**[SSAFY] 특화 프로젝트 멘토링 활동일지**

**작성일자: 2022 년 03월 07일**

|  |  |  |  |
| --- | --- | --- | --- |
| **도메인** | **빅데이터(추천)** | | |
| **멘토** | **장동인** | **팀 코드** | **D204** |
| **진행 일시** | 2022.03.07 14:20 ~ 14:50 | | |
| **진행 내용** | [프로젝트 주제 소개]  1순위 주제 선정으로 반려식물 추천, 2순위 주제 선정으로 칵테일 추천을 선정하였음.  [피드백]  흥미가 있는 파트를 정하거나 메인 테마를 정해놓고 데이터를 모아야 한다. 데이터 셋을 먼저 모으고 하는 방식은 잘못된 방식이다.  [Q. 데이터 셋이 충분한 가?]  트랜잭션 데이터 셋은 못해도 수천개는 있어야 함. 아이템 데이터는 300개여도 상관이 없음.  [Q. 크롤링 문제]  모로 가도 상관이 없으며, 데이터를 모으는 것이 우선이다.  크롤링 관련 문의에 대해서 사이트를 정리해서 컨설턴트님이나 사무국에 문의를 한번 넣어보는 것을 추천한다. [feat. 크롤링 문제 프로님 답변]  [Q. 데이터 수집에 대한 어려움]  실제 파는 가게에 가본적은 있는가? 실제로 한번 가보고 사람들이 뭘 자주 찾는지, 그런 것들을 파악 하는게 중요하다. 단순한 과제로 생각하지 말고 독특한 테마를 선정한 것은 아주 좋은데, 인풋 데이터가 뭔지를 알아야 한다.  키우고 싶은 사람과 좋아하는 사람의 연관성을 찾아내야 한다. 반려식물을 좋아하는 사람들의 특성을 먼저 알아야 하는 것이 우선이다.  [Q. 룰 기반 추천?]  룰 기반 추천으로 가려면은 굉장히 많은 룰을 추가해야 한다.  또한 사용자들이 프로젝트를 사용하여 나온 데이터를 재가공하여 하이브리드로 가는 것은 좋다. 하지만 피드백 데이터없이 단순히 재사용하는 것은 잘못된 방식이다.  [Q. 딥러닝 방식 이용하는 것은 괜찮은 지?]  데이터에 따라 다르며, 지금은 딥러닝에 필요한 데이터가 없다. 딥러닝은 정해진 답이 있는 데이터가 필요하다. 답이 안정해진 데이터를 쓰는 딥러닝도 있지만, 지금 하는 과제에서 딥러닝은 어울리지 않는다. | | |
| **특이 사항** |  | | |