**오라이 최종발표 대본**

안녕하십니까 융복합 프로젝트 2조 발표 시작하겠습니다. 저희는 여름철 아이들 물놀이장 정보 제공 서비스를 준비해보았습니다. 그럼 목차부터 설명드리도록 하겠습니다.

(슬라이드 넘김)

발표는 크게 기획, 기능, 구현, 시연 이렇게 4가지로 나누어 보여드리려고 합니다. 첫번째 기획 부문에서는 주제를 선정하게 된 배경과 서비스의 타겟층, 그리고 기대효과에 대해 말씀드리겠습니다. 두번째 기능 부문에서는 데이터 분석과 AI 파트로 나누어, 저희 서비스에 적용된 각 분야의 기능에 대해서 설명드리겠습니다. 세번째 구현 부문에서는 클라우드와 웹 파트로 나누어 서비스 설계 과정과 그 결과에 대해 말씀드리겠습니다. 그리고 마지막으로 시연 부문에서는 서비스 시연 영상을 보여드리려고 합니다.

(슬라이드 넘김)

네 첫번째로 서비스 기획 과정에 대해 말씀드리겠습니다.

(슬라이드 넘김)

먼저 주제 선정 배경에 대해 말씀드리겠습니다. 현재 코로나 19로 잠시 가동을 멈췄던 우리 동네 물놀이장이 각 지자체의 운영 아래 가동되고 있습니다. 기존에 있었던 공원 내의 놀이터가 여름철에는 물놀이장으로 바껴서 운영되는 곳이 많은데요, 아이를 데리고 멀리 나가려면 챙겨야 할 것도 많고 또 요즘처럼 무더운 날씨에는 멀리 나가는 것 자체가 부담이 되는 일이기 때문에 이렇게 큰 맘 먹고 여러가지를 준비하고 나가지 않아도 되고 게다가 이용 요금도 무료인, 접근성이 뛰어난 동네 물놀이장을 선호하시는 분들이 많습니다.

(슬라이드 넘김)

실제로 여름철에 물놀이장을 가려고 검색하시는 분들이 얼마나 많은지, 검색 키워드량 증가율에 관해서 찾아보았습니다. 이틀 전인, 8월 6일 정오 기준 월간 검색량은 6만 6백건에 달했고, 연령별 검색 비율은 자녀를 키우고 있을 확률이 높은 30대와 40대가 도합 85% 정도의 비율을 차지하는 것을 볼 수 있습니다. 또한 여름철에 물놀이장을 검색하는 횟수가 지속해서 증가 추이를 보이고 있는 것을 확인할 수 있습니다.

(슬라이드 넘김)

또한 아이들과 함께 가기 좋은 물놀이장을 추천해달라는 글도 많이 보입니다. 화면의 글은 이틀 전 정오 기준 네이버 지식in 사이트에 서울 물놀이장을 검색했을 때 나오는 질문들을 스크랩한 자료입니다. 이렇게 여름철에 아이들과 함께 갈만한 물놀이장에 대한 수요도 매우 많은 것을 확인하였습니다.

(슬라이드 넘김)

하지만 이렇게 물놀이장 정보에 관한 수요는 많지만, 반대로 정보에의 ‘접근성’은 매우 낮은 실정입니다. 각 행정구역 별로 물놀이장의 운영공지가 올라오는 사이트도 다르고, 또한 지자체에서 운영하는 물놀이장이 아닐 경우에는 해당 행정구역의 웹 페이지에 들어간다고 해도 정보를 얻을 수 없는 경우도 있습니다. 운영시간은 몇시부터 몇시까지인지, 개장과 폐장일정은 어떻게 되는지, 탈의실이나 샤워시설이 있는지, 우리 집 근처에 물놀이장은 어디어디 있는지 한 번에 알아볼 수 있는 정보 제공처가 없는 것을 파악하고 저희는 그 부분에서 니즈를 찾았습니다.

따라서 저희는 여름철 아이들 물놀이장 정보 제공 서비스를 생각하였습니다.

