



파이썬 추천 알고리즘을 활용한 메뉴 선정 프로그램



서론

본 연구는 “오늘 뭐 먹지?” 질문에 대해 명쾌한 해답을 제시해보자는 동기를 갖고 시작되었다. 개인의 식사 패턴, 현재 식사 상황 등의 복합적인 요소들을 고려하여 적절한 선택지들을 제시하는 메뉴 추천 서비스를 제작하는 것으로 목표를 설정하고 본 연구를 진행하였다.



데이터 수집

1. 데이터 수집 방법

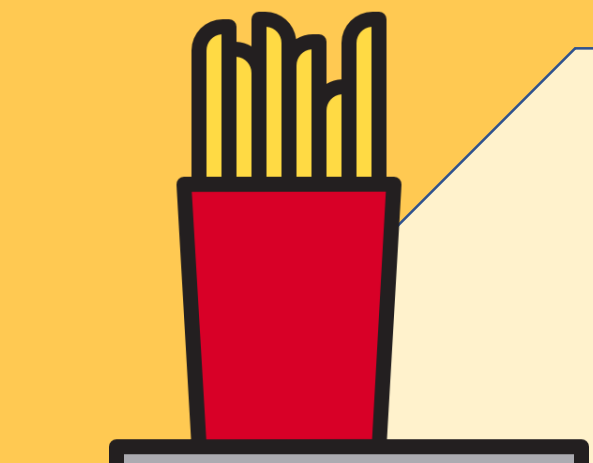
- 초기 데이터**
사용자에게 70여개의 음식들에 대해서 선호도를 조사
- 상황에 맞는 음식 추천**
혼자 먹을 때, 친구와 먹을 때, 데이트할 때 이 세가지의 상황들에 놓여있을 때 주로 먹는 음식들에 대해서 주관식으로 설문 받음
- 추적 데이터**
사용자의 식습관을 분석하기 위해 조원 내부에서 아침,점심,저녁 먹은 음식을 주관식으로 설문 받음.
- 데이터 가공**
주관식으로 응답 받은 데이터들을 활용하기 쉽도록 만들기 위해 기존에 지정했던 음식 이름으로 변경해주는 작업을 거침

2. 레시피 크롤링

- 우리가 지정한 70여종의 음식들에 대해 각각의 레시피를 웹에서 크롤링하여 분석함으로써 음식의 특성 분석

3. Attribute 직접 부여

- 음식 DB 관리자가 각 음식에 대해서 특징을 직접 부여하는 방식
- “뜨거움”, “한식” 등의 다양한 특성들을 직접 입력



웹 & 데이터베이스

• 목적

사용자 맞춤 추천을 위해서는 사용자 구별이 가능해야 했으므로, 로그인/ 회원가입 기능을 구현하였으며 각 사용자 별 데이터(로그인을 위한 데이터 + 음식 추천을 위한 데이터)를 저장하기 위해 MySQL 데이터베이스를 이용하였다.

• 구현

1) DB

GoEat에 해당하는 schema를 만든 후, table import wizard를 통해 자료형과 column 이름을 맞추어 import 하였다. 이를 통해 구현된 DB의 Relational model은 다음과 같다.

