4차 4기 실무 Project

|  |
| --- |
| **OCR을 활용한**  **카페메뉴 추천 웹** |

**2019년 05월 23일**

**지능형 IoT 봇을 활용한 빅데이터 개발 전문가**

**오자매 조**

김란아

문수민

신동수

이지원

허무지

**목 차**

**1. 프로젝트 개요** **1**

1.1 프로젝트 기획 배경 및 목표 1

1.2 구성원 및 역할 1

1.3 프로젝트 추진 일정 2

**2. 프로젝트 현황** **3**

2.1 시장 분석 3

2.2 경쟁 제품 장단점 분석 3

2.3 차별화 핵심 전략 기술 3-4

**3. 프로젝트 개발 결과** **5**

3.1 데이터 수집 및 처리 5-6

3.2 주요 기능 및 상세 동작 6-8

3.3 주요 다이어그램 및 분석서 9

**4. 기대 효과** **9**

4.1 향후 개선 사항 10

4.2 기대 효과 10

**5. 개발 후기** 11-13

**6. 강사 의견** 14

**1. 프로젝트 개요**

1.1 프로젝트 기획 배경 및 목표

- 카페 메뉴판의 수많은 메뉴 중에 어떤 것을 먹어야 할지 선택을 못했을 때 해당 카페에서 어떤 메뉴가 가장 맛있는 메뉴인지 모를 때 먼저 메뉴들을 먹어본 사람들의 평가 데이터를 종합하여

선택해주는 메뉴 추천시스템을 개발하고자 했다.

- 지능형 IOT봇을 활용한 빅데이터 개발 전문가 과정을 통해 배운 DataBase, 인공지능 API 활용 기술을 접목한 서비스를 개발하고자 한다. 메뉴판의 메뉴를 인식하고, 기존의 메뉴 평가를 점수화하여 추천 메뉴를 소개하는 서비스이다.

- API 기술 중 광학 문자 처리 (OCR: Optical Character Recognition) 과 자연어 처리(NLU:Natural Language Understanding)를 사용한다. 메뉴판을 인식할 때 광학 문자 처리를 통해 사진에서 글자를 추출한다. 또한, 리뷰의 감정 분석을 통해 단순한 별점을 넘어 사람들의 메뉴에 대한 인식을 분석한다.

1.2 구성원 및 역할

|  |  |  |  |
| --- | --- | --- | --- |
| 이름 | 전공 | 역할 | 구현 부분 |
| 김란아 | 모바일시스템공학 | 팀원 | Selenium을 이용한 크롤링 작업, Django로 웹 구현  , 웹 디자인, 보고서 작성 및 PPT |
| 문수민 | 산업정보시스템공학 | 팀원 | OCR google api를 활용한 데이터 추출, 데이터 전처리, Django로 웹 구현, 보고서 작성 |
| 신동수 | 정보통신공학 | 팀원 | Selenium을 이용한 크롤링 작업, OCR google api를 활용한 데이터 추출, 데이터 전처리, Django로 웹 구현 |
| 이지원 | 경영학 | 팀원 | Selenium을 이용한 크롤링 작업, Django로 웹 구현, 웹 디자인, 보고서 작성 |
| 허무지 | 정치외교학 | 팀원 | OCR google api를 활용한 데이터 추출, 데이터 전처리, 보고서 작성 및 PPT  ,발표 |

1.3 프로젝트 추진 일정

|  |  |  |  |
| --- | --- | --- | --- |
| 구분 | 기간 | 활동 | 비고 |
| 사전  기획 | 4/26(금)  ~  4/29(월) | 프로젝트 기획 및 팀 구성 |  |
| 4/26(금) | PJT주제 선정, 팀(PM/팀원) 구성 |  |
| 4/29(월) | 프로젝트 멘토링  [프로젝트 방향 설정 및 현업프로젝트 소개] | 현업 멘토 참여 |
| PJT  수행  /  완료 | 4/30(화)  ~  5/23(목) | 프로젝트 수행 |  |
| 4/30(화)  ~  5/2(목) | 프로젝트 설계 |  |
| 5/2(목) | 프로젝트 멘토링  [프로젝트 점검 및 기술 자문 ] | 현업 멘토 참여 |

**2. 프로젝트 현황**

2.1 시장분석

-기존 식음료 추천 서비스는 각 매장별 리뷰를 제공한다. '망고플레이트', '트립어드바이저' 등의 사이트를 통해 지역 별, 카테고리 별 추천 내용을 볼 수 있다.

