등을 통한 정확도 향상은 실험 과정에서 기본 모델을 튜닝하여 진행한다.

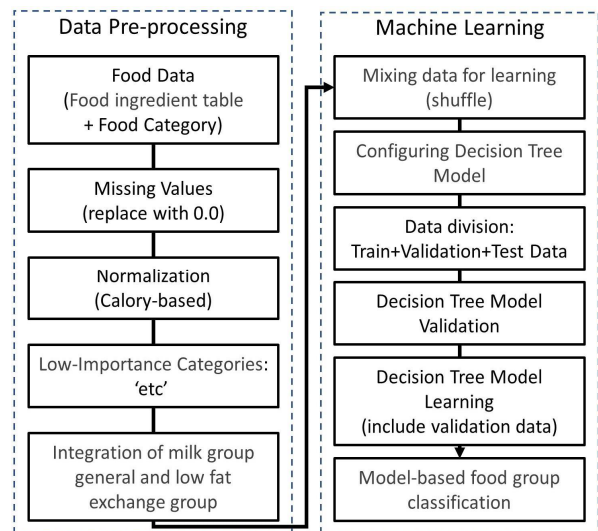


Fig. 2. Data Pre-processing and Model Configuration

IV. Food Classification Model and Evaluation

식품군 분류를 위한 인공지능 모델은 의사결정트리를 적용하였다. 의사결정트리를 사용한 이유는 식품 분류과정에서 각 노드의 불순도와 하위 트리의 개수 확인이 쉽고

트리의 깊이별 분류과정을 직관적으로 파악할 수 있기 때문이다.

기계학습 모델 개발 및 실험은 GPU를 갖춘 시스템을 사용하였으며 개발환경은 Table 4와 같다. 기계학습 모델 구현은 scikit-learn[16]을 사용하였으며 데이터는 기존 식품교환표 및 국가표준식품성분표를 기준으로 전처리 후 사용하였다.

기계학습을 위한 데이터 구성은 Fig. 3의 2~13행과 같이 전체 데이터 1,192개를 훈련 데이터(716), 검증 데이터(238), 테스트 데이터(238)로 구분하여 각각 60:20:20의 비율로 나눴다.

소스 코드에서 train80_input과 test20_input은 전체 데이터 중 훈련 데이터와 테스트 데이터를 80%와 20%로 구분한 것이다. 이후 검증을 위해 train60sub_input과 valid20_input으로 한 번 더 구분하였다. 의사결정트리 구성은 14행에서 최대 깊이 max_depth를 15로 정하였다. 모델 훈련은 18행과 같이 60%의 훈련 데이터와 라벨을 대상으로 진행하였으며 모델의 정확도는 19~20행과 같이 확인하였다.

Table 4. Classification Model Development Environment

Item	Contents
CPU	Intel Xeon Silver 4114 2.20GHz * 2EA
RAM	128GB
GPU	RTX 2080 Ti * 2EA
Language	Python (Jupyter Notebook)
OS	LINUX Ubuntu
Libraries	scikit-learn(Machin Learning), KoNLPy(Morpheme analysis)
Learning Data set	Food Exchange Table(2010), National standard food composition table(2022), Total food items: 1,192.

의사결정 트리에서 깊이별 정확도는 max_depth를 변경하면서 확인하였으며 트리의 깊이가 1일 때, 훈련 데이터와 테스트 데이터의 정확도가 각각 0.757, 0.737로 과소 적합으로 나타났다. 깊이가 증가할수록 훈련 데이터의 정확도는 올라갔으나 테스트 데이터의 정확도와 차이가 많은 과대 적합을 보인다.

```

1 # train and test date
2 train80_input, test20_input
3 , train80_target, test20_target
4 = train_test_split(data, target, test_size=0.2
5 , random_state=55)
6
7 # train and validation data
8 # The ratio is 60 vs 20 from train80_input data
9 train60sub_input, valid20_input
10 , train60sub_target, valid20_target
11 = train_test_split(train80_input, train80_target
12 , test_size=0.2
13 , random_state=55)
14 dtree = DecisionTreeClassifier(max_depth=15
15 , random_state=55)
16
17 # training with train data
18 dtree.fit(train60sub_input, train60sub_target)
19 print(dt.score(train60sub_input, train60sub_target))
20 print(dt.score(valid20_input, valid20_target))
    
```