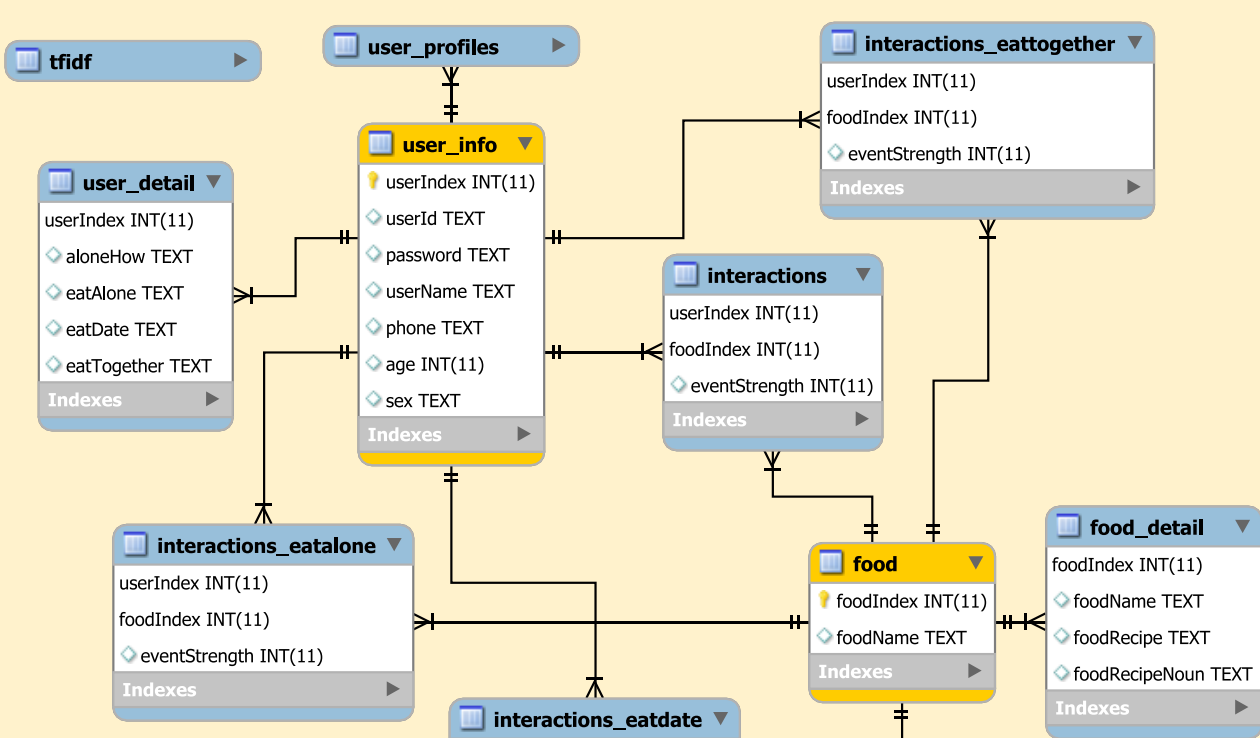


Figure: DB ER Relationship

2) Web

GoEat을 사용자가 편리하게 사용하기 위해 html 기반의 웹으로 프론트를 만들었다. DB와의 연결을 위해 MySQL Connector 라이브러리를 사용하였으며 Tomcat을 이용하여 서버를 구동하고 MySQL 서버와 데이터를 주고 받으며 기능을 구현하였다.

