**비정형 데이터 활용 분석 프로젝트 기획안**

**2023년 4월 25일**

**과정명: [멀티잇]서비스 산업 데이터분석가 취업캠프(Python)**

|  |  |
| --- | --- |
| 프로젝트 주제 | 올리브영 탑리뷰어 리뷰분석을 통한 탑리뷰어 선정모델 개발 |
| 팀 명 | ONE-LIVE YOUNG |
| 팀 원 | 류영주(팀장), 김민제, 서민진, 이석호 |
| 프로젝트 내용  해결하고자 하는 문제  최종 산출물의 청사진 | 1. **프로젝트의 필요성 및 목적**   **(1) 프로젝트 목적**  - 올리브영 탑리뷰어의 선정기준의 정확도를 높여 리뷰의 신뢰도 및 리뷰에 대한 소비자의 인식을 개선시키고자 한다.  **(2) 실태 조사 및 문제 상황**  - 온라인 상권의 활성화 및 코로나19로 인해 비대면 거래가 활발해지면서 소비자의 선택에 있어 이용후기가 크게 영향을 미치고 있는 것으로 나타났다.  - 다른 이용자들의 반응과 경험이 소비자 선택에 중요한 역할을 함에 따라 이용후기의 영향력이 증가하면서 인한 문제점도 늘어나고 있다.  - 구체적으로 이용 후기를 조작 및 소비자가 올린 이용 후기를 삭제하는 등 소비자 피해가 발생하고 있다.  **소비자 이용후기 확인률 97.2%, 신뢰도 70.2%** [표=한국소비자연맹] 출처 : 소비자경제(<http://www.dailycnc.com>)  온라인쇼핑 이용 후기에 대한 소비자를 조사한 결과 소비자의 대다수(97.2%)가 **구매 전 이용후기를 확인**한다고 답했다. 하지만 이용후기를 신뢰한다는 비율은 70.2%로 이용 후기 확인 비율에 비해 상대적으로 낮게 조사되었다.  **소비자 선택에 이용후기 영향 큼** [표=한국소비자연맹] 출처 : 소비자경제(<http://www.dailycnc.com>)  이용후기 누적수가 구매선택에 영향을 미친다는 것을 확인할 수 있다.  이용 후기가 없을 때 대체로 구매를 안하는 경우가 많다.  **‘실제 구매 후 작성’ 중 ‘대가성’ 이용후기 많다.** [표=한국소비자연맹] 출처 : 소비자경제(<http://www.dailycnc.com>)  소비자는 ‘실제 구매’한 소비자가 작성한 정보성 있는 이용후기를 원하지만 실제로는 구매를 하더라도 일부 상품(ex.체험단 상품)이나 적립금 등 ‘경제적인 대가’를 받고 작성하는 이용후기 비율이 높았다.  **‘신뢰할 수 있는 이용후기’ 소비자 니즈 증가** [표=한국소비자연맹] 출처 : 소비자경제(http://www.dailycnc.com)  이용후기와 관련해 신뢰할 수 있는 이용후기에 대한 소비자의 요구가 증가하고 있다.  **(3) 올리브영 ‘탑리뷰어’ 도입**  - 올리브영은 화장품 구매 결정 단계에서 후기가 큰 영향을 미친다는 점을 고려해 온라인몰 내 상품 리뷰 콘텐츠를 강화하고 있다.  - 고객이 자발적으로 후기를 작성하고 소통하는 문화를 조성하고자 ‘탑리뷰언서’를 도입하여 리뷰어 리워드(보상)제도를 시행하고 있다.  - ‘탑리뷰어’는 올리브영에서 구매한 상품에 대한 양질의 리뷰를 작성한 고객들을 1위부터 1000위까지 매주 선정해 공개하는 서비스이다.  