Fig. 3. Decision Tree Learning Model for Food Classification

모델의 정확도를 시뮬레이션하기 위해 깊이를 1부터 15까지 변경하여 점수를 확인 결과는 Fig. 4와 같다. 한 쌍으로 작성된 그래프 중 왼쪽 막대는 훈련 데이터인 train80_input의 결과이며 우측 막대는 테스트 데이터인 test20_input의 결과이다. 깊이가 낮을 때는 과소 적합을 보이다가 여덟 번째 막대 그래프가 나타내는 깊이 8에서 두 데이터의 차이가 가장 작아지며 최적의 결과를 나타낸다. 이후에는 훈련 데이터와 테스트 데이터의 차이가 벌어지면서 과대 적합을 보인다. 깊이 8에서 과대 및 과소 적합이 해소된 결과를 나타내지만, 적합도는 0.845, 0.841로 여전히 높은 점수는 아니다. 시뮬레이션 결과에 따라 의사결정트리의 성능을 개선하기 위한 튜닝은 모델 학습과정에서 사용되는 파라미터를 통해 진행하였다.

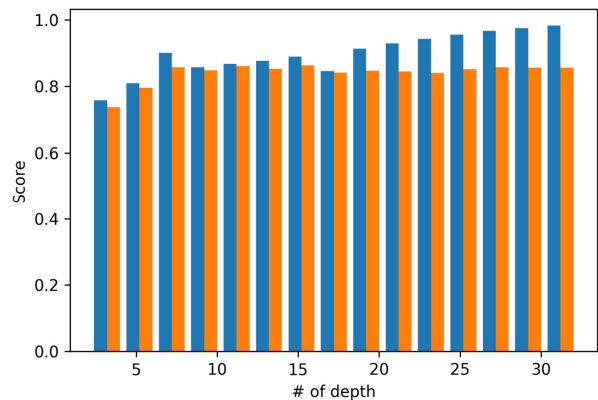


Fig. 4. Scores in Depth of Decision Tree

본 학습 과정에서 성능개선을 위한 튜닝은 최적 하이퍼파라미터를 찾기 위해 scikit-learn의 RandomizedSearchCV() 함수를 이용한 랜덤 서치를 사용하였다. RandomizedSea

rchCV()는 파라미터에 따라 최적의 값을 찾아주므로 의사결정트리와 같이 깊이가 깊으면 자동 탐색이 가능하다.

기계학습 라이브러리인 scikit-learn에서 의사결정트리 모델 생성 시 사용한 훈련 모델의 학습 파라미터는 Table 5와 같다. 트리의 노드 생성에서 하위 노드로 분기는 최소 데이터가 2개일 때 진행하도록 min_samples_split을 지정하였으며 각 노드 사이의 분기 품질 측정은 gini를 사용한다. splitter는 각 노드에서 분할을 선택하는 데 사용되는 전략이며 단말(leaf) 노드의 최소 데이터 샘플 수는 1이다. 최소 불순도 min_impurity_decrease는 0.0이며 트리의 성장을 멈추는 임계치 min_impurity_split는 0이다.

랜덤 서치를 사용한 튜닝 파라미터는 Table 6과 같이 진행하였으며 Fig. 5와 같이 n_iter 변수에서 150번 반복하였다. CPU의 코어는 최대한 사용할 수 있도록 n_jobs는 -1을 할당하였으며 훈련 데이터는 전체 데이터의 80%를 할당하였다. 모델 학습 후 하이퍼파라미터의 최적값은 max_depth: 40, min_impurity_decrease: 0.00042167, min_samples_leaf: 7, min_samples_split: 15로 계산되었다. 의사결정 모델은 튜닝된 하이퍼파라미터에 따라 재훈련 후 식품분류에 사용하였다.

