**Weekly report**

날짜 : **2022-09-16**

연구원 : **최윤석**

* **수행결과**
* **차량 이미지 분류 :** 모델 학습시 발생하는 오류의 해결중
* **관계형 테이블 임베딩** :
  + 캐글에서 텍스트 데이터와 수치형 데이터가 혼합된 테이블 데이터셋(혼합형 데이터셋)을 수집 (현재 총 47개 데이터)

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 1. 혼합형 데이터셋 모음



Figure 2. 혼합형 데이터셋 예시

* + 단순 계급 값 이산화 (1-5사이 값으로 이산화), 이도훈 연구원의 방식으로 이산화 (int(), 단순 계급 값 이산화한 것을 범주형 데이터로 (very high(5) / high(4) / medium(3) / low(2) / very low(1)) 변경하는 3가지 방식의 알고리즘으로 비교

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 3. 단순 계급 값 이산화

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 4. 이도훈 연구원의 방식으로 이산화

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 5. 이산화한 것을 범주형 데이터로 변경

* + Fig 5를 보면, Fig 3,4와 다르게 Cars – Seattle – Food\_Production 테이블이 가깝게 위치함.
  + 하지만 실제 테이블 데이터를 보면 전혀 유사하지 않은 3개의 테이블임.
  + 테이블의 대부분을 수치형 컬럼을 가진 테이블들이 이산화되어 레코드들이 이산화된 5개의 값으로 한정됨.
  + 최종 임베딩 벡터는 전체의 0.6을 레코드 임베딩 벡터에 의해 영향을 받기에, 크게 관련이 없는 테이블들이 벡터공간에 비슷하게 위치하게 되는 것으로 생각됨.

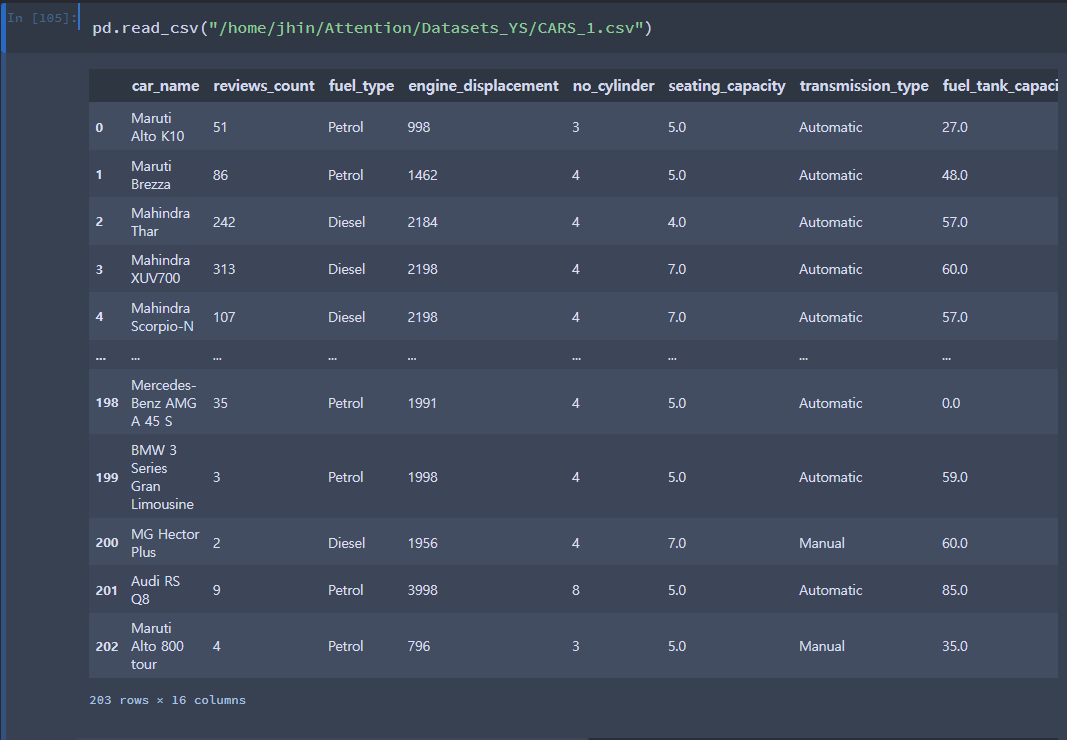


Figure 6. CARS\_1.csv 신차에 관한 평가 데이터

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트, 전자기기, 스크린샷, 컴퓨터이(가) 표시된 사진

자동 생성된 설명

Figure 7. seattle\_01.csv 숙박 공유 사이트 Airbnb의 Seattle에서 일어난 숙박 공유 데이터

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 8. Food\_Production.csv 특정 음식 생산이 환경에 미치는 영향을 정리한 데이터

텍스트이(가) 표시된 사진

자동 생성된 설명

* **수행계획**
* **차량 이미지 분류** : 학습 도중 발생하는 오류 해결 및 모델 학습
* **관계형 테이블 임베딩** :
  + 지속적으로 혼합형 데이터셋을 추가 (목표 전체 데이터셋 약 150개)
  + NDCG 계산에 필요한 Dataset\_rel.csv 파일 생성 및 NDCG로 임베딩 모델의 정량적 평가 진행
  + 이산화한 것을 범주형 데이터로 바꾸는 임베딩을 보완하는 방법 찾고 실험
* **기타사항**
  + **데이터 융합 프로세스**