(슬라이드 넘김)

저희 서비스의 타겟층은 이러합니다. 다양한 운동 기능이 발달하여 몸의 균형이 증가하는 세 살부터 초등 고학년이 되기 전인, 열 살까지의 자녀를 둔 부모 또는 보호자를 타겟으로 삼았습니다. 서비스 제공 지역은 현재는 서울특별시 내에 위치한 물놀이장에 대해서 먼저 베타 서비스할 예정이고요, 후에 전국으로의 확장 가능성을 염두에 두고있습니다. 또한 물놀이장의 종류는 동네 물놀이장과 바닥분수, 그리고 한강 야외 수영장. 이렇게 세 가지 종류로 서비스 할 계획입니다.

(슬라이드 넘김)

저희 서비스의 기대효과는 이러합니다. 첫번째로는 기존에 없던 서비스라는 점입니다. 거기에 사용자 위치 정보를 이용하여 현재 위치 주변의 장소 검색이 가능하게끔 해주었고, 물놀이장에 대한 기본 정보와 함께 근처의 편의시설 정보도 함께 제공해줍니다. 또한 저희 서비스의 차별점은 물놀이장에 대한 편의지수를 도출하고, 범주화하여 편의 등급을 신호등의 형태로 보여준다는 점입니다. 또한 인공지능을 이용해 리뷰 데이터를 분석한 결과를 함께 제공할 예정입니다. 또한 유저도 평점과 리뷰를 남길 수 있고, 마음에 드는 장소를 북마크 할 수 있는 기능을 포함하여 유저들이 지속적으로 서비스를 이용할 수 있도록 하였습니다.

(슬라이드 넘김)

그렇다면 이제 앞서 말씀드렸던 저희 서비스의 차별점이 될, 데이터 분석과 AI 기능에 대해서 말씀드리도록 하겠습니다.

(슬라이드 넘김)

먼저 데이터 분석을 통한 편의지수 개발 과정에 대해 말씀드리겠습니다.

(슬라이드 넘김)

저희는 첫번째로 인덱스 도출에 필요한 데이터 마트를 구축해주었습니다. 저희가 서비스할 물놀이장 3종류에 대한 데이터를 수집하였고, 편의지수를 도출하는데 시설 간의 거리를 계산하는 과정이 필요했기 때문에, 위도와 경도 좌표 데이터를 가지고 있는 편의시설 데이터를 수집하였습니다. 또한 인덱스 도출 관련해서 레퍼런스가 될만한 기존의 자료와 선행 연구 논문을 수집하였습니다. 이렇게 수집된 약 70만개의 데이터를 가지고 결측치를 제거하는 등의 기본적인 전처리 과정을 진행하여 데이터를 1차적으로 정제해주었고, 그 뒤에 좌표 간의 거리를 계산하기 위해 하버사인 함수를 이용하여 기존 데이터프레임에 파생변수를 생성해주었습니다. 또한 인덱스 도출과정에 필요한 값을 얻기 위해 선형 회귀 모델을 이용한 예측 데이터 또한 가공열로 추가하여 지수 개발을 위한 데이터 마트를 구축해주었습니다.

(슬라이드 넘김)

좌측의 장표는 저희가 수집한 데이터의 리스트입니다. 아래와 같은 플랫폼에서 데이터를 수집하여주었고, 오른쪽은 그 rawdata의 모습인데요, 빨간색으로 강조된 표시와 같이, 수집한 약 70만개의 데이터 모두 위도와 경도 좌표를 가지고 있는 데이터들로 수집하여 주었습니다.

(슬라이드 넘김)

그리고 하버사인 함수를 이용하여 각 장소 별 좌표 간 거리를 계산해주었는데요, 화면에 보이는자료는 편의지수 개발 전, 전처리가 완료된 최종 데이터프레임의 형태입니다.

(슬라이드 넘김)

이제부터는 저희 팀의 인덱스 개발 과정에 대해 말씀드리려고 합니다. 저희는 편의지수를 개발하기 위해 기존 연구 논문들을 참고하였는데요, 레퍼런스가 된 연구 모형들은 좌측의 노선가식 평가방법의 접근 계수 공식과 우측의 아오야마 연구 모형이 있습니다. 이 모델들이 시사하는 점은 아래와 같습니다. 개인이 시설을 이용함으로써 얻는 가치는 해당 시설 서비스의 이용가치로부터 이동으로 잃는 가치의 차이로 나타나는데, 이 가치가 0 이하가 될 때, 이론적으로 시설은 장점을 잃어 이용하지 않게 된다는 것입니다. 그렇다면 얼마만큼의 거리까지가 효용가치가 0이 되기까지의 한계 거리일까요? 거기에 서비스의 이용가치를 산정하기 위한 단위거리 당 심리적 부담은 어떻게 구하면 좋을까요? 저희는 이것을 고민하며 오른쪽의 아오야마 모형을 토대로 편의시설 접근성을 연구한 선행 논문을 수집하였습니다.