-어플리케이션에서는 무작위 선정(랜덤)방식으로 메뉴를 제안한다. '오늘 뭐먹지?'의 경우에도 화면에서 버튼을 클릭하면 일시정지한 메뉴를 추천하는 방식이다.

- 사용자에게 업체 내 메뉴를 추천하는 서비스는 스타벅스에서 제공하고 있다. 스타벅스 어플리케이션에서는 해당 시간, 날씨, 이전 구매 내역을 기반으로 메뉴를 추천하고 있다.

2.2 경쟁 제품 장단점 분석

- 기존 식음료 추천 서비스의 경우 많은 사용자들이 이용하고 있어 리뷰와 평점의 수가 많다. '망고플레이트'의 경우 평점은 '맛있다', '괜찮다', '별로' 로 나뉜다. 평점별로 분류하여 평점의 개수를 보거나, 각 리뷰에 담긴 반응을 읽을 수 있다. 그러나 평점은 매장별로 제공되는 정보이기 때문에 메뉴 각각에 대한 정보를 알 수 없다. 또한 각 리뷰는 한두 가지 메뉴에 대한 정보를 알 수 있어서 매장 내 메뉴 별 비교가 어렵다.

- 무작위 추천 방식은 단순한 구조로 서비스를 제공하여 선택하기 편하다. 여러 선택지 중 하나를 제안해주지만, 각 메뉴를 제안해주는 근거가 충분하지 않다. 여러 메뉴 중 어느 것을 추천하더라도 무작위라는 점에서 이유가 같다.

- 스타벅스 어플리케이션은 매장 내에서 여러 메뉴 중의 하나를, 객관적인 데이터에 따라 제공한다는 점에서 의미가 있다. 다만, 메뉴에 대한 반응(리뷰)을 반영할 수 없다는 문제가 있다.

2.3 차별화 핵심 전략 기술

-OCR을 이용한 매장 확장 가능성

스타벅스는 스타벅스의 데이터베이스 내에 있는 메뉴를 기반으로 메뉴를 불러온다. 다양한 카페에서 매장에 방문 했을 때의 상황에 맞춰 메뉴를 추천하기 위해서 메뉴판에서 메뉴를 인식해올 필요가 있다. Google 의 Vision API를 활용하여 메뉴 사진에서 한글 텍스트를 읽어오면 사용자가 메뉴판 사진을 찍는 단순한 과정으로도 메뉴 선택지를 만들 수 있다는 장점이 있다. 현 프로젝트에서는 규모를 유지하기 위해 파스쿠찌를 대상으로 서비스를 제공한다. 프로젝트를 확장한다면 다양한 카페의 메뉴 사진을 활용해 유사한 메뉴에 대한 사용자의 추천 의견을 받을 수 있을 것으로 예상된다.

-NLU를 이용한 비정형 데이터 분석

기존의 추천 데이터는 평점과 리뷰의 연결고리가 적다. 서비스에 활용하기 위해 ‘요기요’에서 크롤링한 리뷰 데이터를 보면 평점이 5점(최고점)인데도 실제 리뷰 내용은 불만을 제기하는 등 Null 값을 제거하기 위해 임의로 평점 값을 채우는 경우가 있다. 또한 리뷰 데이터가 많음에도 불구하고, 사용자는 상위 페이지에 노출된 몇몇 리뷰를 읽고 판단을 내려야 하거나, 메뉴 별로 리뷰에 대한 평가에 접근하기에는 복잡하다. IBM의 NLU API를 사용해 리뷰 자체에 대한 긍/부정 지수를 제공하여 사람들의 구체적인 평가를 점수로 환산하고자 한다. 메뉴별로, 종합적인 평가를 얻을 수 있다는 점에서 리뷰 데이터 전체에 대한 분석을 할 수 있다.