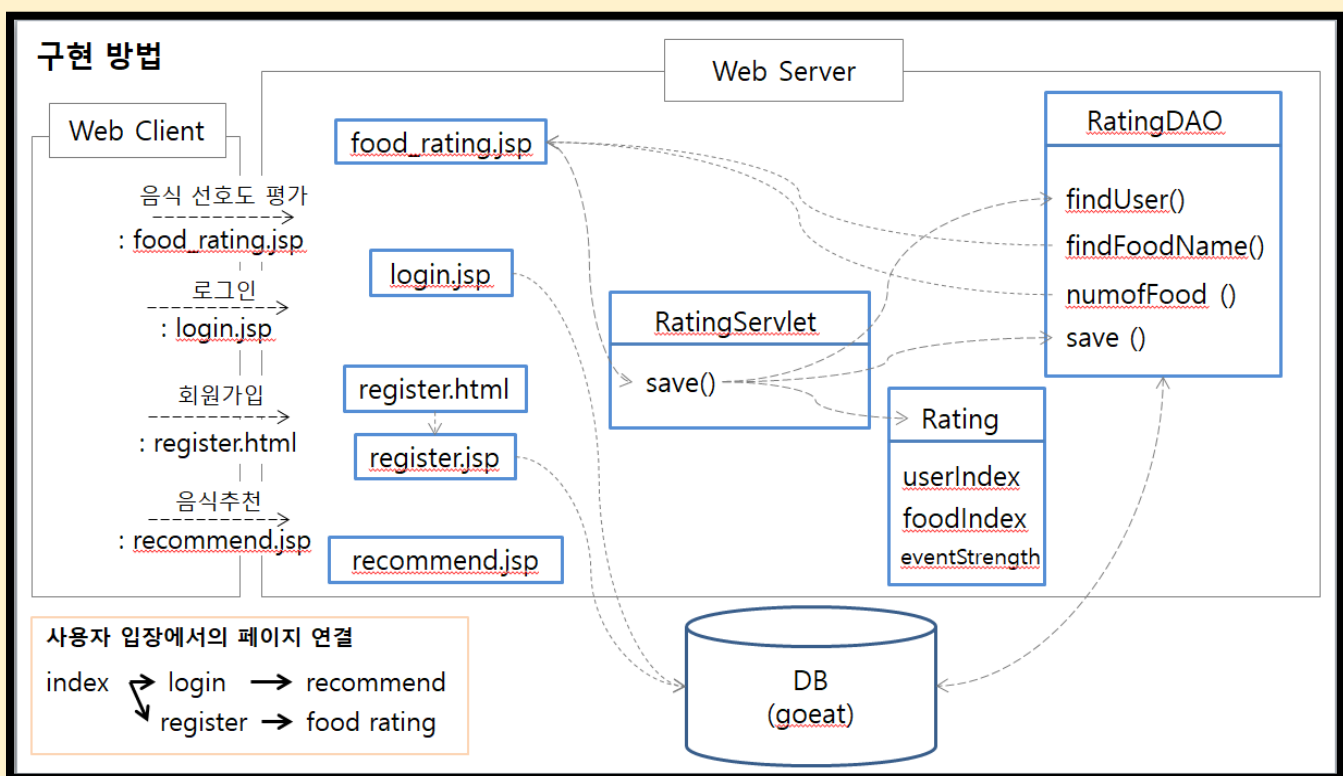
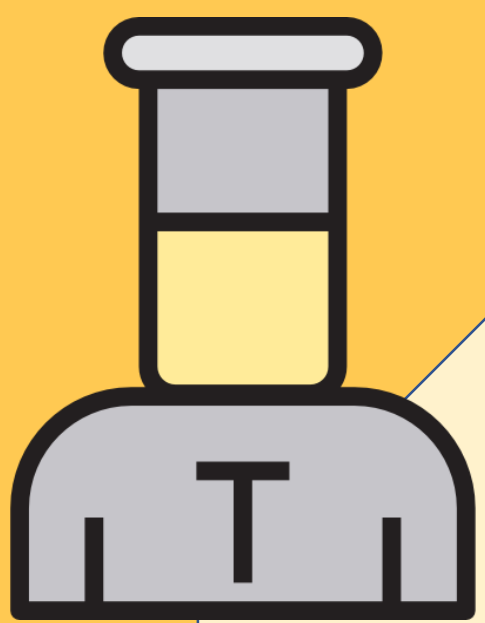


Figure: Web & DB 구현도

Figure: 회원가입 화면

Figure: 음식 평가 화면



추천 알고리즘

1. 추천 시스템

협업 필터링(Collaborative Filtering)

다른 사용자들의 선호나 취향에 대한 정보를 모은 것을 토대로 한 사용자의 관심사를 예측하는 방법이다.

콘텐츠기반 필터링(Content-Based Filtering)

사용자가 이전에 소비한적이 있는 아이템에 대한 설명이나 특성에 대한 정보를 이용하여 사용자의 선호를 모델링하는 방법이다.

융합형 필터링(Hybrid Filtering)

협업 필터링과 콘텐츠 기반 필터링을 합하여 추천결과를 내는 알고리즘이다.