Table 5. Decision Tree Model Parameters

Parameter	Value
criterion	gini
splitter	best
max_depth	15
min_samples_split	2
min_samples_leaf	1
min_weight_fraction_leaf	0.0
max_features	None
random_state	55
max_leaf_nodes	None
min_impurity_decrease	0.0
min_impurity_split	0
class_weight	None
presort	deprecated
ccp_alpha	0.0

Table 6. Random Search Tuning Parameter

Parameter	Value
min_impurity_decrease	0.0001~0.01
max_depth	1~50
min_samples_split	2~30
min_samples_leaf	1~30

```

1 params
2 = {'min_impurity_decrease': uniform(0.0001, 0.01),
3   'max_depth': randint(1, 50),
4   'min_samples_split': randint(2, 30),
5   'min_samples_leaf': randint(1, 30),
6   }
7 rs_tree = RandomizedSearchCV(
8   DecisionTreeClassifier(
9     random_state=55), params, n_iter=150
10  , n_jobs=-1, random_state=55)
11 rs_tree.fit(train80_input, train80_target)
    
```

Fig. 5. Training Model with Random Search

의사결정트리를 사용한 지도학습에서 라벨링(labeling)을 위한 속성에는 지방, 단백질, 탄수화물, 총 식이섬유, 총당류 및 카테고리 항목을 사용하였으며 학습 결과는 각 식품에 대한 식품군으로 분류되었다.

식품군 분류를 위한 의사결정트리는 기존 식품 1,192개 식품으로 생성하였으며 트리의 구성은 Fig. 6과 같다.

Fig. 5에서 최적의 파라미터를 찾기 위한 RandomSearchCV()의 n_iter를 Fig. 7과 같이 10~160까지 변경하여 실행한 결과 150번 반복할 때 최적의 파라미터 값이 계산되었으며 파라미터를 모두 적용한 예측 점수는 97.45%로 나타났다.

분류가 충분하지 못한 2.55%의 식품은 식품구성표와 영양소 측정 시기의 차이에 따라 분류가 안 된 것으로 보인다. 즉, 데이터 구성 시 2010년도 식품교환표에서 누락된 식품 성분은 농촌진흥청의 국가표준식품성분표 제10개정판을 참고하였으므로 두 식품에서 성분의 차이가 기계학습 모델 구축에 영향을 준 것으로 판단된다. 특히, 선행 연구인 [9]에 나타난 95.37%의 정확도는 본 연구에서 향상되었는데, 이유는 랜덤 서치를 통한 최적의 하이퍼파라미터 탐색에 따른 것으로 보인다.

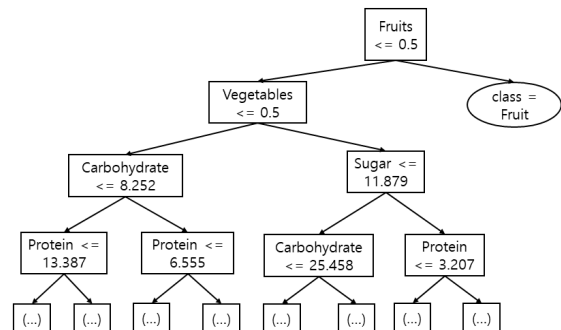


Fig. 6. Decision Process of Decision Tree

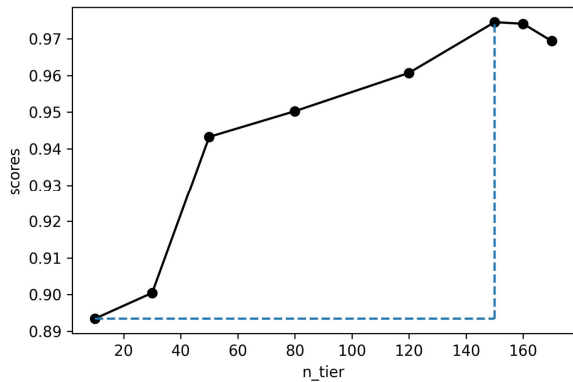


Fig. 7. n_tier of Random Search

검증 데이터를 사용한 식품 분류 결과는 Table 7과 같이 (1) 곡류군, (2) 과일군, (3) 어육류군. 저지방, (4) 어육류군. 중 지방, (5) 어육류군. 고지방, (6) 우유군, (7) 지방군, (8) 채소군의 8개 식품군에서 총 238개로 나타났으며 예측값과 식품군이 일치하였다.