**3. 프로젝트 개발 결과**

3.1 데이터 수집 및 처리

3.1.1 데이터 수집

* 리뷰데이터 - Selenium 웹 크롤링
* ‘yogiyo’ 사이트를 크롤링하여 메뉴 별 리뷰를 수집했다. 배달 사이트의 경우 주문을 한 사람에 한해 리뷰를 남길 수 있다. 주문을 한 메뉴를 표시할 수 있어 메뉴 별 반응을 얻기가 용이했다.
* 사용자가 직접 동작하지 않으면 정보가 공개되어 있지 않도록 웹 페이지가 설계되어 있는 경우가 있다. ‘더보기’버튼을 눌러야만 모든 리뷰를 볼 수 있어 브라우저를 직접 조작할 필요가 있다. Selenium은 웹 브라우저를 컨트롤하여 크롤링하는 방식이다. WebDriver를 통해 Selenium이 자동화하여 웹사이트를 탐험한다.
* 파스쿠찌 홈페이지에서 각 지점별 주소를 가져왔다. 주소를 입력하면 배달이 가능한 매장을 확인할 수 있다.
* 매장 명, 유저ID, 메뉴, 리뷰, 맛-양-배달 평점, 날짜를 가져왔다. CSV 파일로 저장한 뒤 Insert문을 사용해서 OracleDB에 연결했다.

