

머신러닝을 활용한 선호도 기반 레시피 추천 시스템

김나희¹, 박규리¹, 이민경¹, 현소정¹, 김성욱²

¹서울여자대학교 정보보호학과 학부생

²서울여자대학교 정보보호학과 교수

swumint@swu.ac.kr, xxstargyu@gmail.com, emk0528@swu.ac.kr,

gusth1111@swu.ac.kr, kim.sungwook@swu.ac.kr

Preference-based Recipe Recommendation System Using Machine Learning

Na-Hui Kim¹, Gyu-Ri Park¹, Min-Kyeong Lee¹, So-Jung Hyun¹, Sung-Wook Kim²

¹Dept. of Information Security, Seoul Women's University

요 약

COVID-19 유행 이후 배달 음식의 수요가 증가했다. 이러한 배달 음식은 재료나 영양성분 파악이 어려운 경우가 많아 원하는 음식 조건이 있거나 영양 균형을 필요로 하는 개인이 곤란을 겪는다는 문제점이 있다. 따라서 맞춤형 음식을 원하는 개인이 손쉽게 요리를 할 수 있는 방안을 마련하고자, 여러 머신러닝 알고리즘을 결합한 하이브리드 모델을 이용한 레시피 추천시스템을 구현했다. 구현 후에는 웹사이트를 제작하여 직접 적용해봄으로써 그 활용성을 확인했다.

1. 서론

COVID-19 이후 배달 음식을 애용하는 사람들이 급격하게 증가했고 사태가 완화된 이후에도 배달 음식의 간편성, 현대인들의 바쁜 생활패턴 등으로 인해 사람들의 배달 음식 의존도가 여전히 높다. 이러한 배달 음식 중에서는 건강에 좋지 않은 패스트푸드나 어떤 재료가 들어갔는지, 영양정보가 어떤지 파악하기 어려운 경우가 많다.

이러한 문제를 개선하기 위해 개인 맞춤형 음식의 레시피를 추천해 본인에게 필요한 요리를 간편하게 찾을 수 있도록 해주는 시스템을 구축하였다. 본 연구에서는 재료 유사도와 사용자 선호도에 따른 추천 기능과 카테고리별 검색 기능을 구현해 사용자의 상황에 맞는 다양한 레시피와 정보를 제공함으로써 사용자들의 요구에 부응하도록 했다.

2. 데이터셋

2.1 데이터셋 출처

본 논문에서 사용한 데이터는 식품의약품안전처와 농촌진흥청에서 제공한 공공데이터포털의 조리식품의 레시피 DataBase [1], 이달의 음식 정보 데이터[2]이다.

2.2 데이터 전처리

주어진 데이터셋은 총 1260개의 행으로 구성되어 있으며, 각 행은 레시피에 관련된 정보를 포함하고 있다. 전처리 과정의 주요 목표는 결측값이나 결함이 있는

데이터를 삭제하고, 각 레시피가 특정 종류에 속하는지 여부를 나타내는 새로운 분류 기준을 도입하는 것이다. 이를 위해 총 4가지의 새로운 컬럼 ('LOW_NA', 'HIGH_PRO', 'VEGAN', 'DIETS')을 추가하여 각 레시피가 해당 기준에 부합하는지를 나타내는 이진값(1 또는 0)을 할당한다. 아래 기준표는 식품의약품안전처고시[3]를 참고하였다.

Column	Threshold
LOW_NA	≤500mg
HIGH_PRO	≤55g
VEGAN	No animal-derived ingredients
DIETS	Fat, carbohydrate, and calorie contents standardized and clustered using K-means algorithm

<표 1> 각 영역에 대한 기준

재료 데이터에서 불용어를 제거하고 Doc2Vec을 사용하여 각 행의 단어 특성을 추출하고 문맥 정보를 응축한다.

3. 레시피 추천 서비스 'BOB.D'

메인화면에서는 사용자가 선택한 비건식, 저염식 등