기존 식품교환표의 식품을 의사결정 모델로 분류한 결과는 모두 8개 식품군에 속하는 것으로 나타났다. 또한, 형태소 분석으로 식별된 식품도 각 식품의 영양소를 기준으로 기존 식품교환표의 범위를 벗어나지 않고 분류되었다. 결과적으로, 식품들이 의사결정 모델을 통해 식품군으로 분류된 것은 새로운 식품에 대한 속성을 기준으로 신속한 분류가 가능하다는 것을 의미한다. 각 식품군에 포함되는 식품 분류 결과는 Table 8과 같으며 지면 관계상 일부만 표기하였다.

Table 7. Food Group and Categorized Foods

No	Food Groups	Cnt.
1	Grain	43
2	Fruit	31
3	Fishes&Meat/Low Fat	35
4	Fishes&Meat/Mid Fat	18
5	Fishes&Meat/High Fat	14
6	Milk	4
7	Fat	29
8	Vegetable	64

V. Conclusions

식품의 다변화에 따라 선별된 식품 섭취는 병원이나 요양원의 환자들에게 치유를 보조하는 중요한 항목 중 하나이다. 또한 건강관리, 다이어트를 위해 식품 선호에 기반한 교환 섭취가 필요하다.

본 연구는 기존 식품과 새로운 식품에 대한 식품군을 분류함으로써 식품교환표에 신속한 식품 추가를 위한 기계 학습 기반의 식품 분류 기법을 제안하였다.

인공지능의 기계학습 중 의사결정트리를 이용한 지도학습으로 식품의 군을 분류하였으며 선행 연구에서 나타난 95.37%의 정확도는 97.45%로 향상되었다. 이처럼 향상된 결과를 얻기 위한 시도로 기존 연구를 확장하여 웹 크롤링으로 수집한 데이터에서 형태소 분석을 사용하여 새로운 음식을 발굴하고 기존 식품 데이터로 학습한 의사결정 모델로 분류하였다. 또한, 의사결정트리에서 최적의 하이퍼 파라미터를 찾기 위해 랜덤 서치를 적용하였다.

Table 8. Food Groups Classification Results

Foods	Protein (g)	Fat (g)	Carbohydrate (g)	Total sugar (g)	Dietary fiber (g)	Category	Food Group	Predicted Group
Cooked glutinous rice, White rice	1.973684	0.065789	21.84211	0	0	Cereal and Product	Grain	Grain
Rice with glutinous rice, Brown rice	1.976048	0.11976	22.33533	0	0	Cereal and Product	Grain	Grain
Rice porridge, White rice	1.971831	0	21.97183	0	0	Cereal and Product	Grain	Grain
Glutinous rice, White rice, Raw	1.763085	0.110193	21.90083	0	0	Cereal and Product	Grain	Grain
...								
Fruits cocktail, Can	0.253333	0.04	24.69333	25.2	1.6	Fruit	Fruit	Fruit
Peach, White peach, can	0.430769	0.030769	24.67692	21.87692	3.230769	Fruit	Fruit	Fruit
Peach, Yellow peach, can	0.382716	0.024691	24.62963	23.22222	3.08642	Fruit	Fruit	Fruit
Pine apple, solid, can	0.397436	0.051282	24.58974	20,26923	2.179487	Fruit	Fruit	Fruit

분류 모델의 높은 정확도는 기존에 식품교환표를 갱신하기 위한 설문 조사, 수작업 분류와 비교하여 신속한 처리가 가능한 것을 나타낸다.

본 연구 결과, 분류 모델을 통한 식품 분류는 기존 식품 분류와 대비하여 수작업을 줄임으로써 소요 시간과 비용의 감소를 가져올 것이다. 본 의사결정 모델은 새로운 식품 분류에도 성능을 발휘하므로 식품 선호도를 식품교환표에 신속히 반영하기 위한 도구로 활용될 수 있다.

본 연구는 2010년에 최종 개정된 식품교환표를 사용하여 최근의 식품을 분류하기에는 한계가 있다. 향후에, 식품교환표가 최신 음식을 포함하는 방향으로 개정된다면 본 연구를 새로운 식품 분류 모델로 확장할 계획이다.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A8027993).