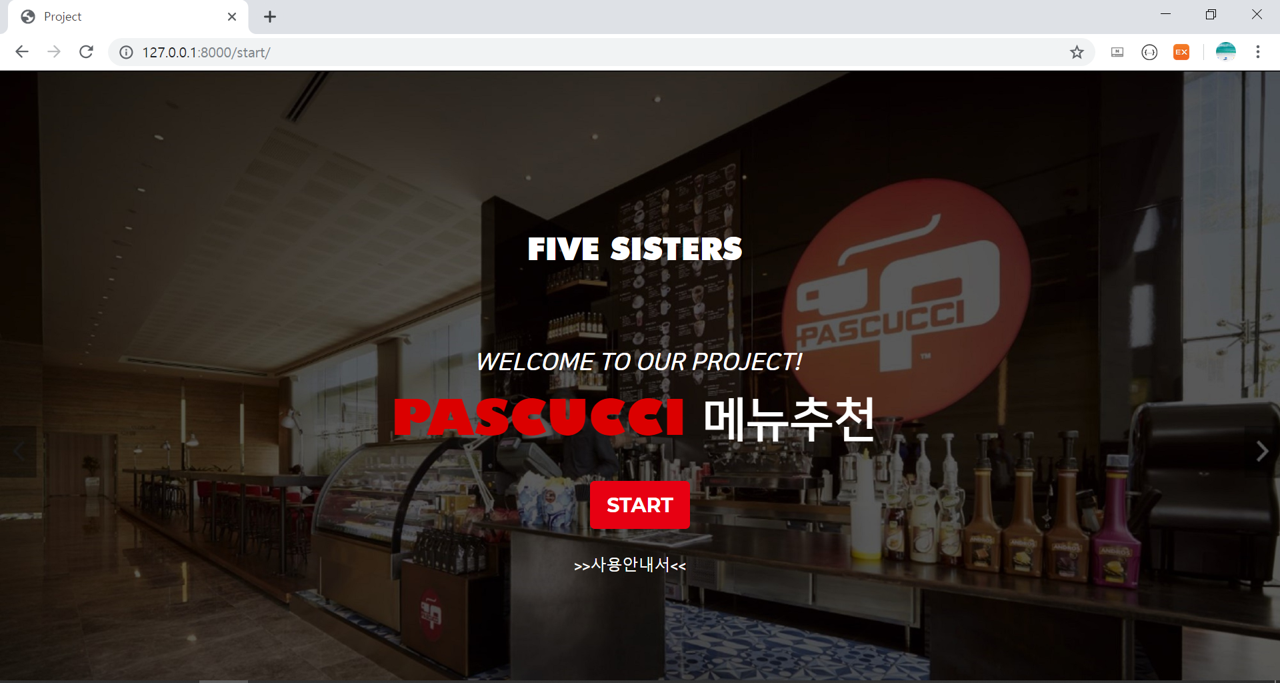
3.1.2 데이터 처리

* 데이터 전처리
* 사용자의 개별 리뷰를 메뉴 단위로 나누는 전처리를 했다. 각 메뉴별로 리뷰를 재구성하여 메뉴 이름별로 데이터를 집계 할 수 있었다.
* 리뷰가 없고 평점만 남아있는 경우 열을 모두 제거 하여 결측값을 처리했다.
* 주문 날짜의 날씨 데이터를 가져오기 위해 리뷰 데이터의 날짜 형식과 기상청 API의 날짜 형식을 맞췄다.
* 기상청 자료는 강수 지속 시간, 1시간 최대 강수량, 일 강수량 값을 제공하고 있다. 기상청에서는 1시간 최대 강수량이 5mm를 넘는 경우 비가 온다고 발표를 한다. 이에 맞춰 강수 지속시간이 하루 3 시간 이상 이거나, 1시간 최대 강수량이 5mm를 넘는 경우, 일 강수량이 5mm를 넘는 경우를 비오는 값으로 처리했다.
* 날짜 칼럼을 기준으로 강수 여부와 리뷰 데이터를 합친 자료를 만들었다.
* 데이터 정규화
* 맛 - 양 평점은 1~5 점으로, 리뷰 데이터를 수치화 한 값은 -1 ~ 1의 값으로 나타난다. 메뉴 주문 빈도수도 리뷰 데이터가 1개 ~200 여개로 다양하다. 데이터 분석을 위해 각 요소의 값을 0~1의 값으로 정규화 시켰다. scikit-learn의 Minmaxscaler 을 활용했다.

