**오픈소스전문프로젝트**

데이터 수집과 분석 계획

**수집 데이터셋**

1) 음식점 데이터

전국 음식점 데이터를 한식, 일식, 중식, 분식, 카페 등으로 분류하여 geocode와 주소, 상호명, 시설총규모 등의 데이터를 포함한다

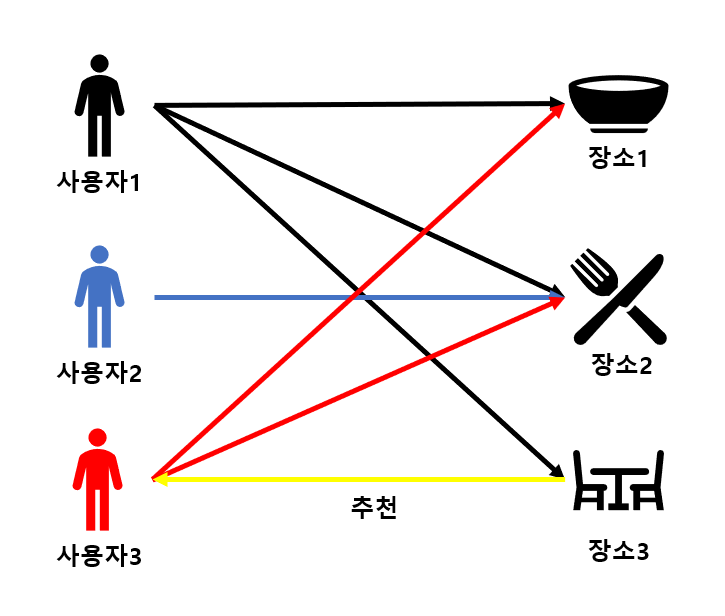
2) 사용자 데이터

어플리케이션 사용자에 대한 정보를 데이터 분석에 활용하기 위해 성별, 연령 등에 대한 정보를 저장

3) 사용자 리뷰 데이터

사용자가 약속장소로 선택한 장소에 대한 리뷰, 평점 등에 대한 내용을 데이터로 활용

**활용방안**

사용자 데이터를 기반으로 특정한 연령, 성별 등 모임이 선호하는 약속장소를 분석하여 또다른 사용자에게 추천할 수 있도록하는 음식점 데이터와 연계한다. 이를 통해 사용자들을 특정한 기준으로 분류하고, 분류 집단이 선호하는 매장을 약속장소 주변에서 찾아내어 추천할 수 있다.  
 모임 이후에 약속장소에 대한 리뷰 및 간단한 평점을 매기도록 유도하여 리뷰 데이터를 수집하고, 이를 기반으로 주변에서 긍정적 평가가 높은 매장을 추천할 수도 있다.  
 음식점 데이터의 시설총규모는 매장의 크기로써, 많은 인원이 모일 때 그 인원들을 수용할 수 있는지에 대한 여부를 판단할 수 있는 간접적인 기준이 되어, 일정 인원 이상의 규모있는 모임에는 시설총규모가 작은 매장은 추천하지 않도록 할 수 있다.

**활용 예시**

수집한 데이터는 위의 그림과 같이 사용자 기반의 협업 필터링을 통해 장소를 추천하는 방법을 적용하는 추천시스템을 구현할 계획이다. 비슷한 연령, 성별, 지역에 따라 사용자 그룹을 형성하고, 그룹 내 사용자1이 방문한 장소와 사용자3이 방문한 장소에 대한 패턴이 유사할 때, 사용자3에게 사용자1이 다녀왔던 장소를 추천한다. 즉, 사용자 간의 이용패턴의 유사성에 근거하여 장소를 추천하는 것이다. 추가로 각 사용자의 재방문 횟수, 장소에 대한 리뷰 등을 추가로 반영한다면, 좀 더 정밀한 추천 시스템의 구현이 가능할 것으로 사려된다.

**분석계획**

현재 단계에서는 시설총규모를 통한 인원수에 따른 장소추천 기능을 수행할 수 있을 것으로 보인다.

위의 데이터들을 활용한 서비스가 원활하게 이루어지기 위해선 최신화 된 음식점 정보와 사용자의 연령, 성별 등에 대한 정보가 확보되어야 하며, 사용자들의 활발한 어플리케이션 활용을 통해 충분한 음식점 리뷰, 선호매장 등 소비자데이터를 확보되야한다.

리뷰와 선호매장에 대한 부분은 실제 어플리케이션 사용자를 충분히 확보할 방법이 없어 현재 단계에선 활용할 수는 없지만, 충분한 사용자가 확보되어 선호매장 등의 데이터를 수집할 수 있다면, “**20대 여성**들의 모임은 주로 **카페**에서 모이며, **OOO매장**을 많이 선호한다.” 같은 정보를 도출할 수 있습니다.

**한계점**

분석계획에서 서술한 바와 같이 선호매장, 리뷰 데이터를 분석하는 경우, 충분한 사용자와 사용기록이 요구되므로 현 단계에서 데이터를 수집할 수 없다. 때문에 활용계획에 어려움이 있다.

시설총규모의 경우 매장의 크기만을 나타내는 데이터이므로, 실시간으로 매장에서 수용할 수 있는 인원에 대한 데이터로는 사용할 수 없다. 이러한 신뢰성의 문제는 매장에서 직접 수용가능 인원에 대한 정보를 어플리케이션에 제공하거나, 매장이 만석인 경우 장소 선택 시에 만석임을 알릴 수 있도록 표기하는 것으로 해결할 수 있을 것이다.

위에서 준비한 추천기능의 경우 사용자 개인의 알레르기, 식성, 취향 등으로 인해 방문하지 못하는 매장을 추천하는 문제가 발생할 수도 있다. 이러한 문제점을 해결하기 위해 사용자 정보에 비선호매장, 음식, 장소 등을 추가로 입력한다면 좀 더 사용자 요구에 맞는 장소를 추천할 수 있을 것이다.