REFERENCES

- [1] K. Lee, S. Kim, and S. Heo, "In-Depth Analysis of Food Consumption in Korea," Korea Rual Economic Institute Reports, pp. 1-254, Oct. 2016.
- [2] S. Hong, "R&D operating strategy for future food industry," Food science and industry, Vol. 53, No. 3, pp. 307-315, Sept. 2020.
- [3] J. Park, "Individualized Medical Nutrition Therapy for Diabetic Patients according to Diabetes Medication," Journal of Korean Diabetes, Vol. 23, No. 1, pp. 50-56, Jan. 2022. DOI: 10.4093/jkd.2022.23.1.50.
- [4] National Nutrition Statistics 2020, Korea Health Industry Development Institute: <http://www.khidi.or.kr>
- [5] J. Yoon, and H. Jang, "Diet Quality and Food Patterns of Obese Adult Women from Low Income Classes," The Korean Society of Community Nutrition, Vol. 16, No. 6, pp. 706-715, Oct. 2011.
- [6] Y. S. Kim, "How to use the food table," Korean Institute of diabetes, Vol. 123, pp. 44-45, 2000.
- [7] D. L. Ju, H. C. Jang, Y. Y. Cho, J. W. Cho, H. S. Yoo, K. S. Choi, M. H. Woo, C. M. Sohn, Y. K. Park, and R. W. Choue, "Korean Food Exchange Lists for Diabetes: Revised 2010," Journal of Korean Diabetes, Vol. 12, No. 4, pp. 228-244, Dec. 2011. DOI:10.4093/jkd.2011.12.4.228
- [8] J. M. Choi, J. D. Kim, C. Y. Park, and Y. S. Kim, "Automatic Word Spacing of Korean Using Syllable and Morpheme," Applied Sciences, Vol. 11, No. 2, pp. 1-10, Jan. 2021. DOI:10.3390/app11020626
- [9] J. Kim, M. Lee, H. Jeon, K. Kim, J. Kim, N. Park, C. Jin, J. Kwon, and J. Kim, "Food Exchange Table Organization Model Based on Decision Tree Using Machine Learning," The KIPS Spring Conference 2020, Korean Information Processing Society, Vol. 27, No. 2, May 2020.
- [10] D. Benware, "Machine Learning Algorithms: How The Machine Learning Algorithms Work Behind The Scenes: Random Forest Algorithm Geeksforgeeks," Independently Published, 2021.
- [11] R. Rafael, C. Juana, M. Efrén, and A. Marco, "Induction of decision trees as classification models through metaheuristics," Swarm and Evolutionary Computation, Vol. 69, No. 1, p. 101006, 2022. DOI:10.1016/j.swevo.2021.101006
- [12] E. Park, and S. Cho, "KoNLPy: Korean natural language processing in Python," 26th Annual Conference on Human and Language Technology, pp. 133-136, Oct. 2014.
- [13] N. Kim, and S. Yoo, "Stock Information Retrieval and Analysis using Korean Natural Language Query," Journal of Korean Institute of Information Technology, Vol. 20, No. 9, pp. 13-18, 2022.
- [14] S. Hong, "Diabetes Nutrition Management and Food Exchange Table," University of Ulsan Publisher, 2019.
- [15] Rural Development Administration, "10th revision Korean Food Composition Table," National Institute of Agricultural Sciences, 2022.
- [16] Decision Tree Libraries of Machine Learning: <https://scikit-learn.org>

Authors



Ji Yun Kim received the B.S. degree in Food and Nutrition from Sahmyook University, Korea, in 2021. Ji Yun Kim has been working as a manager at Wolgye E-Mart branch of Rolling Pasta, THEBORN INC. since Jan. 2022 (www.theborn.co.kr).

She is interested in Food Management Systems using ICT, Pre-processing Food Big Data for Machine Learning.



Jongwan Kim received the Ph.D. degrees in Computer Science and Engineering from Korea University. Dr. Kim joined the faculty of the Smith College of Liberal Arts in 2016 and works for Software Convergence Education

Center at Sahmyook University from 2021, respectively. He is currently a professor and he is interested in Big Data, Machine Learning and Distributed Computing.