(슬라이드 넘김)

해당 논문에서는 20대부터 50대까지 2250명을 대상으로 설문조사를 진행하여, 각종 편의시설에 대한 만족거리 평균값과 k값을 구하였습니다. 해당 논문에서 정의한 만족거리와 k값의 의미는 다음과 같습니다. 만족거리는 시설로부터 멀어질수록 이용자의 시설 만족도라는 효용이 떨어져 불만을 일으키게 되기까지의 거리입니다. 그리고 K값은 시설의 만족거리에 따른 민감도입니다. K값은 값이 작을수록 거리가 멀어짐에 따라 만족도가 급격하게 감소하는 양상을 띕니다. 저희는 이 설문조사 데이터를 참고하여 저희 서비스가 제공하는 편의시설들에 대한 만족거리와 K값을 산정하는 기초 토대로 사용하였습니다.

(슬라이드 넘김)

이 과정에서, 저희가 서비스 하는 장소는 ‘아이들’이 이용하는 물놀이장이라는 점을 고려하여 저희 주제에 맞게 몇 가지 변형시켜주었습니다. 어린이 이용가능 화장실과 수유실에 대해서는 현재 행정안전부가 지정하는 어린이와 노인 및 장애인 보호구역의 지정 및 관리에 관한 규칙에 제정되어있는 기준을 참고하여 어린이 보행속도 0.7m/s를 고려하여 만족거리, 즉 최대 한계 거리를 8분을 곱하여 300m로 산정하였고, 기존 설문조사 데이터에 포함되어있지 않았던 편의시설인 주차장의 경우에는 성인의 보행속도와 다른 편의시설의 만족거리를 고려하여 10분을 곱하여 600m로 정하였습니다.

(슬라이드 넘김)

또한 이렇게 임의로 주제에 맞게 데이터를 변형시켜주었기 때문에, 그에 맞춰 K값도 선형 회귀 모델을 이용하여 저희가 저희 주제에 맞춰 새롭게 정한 만족거리에 해당하는 K값을 예측하여 사용하여주었습니다. 좌측의 그래프는 만족거리와 K값의 산점도와 회귀선을 나타낸 그래프입니다. 그래프를 통하여 데이터의 분포와 회귀선을 그려본 다음, 머신러닝을 통하여 회귀 모형을 만들고, 저희가 지정한 만족거리에 따른 K값을 예측하여 도출해주었습니다. 저희는 이 K값을 편의시설 별 가중치로 설정해주었는데요, K값이 작을수록 만족도가 감소하는 경향을 띄어야 하기 때문에 저희는 그 역수인 1/K 을 계수로 사용해주었습니다.

(슬라이드 넘김)

저희의 편의지수 도출 알고리즘을 요약하면 다음과 같습니다. 물놀이장 1개와 편의시설 1개에 대한 개별 점수를 도출한 후 합산하고, 각 편의시설 별 점수를 정규화 한 다음 시설별 민감도 (계수)를 곱해주었습니다. 그리고 도출된 모든 물놀이장의 점수를 0부터 100 사이의 값으로 다시 한번 표준점수화 해주었습니다.

(슬라이드 넘김)

저희는 이렇게 도출된 편의지수를 범주화하여 총 3단계의 등급으로 나누어주었는데요, 이를 위해 먼저 우측의 커널 밀도 함수 그래프를 그려서 점수의 분포를 확인해주었습니다. 그리고 기초 통계량을 확인하여 범주화 기준을 정해주었는데요, 가장 하위의 등급인 soso 등급은 최솟값부터 제 1 사분위수까지, 중간 등급인 good 등급은 제 1 사분위수부터 중위수까지, 상위 등급인 Excellent 등급은 중위수부터 최댓값으로 범위를 나누어 줬고, 각 등급은 물놀이장을 지도에 나타내는 마커에 색깔로 표시해주어 사용자가 저희 서비스에 처음 들어와서 보이는 메인페이지에서부터 장소의 마커만 보고도 그 색깔로 편의지수 등급을 한 눈에 파악할 수 있게 해주었습니다.