2. 협업 필터링 구현

특이 값 분해를 이용한 협업 필터링은 원본 사용자-선호도 행렬 R을 다음과 같이 분해 한다.

$$R \approx U \cdot \Sigma \cdot V^T$$

U : 사용자가 각각의 속성을 얼마나 좋아하는지를 나타내는 행렬
 V^T : 속성들이 음식과 얼마나 연관이 있는지를 나타내는 행렬
 Σ : 속성들의 가중치에 대한 직교행렬

Σ 의 요인 개수가 더 많을 수록 원본 행렬에 더 가깝게 복원 되지만 새로운 데이터에 대해서는 예측이 어려우며, 요인의 개수가 적을 수록 더 일반화가 많이 일어난다. 따라서 적정 요인 개수를 찾는게 중요하다. 우리는 Grid Search를 통해 적정 요인개수를 찾아보았다.

3. 콘텐츠 기반 필터링 구현

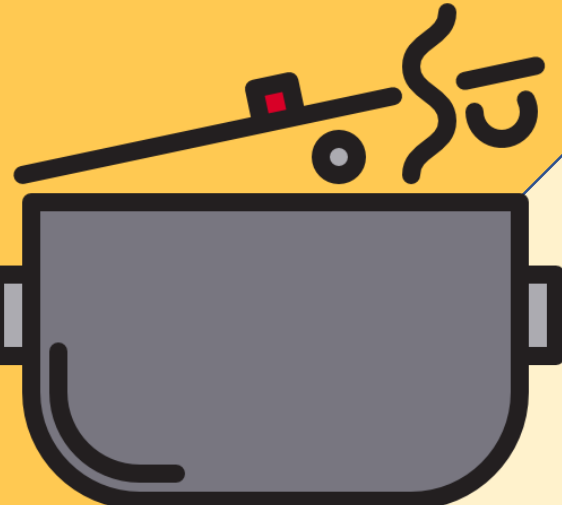
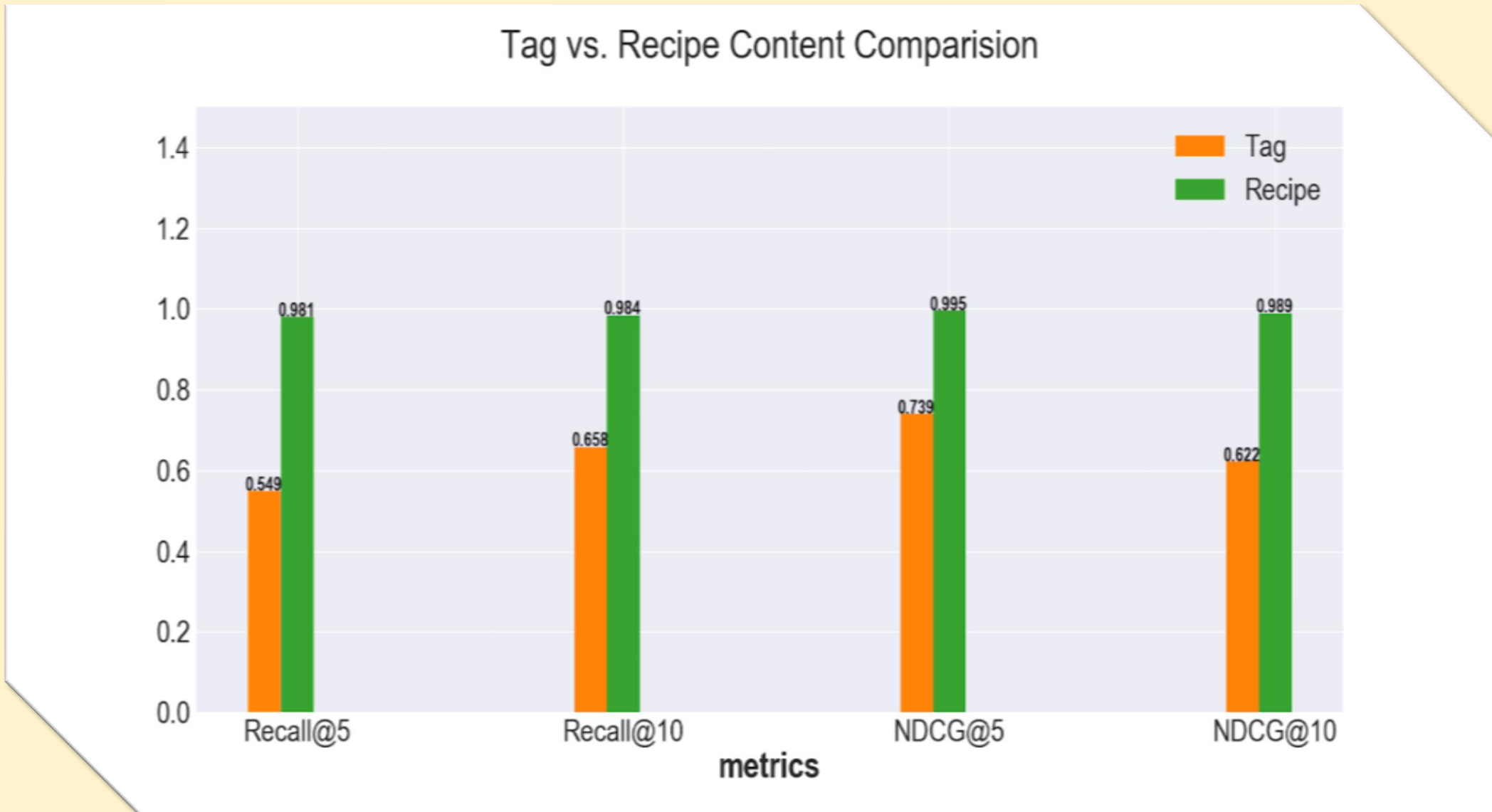
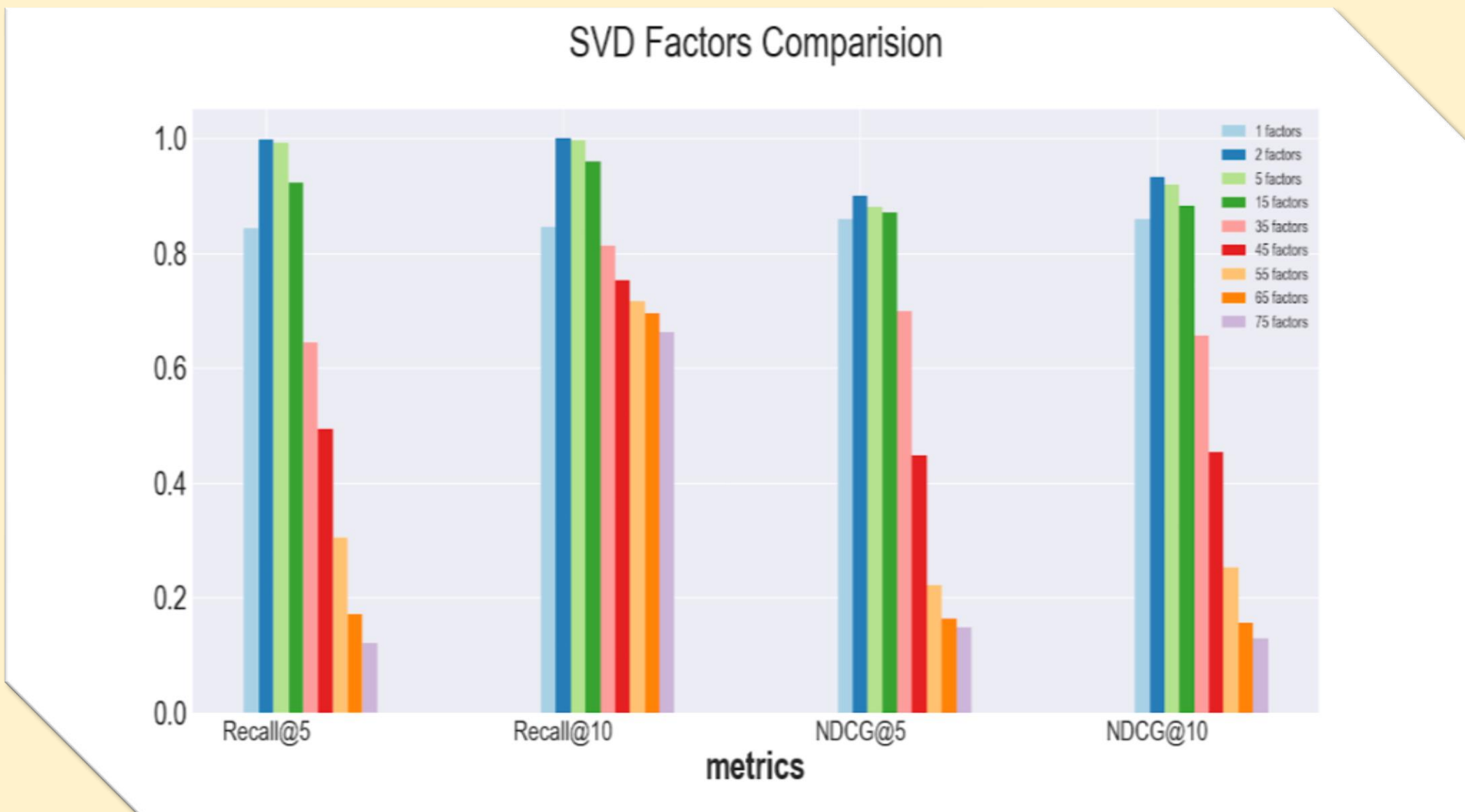
TF-IDF는 여러 문서로 이루어진 문서군에 대하여 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 수치이다. 이를 사용하여 음식의 속성을 정의하였으며 사용자가 해당 음식을 얼마나 선호하느냐에 가중치를 부여하여 사용자 프로필을 만들었다.

