



주제분석 계획

- 주제분석
 - 화 8시 회의 각자 할일 정하기+ 주말까지

시각화 - 카테고리 별로 4명이 나눠서

NA처리 / 클래스불균형 - 예진 효주

독립성검정 /클러스터링 - 장희 혜현

★ 2주차 내용

- 데이터 전처리 (NA처리, 범주형변수 인코딩, 파생변수 가능?)
 - 결측치 처리 (설문조사데이터를 다룰 때 주의할 점)
- 시각화 - 어떤 게 가능할까? 변수별 백신점종여부 그래프로 확인
- 범주형 변수끼리의 독립성 검정 + 상관관계 파악
- 클러스터링 통해서 유형을 나눠 볼 수 있을까? + 아니면 다른 방식 뭐가 있을까?
- 클래스 불균형 (계절독감여부) 샘플링 ⇒ SMOTE MSMOTE 등

- 2주차는 전처리 EDA가 메인! ⇒ 어떤 변수들이 백신 접종에 큰 영향을 주는지 알아보기(핵심)
 - 도메인 지식 공부 - 사람들이 백신 접종을 맞는 데에 영향을 주는 요인들이 뭘까?
 - 피피티 템플릿 만들기
 - 다중라벨 분류 공부

♥ 4주차 내용

- 모델링!
 1. Binary 분류 모델 (패널티 로지스틱 회귀, 나이브베이즈안분류, LGBM, 캣부스트 등 등)
 2. 다중라벨 분류방법 적용 (adapted algorithm, ensemble approach 등)
 3. 사후 모델링 해석 방법 적용 -

h1n1_concern

h1n1_knowledge

<행동> - 효주

behavioral_antiviral_meds

behavioral_avoidance

behavioral_face_mask

behavioral_wash_hands

behavioral_large_gatherings

behavioral_outside_home

behavioral_touch_face

<주변 의료적 환경> 장희

doctor_recc_h1n1

doctor_recc_seasonal

chronic_med_condition

child_under_6_months

health_worker

health_insurance

<의견> - 혜현

opinion_h1n1_vacc_effective

opinion_h1n1_risk

opinion_h1n1_sick_from_vacc

opinion_seas_vacc_effective

opinion_seas_risk

opinion_seas_sick_from_vacc

<인적 경제적 정보> -지연 예진

age_group

education race

sex

income_poverty

marital_status

rent_or_own

employment_status

hhs_geo_region

census_msa

household_adults

household_children

employment_industry

employment_occupation