(슬라이드 넘김)

그리고 도출된 편의지수 결과를 가지고 또 여러가지 분석과, 유저에게 보여줄 수 있는 시각화 부분을 고안해보았습니다. 왼쪽은 서울특별시 내 행정구역 별 물놀이장의 점수 평균을 나타낸 것입니다. 그 결과 종로구와 중구가 1,2위를 보이는데요, 아마 유동인구가 많고 문화 유적 및 관광지가 많아 상대적으로 편의시설이 많이 위치해 있어 높은 점수를 얻었으리라고 추측됩니다. 오른쪽은 folium 라이브러리를 사용하여 지도상에 시각화를 해보았습니다. 지도 상의 검은색으로 표현된 자치구는 데이터의 개수가 많지 않고 분포상 큰 차이를 보이기 때문에 제외하여주었습니다.

(슬라이드 넘김)

다음으로는 유저에게 보여줄 랭킹 자료입니다. 편의지수를 토대로 서울특별시 내의 물놀이장 편의지수 랭킹을 나타내보았습니다. 현재 화면에는 서울특별시의 전체 랭킹이 나와있는데, 후에 개별 유저의 위치에 따라 해당 행정구역의 편의지수 랭킹을 보여주는 기능도 추가 가능하겠습니다.

(슬라이드 넘김)

다음으로는 레이더 차트를 이용하여 편의지수를 상세화하였습니다. 유저는 이 그래프를 보고 해당 장소가 어떤 부분에서 높은 점수, 혹은 낮은 점수를 얻었는지 알 수 있고, 본인이 중요하게 생각하는 편의시설에 대해서 세부적으로 파악하기도 용이합니다. 또한 편의지수를 도출하는 알고리즘에 대한 설명을 세세하게 글로 적어놓지 않아도 이 그래프를 통해 점수가 어떻게 도출되었는지 대략적으로 파악할 수 있겠습니다.

(슬라이드 넘김)

다음으로는 저희 서비스에 들어간 AI 기능에 대해 말씀드리겠습니다.

(슬라이드 넘김)

저희는 자연어처리 기술을 사용하여 긍정과 부정을 분류하는 감성분석 기능을 넣어주었습니다. 저희는 이 기능을 통해 물놀이장의 리뷰들의 긍정과 부정 비율을 도출하고자 했는데요, 먼저 구글 지도에 나와있는 모든 물놀이장 장소에 대한 리뷰를 크롤링하여 수집하였습니다. 이후 수집한 모든 raw데이터에 라벨링을 진행하였습니다. 그리고 긍정데이터와 부정데이터의 개수의 비율을 맞춰서 학습 데이터를 구축하고 전처리해주었습니다. 그 이후 KoNLPy의 Okt 모델을 통해 형태소를 분석하고, 데이터를 토큰화해주었습니다.

(슬라이드 넘김)

이렇게 토큰화 작업을 끝내고 정수 인코딩과 패딩 작업을 거쳐 입력 데이터를 구축하고, Keras의임베딩 층을 입력층으로 사용하였습니다. 그리고 은닉층에 LSTM 모델을 적용해주었고, 시그모이드 함수를 이용하여 결과값을 도출하였습니다. 그 결과 모델은 약 81%의 정확도를 보였습니다.

LSTM : RNN의 한 종류, 기존의 RNN 보다 더 과거의 데이터를 고려하여 미래 데이터를 예측

(슬라이드 넘김)

화면에 보이는 캡처 파일은 저희 딥러닝 모델에 리뷰 데이터를 적용해본 결과입니다. 이번 프로젝트에서는 구글 맵의 리뷰를 가져와서 사용하였지만, 실제 서비스 단계에서는 사용자가 직접 작성한 리뷰 데이터도 포함하여 긍/부정 비율을 도출할 수 있겠습니다.