3.1.3 데이터 분석

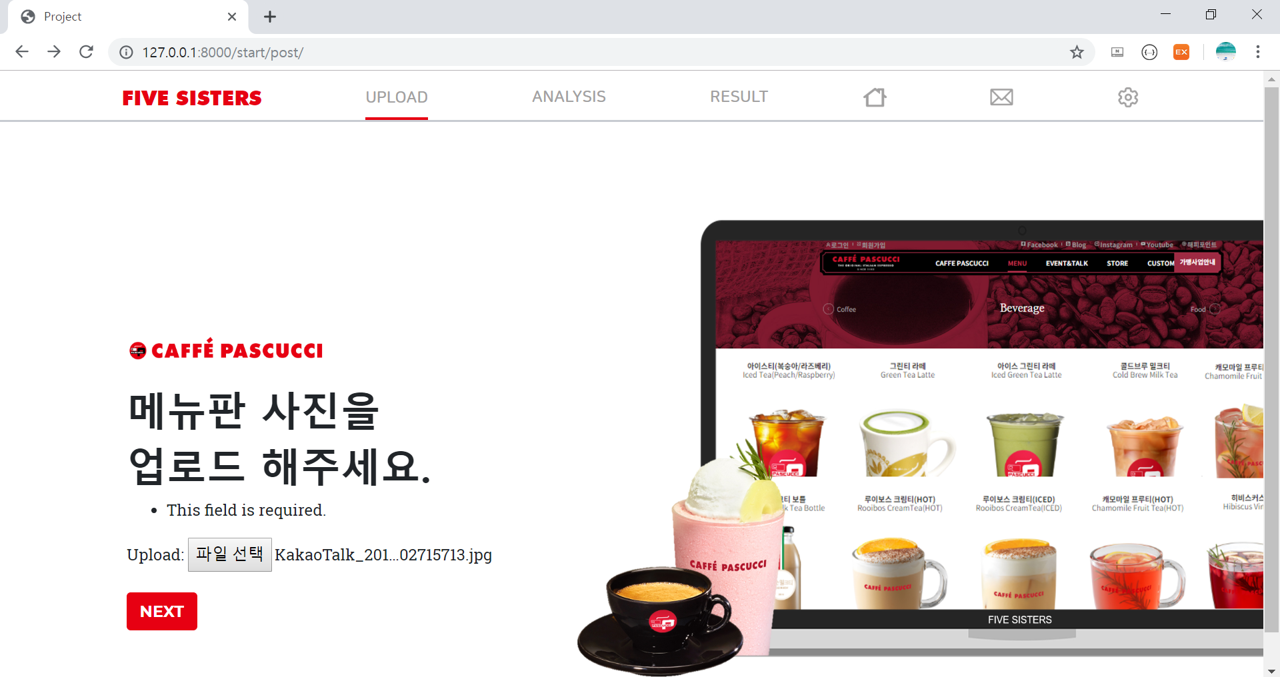
* 빈도수, 리뷰 감정분석, 맛 평점, 양 평점, 비 여부를 기준으로 각 수치를 0~1 로 환산했다. 각 기준에 가중치를 부여한 뒤 메뉴 별 추천 지수를 만들었다.
* 가중치는 AHP(Analytic Hierarchy Process)를 이용했다. 계층화 분석법은 요소를 쌍대비교 한 뒤 행렬로 정리하여 값을 산출한다. 쌍대비교는 a가 b보다 n배 중요하다는 논리이다. 쌍대비교 행렬 곱을 한 다음 각 행렬의 각 열을 합하고, 전체 합으로 나누어 상대적 중요도를 도출한다. 각 가중치의 합은 1이 된다.
* 가중치를 계산한 결과를 반올림하여 빈도수 0.5, 리뷰 0.25, 맛 0.15, 비 0.07, 양 0.03을 곱했다. 당일 날씨와 맞추기 위해 비가 올 경우와 비가 오지 않을 경우로 나누어서 따로 결과값을 만들었다. 비가 오지 않을 경우에는 비가 온 때의 평균에서 1을 뺐다.