음식의 레시피는 웹에서 크롤링하여 명사만 추출해 사용하였다.

4. 전체 모델 비교

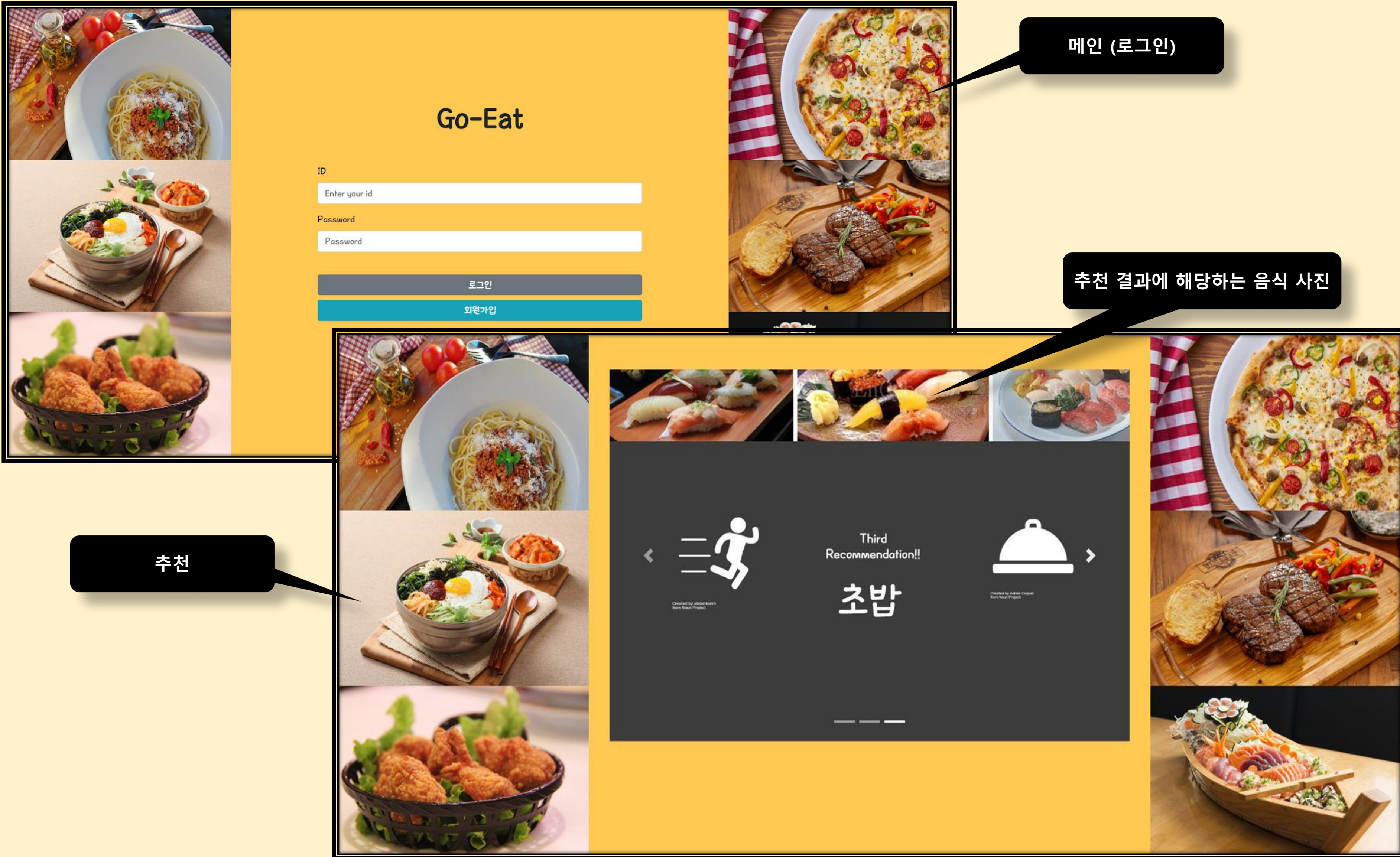
협업 필터링과 콘텐츠 기반 필터링을 기반으로 둘의 값을 단순 곱한 융합형 모델과 조화평균을 낸 융합형 모델을 만들어 내고, 그 결과를 비교하였다.