출처 : CJ올리브영  **(4) 올리브영 현황분석**  **현재 올리브영 탑리뷰어 선정방식**  - 3가지 항목만으로 탑리뷰어 선정 (정확도 문제)   * 리뷰 도움이 돼요 OR 컬렉션 좋아요 받기 * 한달이상사용 리뷰 작성 * 리뷰 하나당 포토3장 이상 등록     출처 : CJ올리브영  - ‘탑리뷰언서’가 되기 위한 리뷰경쟁 (신뢰도 문제)  [사진=네이버 캡처]  출처: 녹색경제신문  할인 쿠폰과 포인트 적립의 혜택을 받기 위해 도용한 사진 업로드를 하는 등 무차별적인 리뷰를 감행하는 소비자가 발생하고 있다. 이는 리뷰의 신뢰도 문제를 야기한다.  **(5) 추가할 선정기준**  **리뷰의 길이 편차**  : 리뷰길이가 길면 제품에 대한 정보뿐만 아니라 사용후기를 구체적으로 제시하여 제품을 구매하는 소비자들에게 제품의 전달력 높일 것이라고 가정  : 반복되고 성의 없는(대략 10글자 이하 작성)리뷰일 경우 리뷰가 짧다는 것으로 가정  **리뷰의 작성 주기**  : 리뷰 작성 주기를 통해 리뷰를 몰아서 작성하는지 확인 가능  : 리뷰를 작성 마감일에 몰아서 작성하는지 확인 가능  **직접 구매 상품과 체험단 신청 상품 리뷰의 차이 (비교)**  : 이를 통해 평소 리뷰에 쏟는 노력의 정도를 알 수 있을 것이라고 예상  **리뷰 성의없는데 ‘도움이돼요’ 개수 많은 경우**  : 현재 올리브영 선정기준에 있는 ‘도움이 돼요’ 개수의 수는 점수 반영비중이 가장 높습니다. 하지만 ‘도움이돼요’ 개수는 많지만 성의 없는 리뷰를 종종 발견하였습니다. 특히 음식종류의 경우에서 많이 나타나는 것을 확인하였습니다. 이를 어떻게 처리해야 될 지 고민중에 있습니다. |
| 프로젝트 추진 일정 | * 4/20 : 주제 선정, 역할 분담 * 4/21 ~ 4/25 : 크롤링 및 데이터 수집 * 4/22 ~ 4/26 : 전처리 코드 작성, 수집된 데이터 칼럼명 수집 * 4/26 ~ 4/28 : 모델링 * 4/28 ~ 5/1 : 비교 분석 * 5/1 ~ 5/2 : 디버깅 & 개선 모델 제안, 발표자료 제작 * 5/3 : 최종발표 준비   \*4/25 : 기획안 발표  \*4/28 : 팀원 평가  \*5/4 : 프로젝트 발표 |
| 프로젝트 수행 방향  수행 방법/도구 | 1. **프로젝트 수행 방법**   **(1) 데이터 수집**  올리브영 ‘탑리뷰어’ 1000명의 11만개 리뷰 데이터  **(2) 수집 방법**   requests를 사용하여 JSON형태로 받아옴  **(3) 수집 범위**  4/21 기준 2달간 탑리뷰어 랭킹 및 상품별 리뷰  **3. 진행 방향**  **(1) 데이터 전처리**  : 비어있는 리뷰 제거, 토큰화 등등  **(2) 분석 방법 및 모델링**  : 감성분석 모델을 사용하여 도움이 많이 된 리뷰와 성의 없는 리뷰를 분류하고 특징들을 추출하고자 한다.  **4. 예상 결과**  - 추가/수정된 ‘탑리뷰어’ 선정기준을 통해 기존의 올리브영에서 측정한 ‘탑리뷰어’ 순위와 우리가 만든 선정기준을 통해 재측정된 ‘탑리뷰어’ 순위를 비교 분석  - 이를 통해 소비자들에게 리뷰에 대한 인식 개선 및 신뢰도 문제를 해결해 나아갈 수 있다. |