# 캡스톤 18 조 김민송 멘토링 활용 보고서

| 프로젝트명 | O24Sec | 팀명 | 멜러리를 찾아서 |
|---|---|---|---|

**팀 멘토링**

* 간단히 팀의 공통사항에 대한 미팅 주제와 개인별로 느낀점을 적었습니다.

◎ [03/30] 팀 미팅 - 윤명근 교수님

- 중간보고회 발표 피드백

⇨ 발표 시 현재 산업기관인 윈스와 자주 소통하고 있는 것도 표현해 주면 좋을 것 같다.

⇨ 피피티 내에 글로 된 설명보다 시각적으로 볼 수 있는 그림 위주면 더 좋을 것 같다.

◎ [04/06] 팀 미팅 - 윤명근 교수님

- 중간보고회 발표 피드백

⇨ 발표 시간이 너무 긴 것 같다 상세한 기술 설명은 제외하고 어떻게 만든 지 큰 틀 위주로 발표하는 게 좋을 것 같다.

⇨ 발표 내용의 단어 수준이 높은 것 같다. 전문적인 단어는 최소화하고 간단하게 설명할 수 있는 방법을 찾으면 좋을 것 같다.

⇨ 1 학년에게도 설명했을 때 이해시키려면 전공적인 단어보다는 누구나 아는 단어로 돌려가며 설명하는게 모두에게 더 이해가 잘 될 것 같다.

◎ [04/08] 팀 미팅 - 윤명근 교수님

- 향후 프로젝트 진행 방향에 대한 토의
- 최신 기술 / 논문 리서치
- 기간별, 출처가 다른 데이터에 대한 강건성 확보

⇨ 유사도 관련한 기술을 몇 개 조언 받을 수 있었다. SSDEEP, Locality Sensitive Hashing 등 우리 프로젝트에 적용 시킬 수 있을지 확인을 해봐야 할 것 같다. 현 상황에서 클러스터링 알고리즘 보다 유사도 쪽에서 수치적으로 확인하는게 더 도움이 될 수 있을 것 같기 때문에 해당 내용은 공부를 해봐야 할 것 같다.

◎ [04/12] 윈스 미팅 - 윤명근 교수님, 최병환 팀장님

- 추가 데이터에 관련한 요청 및 논의
- 클러스터링 향후 방향성에 대한 논의

⇨ 정답 데이터로 쓸 최근에 정합을 시도했던 데이터가 있는지 확인해 봤는데 회사 특성상 확인을 해봐야 할 것 같다고 하셨다.

⇨ 데이터 중 특정 이벤트에 대한 불확실성 때문에 해당 데이터는 제외시키고 프로젝트를 지속하기로 하였다.

| | |
|---|---|
| 개인 멘토링 | [04/12] 윤명근 교수님<br>Q. 현재 진행 중인 클러스터링 실험에서 실험 대상이 되는 내용이 단순히 하이퍼파라미터를 바꿔가면서 클러스터링 결과를 보는 중인데, 하이퍼파라미터의 경우 데이터가 바뀔 경우 하이퍼파라미터도 바꿔주어야 하는데 현재 이 실험 방법이 옳은 것인가요?<br><br>A. 만약 하이퍼파라미터를 크게 수정 안 해줘도 클러스터링 결과가 계속 비슷하고 잘 나온다면 조금이라도 더 나은 결과를 보이기 위해 해 볼만한 실험이긴 한데 현재 우리가 목표하는 부분이랑은 거리가 있어 보인다. 그럼 목표를 클러스터링을 가지고 분류가 아닌 각각 이벤트를 서로 유사도 검사해서 유사도가 높게 나온 것을 뭉치는 방식으로 진행을 해보도록 하자.<br><br>⇨ 클러스터링으로 볼 수 있는 한계를 인식할 수 있었고 이전에 공부하고 있던 유사도 알고리즘에 대해서 다시 확인해봐서 프로젝트에 바로 적용시킬 수 있는 걸 찾아보도록 해야할 것 같다.<br><br>Q. 지금 가지고 있는 페이로드로 단어 벡터를 만들어서 유사도 검사를 진행해야 하는게 현재 과제인데, 이때 각 프로토콜에 대한 규정된 헤더정보 같은 것을 이용해보면 어떨까요??<br><br>A. 프로토콜이 워낙 많다 보니 그 부분은 조금 힘들 수 있을 것 같다. 현재 학기 수업도 따라가면서 프로젝트를 진행하다 보니 전부에 대해 조사하는 것은 과한 시간을 투자하는 것이지 않을까 싶다. 현재 80 번 포트에 대해 한정을 시켰으니 80 번에서 자주 나오는 부분에 대해서 확인할 수 있도록 파싱을 해보자.<br><br>⇨ 80 번 포트의 경우에는 띄어쓰기도 많이 들어 가 있고 특수문자도 많기 때문에 해당 부분을 기준으로 페이로드를 Tokenize 해봐서 결과를 확인해봐야 할 것 같다. |