(슬라이드 넘김)

그리고 저희는 위에서 수집한 리뷰 데이터들을 이용해서, 자주 언급되는 키워드를 추출해서 그것을 워드클라우드로 시각화하는 기능을 넣어주었는데요, 위의 긍/부정 분류에서는 토크나이즈 과정에서 KoNLPy의 Okt 모델 중에 형태소 분석기를 사용해주었지만, 이번에는 명사만을 추출하는 알고리즘을 사용해주었습니다. 그리고 그 과정에서 추출된 특정 키워드들은 불용어로 처리하여 제외해주었는데요, 기본적으로 유저는 아이와 함께 갈만한 물놀이장 장소를 찾기 위해 저희 서비스를 이용하는 것이기 때문에 아이, 공원, 물놀이 등의 키워드를 보는 것을 원하지 않을 것이라고 판단하였습니다. 따라서 저희는 유저의 입장에서 생각하여, 니즈에 맞지 않는 단어들은 제외하고 워드클라우드를 생성해주었습니다.

(슬라이드 넘김)

다음으로는 서비스 구현 과정에 대해 말씀드리도록 하겠습니다.

(슬라이드 넘김)

첫번째로는 클라우드 설계 구성도에 대해 설명드리도록 하겠습니다,

(슬라이드 넘김)

클라우드는 2개의 가용영역(Availability zone)으로 나누어져 있습니다. 각 가용영역 public subnet에는 리액트를 배포할 ec2와 rds 관리를 위한 bastion host로 구성되어 있고, private subnet에는 장고서버를 배포할 2개의 ec2와 rds로 구성되어 있습니다. 사용자는 1번을 통해 웹사이트의 화면을 전달받고 2번을 통해 데이터 분석 및 AI 딥러닝 결과를 axios 형태로 호출합니다.

(슬라이드 넘김)

이번 구성도는 현재의 기본구조에 github Actions를 추가한 구성도로 github respository에 개발과정을 push하면 github actions을 통해 work flow를 자동화 할 수 있도록 설계할 예정입니다. 또한 EKS 클러스터를 추가하여 무중단 서비스를 제공할 예정입니다.

(슬라이드 넘김)

그리고 저희 웹 서비스 구현 과정과 결과에 대해 말씀드리도록 하겠습니다.

(슬라이드 넘김)

먼저 저희가 설계한 DB ERD 입니다. 저희 서비스의 데이터베이스는 MySQL을 사용하였으며, 오른쪽에 빨갛게 강조 표시된 물놀이장 장소 테이블과 각 편의시설 별 데이터 테이블 사이에 거리 테이블을 생성하여 각 테이블을 연결해준 것이 가장 주요한 관계 설정이고, 그 외의 기능들을 구현하기 위해 북마크와 리뷰, 유저 테이블도 만들어서 기존 테이블과 함께 관계를 설정하여 DB를 구축하여 주었습니다.

(슬라이드 넘김)

저희는 시연 영상을 준비했는데요, 영상에 앞서 각 페이지들에 대해 먼저 설명을 드린 다음, 영상을 보여드리려고 합니다.

첫번째로는 저희 웹 서비스를 들어오면 디폴트로 보이는 메인 페이지입니다. 사용자가 저희 서비스에 들어오면 본인의 위치를 중점으로 근방의 지도 화면을 보여줍니다. 현재 제 위치가 찍힌 곳은 저희 대면 수업이 이루어지는 강의장의 위치입니다. 이 근처에 물놀이장, 바닥분수, 야외 수영장에 대해 편의등급의 색깔이 들어간 마커가 표시됩니다. 또한 각 물놀이장 종류는 지도 좌측 상단에 위치한 셀렉트 박스를 클릭하여 마커를 온/오프 할 수 있습니다. 현재는 물놀이장을 클릭하여 화면에 출력하고 있는 상태입니다.

(슬라이드 넘김)

그리고 좌측 상단의 햄버거 모양의 토글 버튼을 클릭하면 사이드 바가 열리면서 편의시설 9종류에 대한 아이콘이 표기가 됩니다. 9종류는 수유실, 어린이 이용가능 공중화장실, 편의점, 다이소 및 드럭스토어, 병원, 약국, 경찰서, 소방서 및 119안전센터, 그리고 공영주차장으로 이루어져 있으며, 해당 편의시설 아이콘을 클릭하면 지도 상의 해당 편의시설의 마커들이 표시되고, 다시한번 누르면 마커를 on/off 할 수 있습니다.