3.2 주요 기능 및 상세 동작



1. 시작 화면

* index.html
* views.py의 index 함수 호출.
* 분석을 원할 경우 Start 버튼을 누르면 다음 화면으로 이동한다.

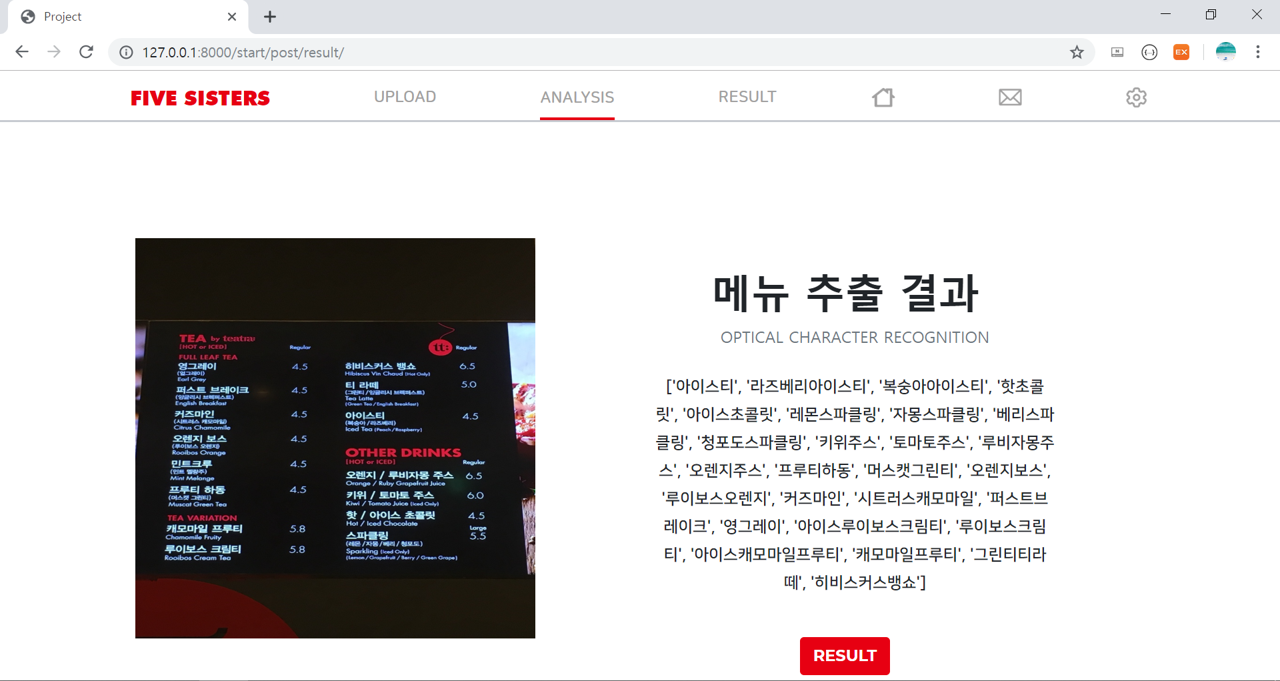


2. 메뉴판 업로드

- post..html

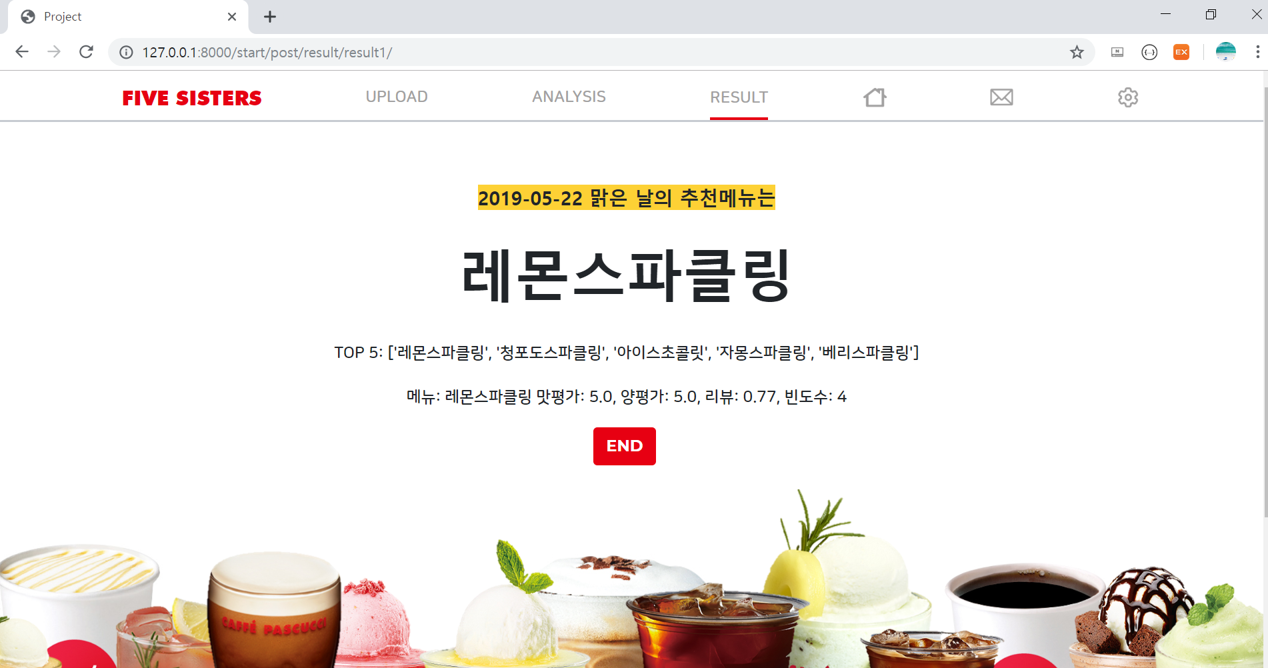
- views.py 의 create postview 함수 이용

- 메뉴판 사진 업로드를 원하면 파일 선택 버튼을 누른다. 이미지 파일이 드라이브에 저장된다. 메뉴판 인식을 원할 경우 Next 버튼을 눌러 다음 화면으로 이동한다.



3. 인식한 메뉴 이름 표시

* index02. html
* Views.py의index01 함수 호출
* OCR을 통해 인식한 글자를 메뉴 DB와 대조하여 추출한 메뉴의 이름을 나열한다. 메뉴 추천을 원할 경우 Next 버튼을 눌러 다음 화면으로 이동한다.



4. 메뉴 추천

* index3. html
* views.index02
* 감성 분석한 리뷰 데이터, 날씨, 맛 평가를 환산한 메뉴를 추천한다.
* 추천 근거 데이터를 표시한다.
* Home 버튼을 눌러 첫 페이지로 돌아간다.

3.3 주요 다이어그램 및 분석서



3.3.1 이미지 업로드

사용자가 메뉴판 이미지를 어플리케이션에 업로드 한다. 어플리케이션에서는 업로드한 이미지 사진을 확인하기 위해 웹 화면에 메뉴판 사진을 보여준다.

