**[ 최종 프로젝트 발표 구성안 ]**

|  |  |  |  |
| --- | --- | --- | --- |
| **훈련 과정명** | (빅데이터분석)머신러닝(MachineLearning)기반의 빅데이터분석과정 A | | |
| **훈 련 기 간** | 2021-09-08 ~ 2022-01-20 (720시간 / 90일) | | |
| **작 성 일 자** | 2022. 01. 19. | **발 표 일 자** | 2022. 01. 20. |

|  |  |  |
| --- | --- | --- |
| **팀 명** | 2팀 | |
| **구 성 원** | **팀 장** | 목지윤 |
| **팀 원** | 김선민, 이재은, 이종호, 홍현 (가나다 순) |
| **발 표 자** | 이재은 | |
| **프로젝트 주제** | 본 기획은 다수의 인원이 약속 장소를 선정할 때 거리 기반으로 합리적 중앙 위치를 선정하여 사용자에게 알려주고, 별도로 가게를 검색하지 않아도 널리 알려진 서비스 기반으로 이미 평가가 검증된 곳을 다수 추천하여 택일할 수 있도록 도움을 주는 서비스이다. | |
| **프로젝트 발표**  **구성 및 내용** | 1. 주제 선정 배경  중간 지점을 찾아주는 기능과 맛집을 추천하는 기능을 접목한 '중간지점 산출+맛집 추천' 서비스로, 오프 모임을 갖는 다수의 사용자에게 필요한 각종 검색 시간을 줄여준다.  2. 데이터 수집  [1] 망고플레이트 가게 데이터  [2] 카카오맵 가게 리뷰 데이터 : 평점 예측 모델 학습용  3. 개발 과정  [1] 리뷰 평점 예측 모델  - 리뷰 자연어 처리: 카카오 맵에서 수집한 리뷰, 평점 데이터를 다음과 같은 과정을 거쳐 전처리 하였다.      [데이터 정제-> 띄어쓰기 단위로 분리-> 토큰화, 패딩]  - LSTM 모델 학습 및 평가: LSTM을 이용해 모델을 생성하고 전처리한 데이터로 학습 및 평가하였다.      [테스트 정확도: 0.948]  [2] 약속장소와 맛집 추천  - 가게 위치 클러스터링 : 밀도기반 클러스터링(DBSCAN)을 이용해 가게의 위치를 군집화하였다.  - 중간 지점 산출 : 위도,경도 좌표를 이용해 각 출발지점으로부터 중간지점을 산출하였다.  - 약속장소 추천 : 중간지점으로부터 가장 가까운 세 클러스터를 선택하였으며, 세 클러스터는 추천하는 약속장소가 된다.  - LSTM 모델을 이용해 누락된 평점 예측 : 망고플레이트 데이터에는 평점이 누락된 가게가 많아 앞서 생성한 LSTM 모델을 통해 평점을 예측하여 추가하고, 위치 좌표 및 클러스터 인덱스를 포함한 새로운 가게 데이터프레임을 만들었다.  - 평점 상위 5개 맛집 추천 : 세 클러스터 각각의 가게에서 평점 상위 다섯 곳을 뽑아 총 15곳의 맛집을 추천한다.  4. 구동 예시  5. 한계점 및 향후 개발방향  -한계점-  [1] 서울로 한정된 데이터 : 서울로 한정하여 다른 지역에는 적용할 수 없다.  [2] 망고플레이트 맛집 데이터 : 망고플레이트는 맛집 추천 사이트이기 때문에 리뷰와 평점이 좋아 평점 예측시 대부분 5점으로 높게 예측되었다.  -향후 개발방향-  [1] 웹/앱 개발:  웹사이트나 앱에서 이용할 수 있도록 Java Script/Flash 혹은 안드로이드 스튜디오를 이용해 개발  [2] 망고플레이트 데이터 수정 :망고플레이트 외에 다른 사이트도 이용하여 데이터의 범위 확장 | |
| **팀별 발표순서** | 2번 | |
| **발표 예상 시간**  **(준비 및 Q&A 포함)** | 15분 | |
| **비 고**  **(특이 사항)** | 중간 위치는 한 장소만 선정하는 것이 아니라 세 곳을 추천하여 선택의 다양성을 높였으며, 맛집 평점의 경우에는 누적 사용자와 리뷰수 등에서 신뢰도가 높은 ‘망고플레이트’ 서비스를 기반으로 상위에서 출력 결과를 필터링 해 만족도를 올렸다. | |