(슬라이드 넘김)

그리고 또 수유실을 클릭하면, 수유실의 마커도 표시가 되는 것을 볼 수 있습니다.

(슬라이드 넘김)

그리고 물놀이장 마커를 클릭하면 장소의 기본 정보들이 포함되어있는 작은 팝업창이 열리게 됩니다. 해당 장소의 이름과 주소, 운영시간 등의 기본 정보와 사진이 나타나고요. 그리고 상세 페이지로 이동할 수 있는 버튼과 네이버 지도의 길찾기 사이트로 연결되는 길찾기 버튼이 있습니다.

(슬라이드 넘김)

그리고 아까의 팝업창에서 상세페이지 버튼을 클릭하면 나오게 되는 디테일 페이지인데요, 먼저 해당 물놀이장의 사진과 상세 정보들이 제공이 되고, 편의지수와 리뷰 긍/부정 비율을 시각화한 자료가 보여집니다.

(슬라이드 넘김)

그리고 스크롤바를 좀 더 내리면, 시설별 상세 편의지수와 해당 행정구역 내 다른 물놀이장들의 평균 점수와 비교하는 그래프, 그리고 리뷰 키워드들로 구성된 워드클라우드 자료가 보여집니다.

(슬라이드 넘김)

그리고 상단 네비게이션 바의 통계 한눈에 보기 메뉴를 클릭하면 데이터 분석을 통한 물놀이장 통계 자료들을 유저에게 좀 더 제공할 수 있는 페이지를 만들어주었습니다.

(슬라이드 넘김)

다음으로는 시연 영상을 보여드리려고 합니다.

(슬라이드 넘김)

새로고침

물놀이장, 바닥분수, 수영장

햄버거바 열기

덕의근린공원 물놀이장으로 이동 후 확대

화장실, 약국, 주차장

햄버거바 닫기

마커 클릭 후 팝업창 보여주기

상세페이지 클릭

스크롤 내리기

바닥분수 클릭

삿갓봉공원 클릭

상세페이지 클릭

수영장 클릭

뚝섬 수영장 클릭

상세페이지 클릭

통계 한눈에 보기 클릭

새로고침

(슬라이드 넘김)

그리고 저희 서비스의 확장 가능성 부분에 대해서 설명드리고자 합니다. 먼저 첫번째로는 personalized 된 서비스 제공 가능 측면입니다. 유저에게 편의시설 중요도를 선택하게 하여 만족거리와 K값을 사용자 맞춤으로 산정하여 개인화 추천 서비스의 성능을 높이고 서비스 이용자들에게 더 큰 편리함을 제공할 수 있겠습니다. 두번째로는 추가로 설문조사를 실시하여 더 큰 빅데이터를 확보하여 인덱스나 리뷰 관련 모델의 성능을 높일 수 있습니다. 세번째로는 지금 물놀이장, 바닥분수, 야외수영장에 한정된 서비스 범위를 넓혀서 계곡과 해수욕장까지 확장 가능합니다. 네번째로는 현재 서울특별시에 한해 서비스 되고 있지만, 데이터 수집 범위를 넓혀 전국으로 확장 가능합니다. 그리고 유저 또한 특정 장소에 평점과 리뷰를 남길 수 있는 기능을 넣어 댓글 기능으로 커뮤니티 적인 기능도 구현이 가능합니다. 마지막으로 사용자의 저희 웹 서비스 이용 데이터가 쌓인다면, 유저의 평점 및 리뷰 데이터와 북마크 한 장소 데이터를 기반으로 코사인 유사도를 통한 여름철 물놀이 장소 추천 서비스도 추가가 가능합니다.

(슬라이드 넘김)

다음은 저희 웹 서비스를 만드는데 참고한 자료와 데이터의 출처입니다.

(슬라이드 넘김)

이것으로 2조의 여름철 물놀이장 장소 정보 제공 서비스에 관한 발표를 마치겠습니다. 감사합니다.