|  |  |  |  |
| --- | --- | --- | --- |
| [2차 프로젝트 수행 일지 ] | | | |
| 서비스 산업 데이터를 활용한 빅데이터 분석 실무 | | | |
| 프로젝트 타이틀 | 제로칼로리 음료의 상품 리뷰 네트워크 분석 | | |
| 프로젝트 팀명 | **Something to Drink?** | **프로젝트 팀원** | (팀장) 최진형  (팀원) 김기쁨, 오한샘 , 조용일 |

|  |  |
| --- | --- |
| 1. 2주차 프로젝트 수행 계획 및 달성도 | |
| 이름 | **수행 계획** |
| 최진형 | * 프로젝트 주제 선정과 데이터셋 찾기   달성률   * **프로젝트 주제 선정 : 완료** * 제로칼로리 음료의 상품 리뷰 키워드를 의미 연결망을 분석하여 네트워크 시각화 서비스를 제공한다. * **데이터셋 찾기 : 23-01-02 완료** * 내용 : 제로칼로리 음료의 리뷰 크롤링 * 크롤링 사이트 : 다나와 , 네이버스마트스토어, 네이버 브랜드샵 , 쿠팡 * 제품 : 제로콜라, 스프라이트제로, 펩시제로 등 (총 15가지 , 음료별 리뷰 10,000개) |
| 김기쁨 |
| 오한샘 |
| 조용일 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2. 2주차 프로젝트 수행 내용 및 문제점 | | | | |
| 이름 | **최진형** | **김기쁨** | **조용일** | **오한샘** |
| 금주  프로젝트 수행 계획 | * 주제 선정 및 데이터셋 구하기 | | | |
| 금주  프로젝트 수행 내용 | * 주제 : 제로칼로리 음료 리뷰 분석 * 분석 방법 : 키워드를 추출하여 의미연결망 분석 기법을 사용한다. * (sematic network analysis) * 데이터셋 : 다양한 쇼핑몰 사이트에 게시된 제로칼로리 음료의 리뷰를 크롤링(selenium, bs) | | | |
| 문제점 | * 주제 선정에 있어서 문제 * 1. 데이터셋을 구하기 힘든 주제가 많았다. * - 애견미용실, 서울시 문화공간 , 제로웨이스트 등(공통된 문제점) * 2. 이미 구체적으로 서비스되고 있는 주제가 많았다. * 크롤링에 있어서 문제 * 1. 크롤링이 불가능한 사이트가 다수 존재 * 2. 상품 판매 홈페이지마다 크롤링 방식이 다르며, 심지어 상품마다 크롤링이 다른 경우가 있었다 – 크롤링 개수 제한 : 네이버의 경우 리뷰를 2천개 까지만 제공 * 쿠팡의 경우 리뷰를 1500개 까지만 제공 | | | |

|  |  |
| --- | --- |
| 3. 3주차 프로젝트 수행 계획 | |
| 이름 | **수행 계획** |
| 최진형 | * - 한 주 계획 정리, 전처리 필요 내용 확인 , 조원 피드백 |
| 김기쁨 | * - 전처리 준비를 위한 모델링 기법 조사(NetworkX) |
| 오한샘 | * - 잔여 크롤링 처리 및 데이터 파일 정돈화 |

|  |  |
| --- | --- |
| 4. 강사님 피드백 | |
| 금주 프로젝트 수행 내용 | **차주 프로젝트 수행 계획** |
|  |  |
| 피드백 일시 년 월 일  강사명 (서명) | |