3.3.2 메뉴판 OCR

어플리케이션에서는 이미지 경로를 입력하여 구글 Vision API로 사진을 보낸다. OCR 을 통해 이미지에서 텍스트를 추출한다. JSON 형태로 전송받은 글자가 모두 있는 메뉴 이름을 추출해서 메뉴 목록을 완성한다. 기상청 날씨 API를 통해 현재 날씨를 받는다. 비가 오는지 여부를 확인하여 비가 올 경우 1, 비가 오지 않을 경우 0을 반환한다.

3.3.3 추천 메뉴 선정

어플리케이션은 CX\_Oracle 모듈로 접근한다. Oracle DB에서 메뉴별 데이터 전처리 결과를 받아온다. 어플리케이션에서는 메뉴 별 추천 지수를 정렬하여 사용자에게 보낸다. 사용자는 가장 추천하는 메뉴, 추천 지수가 높은 상위 5개 메뉴, 가장 추천하는 메뉴의 추천 근거를 받게 된다.

**4. 기대 효과**

4.1 향후 개선 사항

* 리뷰에서 특정 메뉴에 대한 언급을 구분하면 보다 정확한 값을 도출할 수 있다. 그러나 전체 리뷰의 길이가 짧고 실제 메뉴 이름을 리뷰 내에서 구분하여 언급한 경우가 많지는 않아 의미 있는 값을 얻기 힘들었다. 요기요 외에 다른 사이트에서도 데이터를 수집할 수 있다면 특정 단어에 대한 감성을 도출하는 기능을 활용할 수 있다.
* 다른 브랜드로 확장하기 위해서는 메뉴 이름을 인식하는 방법을 개선할 필요가 있다. 에이드, 티 라떼 등 메뉴의 종류를 먼저 언급하고, 하위 목록으로 재료의 종류만 나열한 경우 하나의 메뉴 이름으로 만들어지지 않았다. 정확한 단어 명을 먼저 제시해야 한다면 개인 카페 등의 메뉴는 조사 후 데이터베이스를 마련하는 과정이 필요하다.
* 날씨의 경우 비의 여부에 따라 0과 1로 나누었다. 눈이 오는 날을 어떻게 처리할 것인지에 대한 논의가 있었다.

4.2 기대 효과

* 사용자 기반, 컨텐츠 기반 추천을 하기 위해서는 컨텐츠에 대한 사용자 정보가 필요하다. 그러나 서비스를 처음 시작하거나, 새로운 사용자가 등장할 경우 사용자에 대한 정보를 입력 받는 과정이 필요하다. 이 기간 동안 사용자에게 추천할 콘텐츠를 제공하기 위한 자료로 활용할 수 있다.
* 기존의 추천 방식은 무작위이거나, 충분한 내용의 평가 정보를 제공하기 어렵다는 한계가 있다. 리뷰 데이터의 경우 상위에 노출된 몇 개의 자료만 읽고 판단하여 전체 리뷰를 활용하지 못한다는 문제가 있다. 리뷰 데이터 각각의 감성을 인식하여 수치화 함으로써 사용자에게 편리하게 추천 정보를 제공할 수 있을 것으로 기대 된다.

**5. 개발 후기**

|  |  |  |
| --- | --- | --- |
| 신동수 | 김란아 | 허무지 |
| 사람, 정장, 넥타이, 의류이(가) 표시된 사진  자동 생성된 설명 | 사람, 실내, 여자, 앉아있는이(가) 표시된 사진  자동 생성된 설명 | 사람, 실내, 벽, 테이블이(가) 표시된 사진  자동 생성된 설명 |
| 이지원 | 문수민 |  |
| 실내, 사람, 테이블, 벽이(가) 표시된 사진  자동 생성된 설명 | 사람, 하늘, 의류, 실외이(가) 표시된 사진  자동 생성된 설명 |  |