음식 추천 시에 있어서 가장 중요한 것은 높은 순위의 음식을 얼마나 잘 맞추느냐 이다. 조화평균 융합형 모델이 다른 모든 수치에서 순수 모델들 보다 정확도가 좋지 않았지만, NDCG@5에서 더 좋은 정확도를 보였기 때문에 최종 모델로 조화평균 융합형 모델을 사용하였다.



결론

TF-IDF를 통해 음식의 레시피 문서에서 중요 단어를 파악하고 이를 바탕으로 각 메뉴에 대한 속성을 추출해내는 방식에 차별점을 두었다. 추천 알고리즘으로는 각각 협업 필터링, 콘텐츠 기반 필터링, 융합형 필터링을 적용하였다.



메인 (로그인)

추천 결과에 해당하는 음식 사진

추천

추후 발전 방향

- 앱 개발을 통해 웹 뿐만 아니라 앱 기반으로도 서비스를 제공하여 서비스의 접근성을 높인다.
- 실제로 서비스를 하면서 더 많은 사용자 데이터를 확보하고, 이를 바탕으로 추천 알고리즘의 성능을 개선해 나갈 것이다.
- 감정, 기온 및 날씨 등과 같은 추가적인 변인들을 적용하여 더욱더 정교한 추천이 이루어질 수 있도록 개선한다.

활용 가치

본 연구를 통해 얻게 된 음식 분석 데이터들은 여러 분야에서 활용될 수 있다는 점에서 그 가치가 높다. 식생활 패턴을 분석을 이용해 사업적인 활용이 가능할 것으로 기대한다. 특히 배달 앱 등에서 음식 추천과 판매 기능을 조합하여 유용하게 활용할 수 있을 것으로 전망한다. 또한 본 알고리즘을 토대로 음식 뿐만 아니라 다른 상품들도 사용자에게 적절하게 추천해줌으로써 사용자의 의사 결정을 돕고 만족도를 증대시킬 수 있을 것이다.