

당뇨병 진단 서비스

설문조사 데이터를 기반으로 당뇨병 진단 예측 웹 어플리케이션



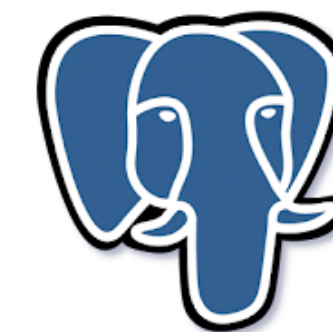
질병관리청
국립보건연구원

한국인 유전체 역학 조사 사업 교육용 데이터
당뇨병, 고혈압, 고지혈증



EDA 진행

로컬 데이터 베이스로 저장



PostgreSQL



Flask

모델을 플라스크에 불러와서
웹 프레임워크 생성



모델을 PKL 파일로 저장



LightGBM

머신러닝 모델 생성



HEROKU

배포

데이터 수집



질병관리청
국립보건연구원

KoGES

질병 연구 기반 구축을 위한
대규모 한국인 코호트 운영

질병관리본부 국립보건연구원 유전체역학과에서는
국내 최대 규모의 코호트 연구인 한국인유전체역학조사사업을 수행하고 있습니다.



교육용 데이터 다운로드

구분	데이터 다운로드	코드북 다운로드
기반조사 교육용 데이터	 base_data1	
	 base_data2	
	 base_data3	

만 40세 이상의 1만명 조사. 병력, 가족력, 생활습관을 포함한 임상검사 자료

데이터 처리

1. EDA

3개의 SAS 파일(SAS7BDAT)을 CSV 파일로 변경 (웹서비스 이용 <https://dumbmatter.com/sas7bdat/>)
PANDAS 라이브러리를 이용하여 하나로 합친후 EDA를 진행

최종적으로 결측값을 처리한 9000여행의 데이터 CSV파일로 추출

2. PostgreSQL

Python을 통해 PostgreSQL에 CSV 파일을 업로드
로컬데이터베이스 생성

머신 러닝

PostgreSQL와 연결하여 데이터 입력

Light gbm을 사용하여 모델링

- target : 당뇨병진단여부
- GridSearchCV 이용 최종 모델의 ROC AUC ~ 0.66

Joblib 를 이용하여 모델을 pkl 형식으로 추출

Flask 웹 개발 및 Heroku 배포

21개의 항목을 입력하면 당뇨병 진단 가능성을 예상하는 웹 어플리케이션을
Flask로 개발

최종적으로 Heroku를 통해 Flask 앱 배포

<https://pj3hamal.herokuapp.com/>

한계 및 보완 계획

데이터 양에 비해 적은 당뇨병진단자 수로 인해 시각적으로 큰 차이를 느낄 수 없었습니다.

배경지식 부족으로 혈액임상데이터 등을 활용하지 못하였습니다.

고혈압, 고지혈증도 포함하는 데이터인데 다 활용하지 못하였습니다.

고혈압, 고지혈증을 포함한 진단 애플리케이션

Flask 앱에 CSS를 적용하여 애플리케이션의 시각적 업데이트