|  |  |
| --- | --- |
| 성명 | 후기 |
| 신동수 | 원하는 프로젝트 결과물을 생각만 하는것과 현실로 구현하는 것의 차이를 크게 느꼈습니다 여러가지 차이들을 조율하고 문제점들을 해결하는 과정에서 많은 것을 배웠고 기초를 다지게 되었습니다  프로젝트를 완성한 경험이 생기니 IT 프로젝트 기획, 코딩, 개발에 대해 큰 자신감이 생겼습니다. |
| 김란아 | 이번 프로젝트를 구성하면서 교육 과정 중 배웠던 내용을 어떻게 적용시킬 수 있을지에 대해 많이 고민했다.  나 혼자서는 디테일하게 적용시키지 못했을 텐데 팀원들의 아이디어들 덕분에 상세하게 구성이 가능했다.  우리가 배웠던 것들이 이렇게 실제로 활용될 수 있는지 직접 만들어보고 느껴보니 짜릿한 기분이 들었다. 수업만 들을 때와는 너무나도 다른 기분이었다.  그리고 팀원들이 각자의 역할을 수행하면서 팀원 하나하나의 어떠한 역량이 우수한지, 그리고 내가 부족한 것들은 무엇인지 보이게 되었다.  저마다의 훌륭한 역량으로 완성된 팀워크로 서로의 부족한 부분을 채워주면서 프로젝트를 완성시켜 나가면서 팀워크의 덕목을 배우게 되었고, 팀워크의 필요성 또한 절실히 느꼈다.  나 혼자만이 하는 일이 아닌 모두의 힘으로 합쳐서 무언가를 완성한다는 것이 나에게 일체감을 주었던 것 같다.  그래서 조금 더 완벽하게 완성시키고 싶어 졌고, 평소와는 다르게 더 열심히 찾아보고 공부하고 매달리는 내 모습을 발견했을 때의 감정은 이로 말할 수 없다.  후에 다른 프로젝트에 임할 때에도 이 감정 이 느낌을 잊지 않고, 끝까지 나아가는 내가 되길 바란다. |
| 허무지 | 코드를 짜고 API를 적용하는 과정에서 생각했던 것과는 다른 난관이 많이 있었다. 메뉴판에서 당연히 메뉴이름이 나올 줄 알았는데 단어가 낱개로 나왔을 때에도 당황했다. 같이 아이디어를 마련해서 문제를 해결하는 과정이 재밌었다. 친구들이 고생을 많이 했다. 함께 고난을 헤쳐가면서 뜻 깊은 프로젝트가 되었다. |
| 이지원 | 프로젝트를 수행하면서 아이디어를 실제로 구체화하고 구현하는 것이 생각보다 훨씬 어렵다는 것을 배우게 되었습니다. 또한 평소에 쉽게 접할 수 있는 서비스들 예를 들면 웹페이지 하나를 만드는데도 많은 노력과 시간이 들어간다는 것을 알게 되었습니다.  혼자가 아닌 여럿이 함께 모여서 프로젝트를 진행하다 보니 다양한 의견과 배경지식을 나눌 수 있었고 원하는 아이디어를 서비스로 완성할 수 있었습니다. |
| 문수민 | 백지 상태에서 팀원이 모여 여러 아이디어에 대해 의견을 나누고 구체화 시켜 최종적으로 우리가 목표했던 서비스를 구현했다는 점에서 만족스러운 프로젝트라고 생각합니다. 수업을 들으며 배우는 것과 그것을 토대로 실제 구현에 적용하는 일이 어렵다는 것을 느꼈습니다. 수업 때 배웠던 python을 이용한 데이터 처리 부분을 다시 한 번 정리할 수 있는 기회가 되어 좋았고, 잘 모르던 웹 사이트 구현 원리를 조금 더 알게 되었습니다. |

**6. 강사 의견**

|  |  |  |
| --- | --- | --- |
| 평 가 요 소 | 배점 | 평 |
| 아이디어 :  유사한 서비스 존재 유무 및  체계성 | /20 |  |
| 2. 개발 : 실제 구현 정도 및 배포 유무,  코드의 무결성 및 난이도, 현업적용도, 실무기술 반영정도 | /30 |  |
| 3. PJT 수행력 : 일정관리 및 역할분담, 목표 일정 달성도, 팀내 참여도 등 | /30 |  |
| 4. 준비도 : 프리젠테이션 및 프로젝트 준비 정도 | /20 |  |
| 계 | /100 | 강사 의견 필수 |