



Health Care Industry

Stroke Prediction



디지털 헬스케어



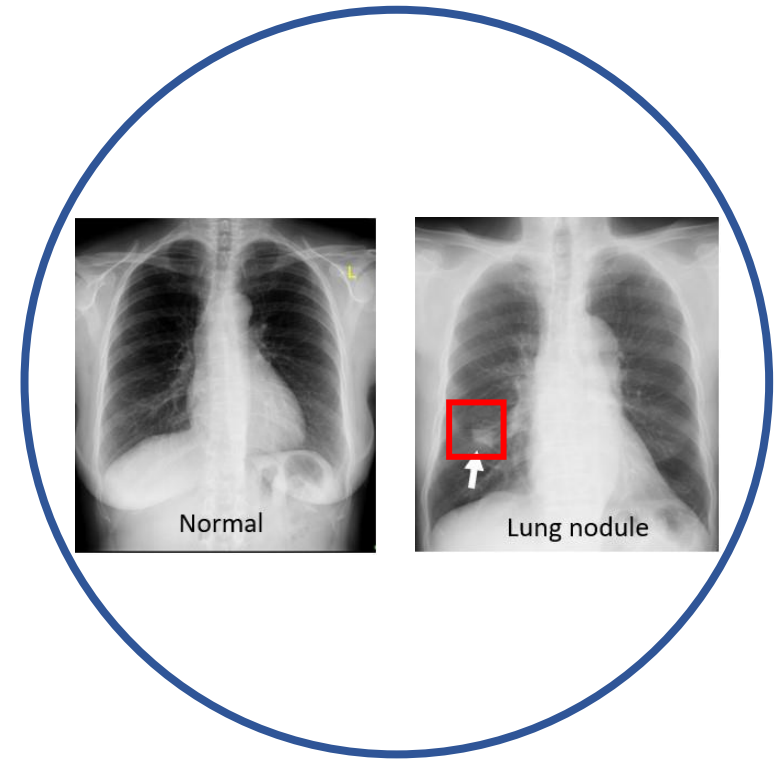
사례



Wearable Device



Telehealth



AI prediction

산업 발전 계기

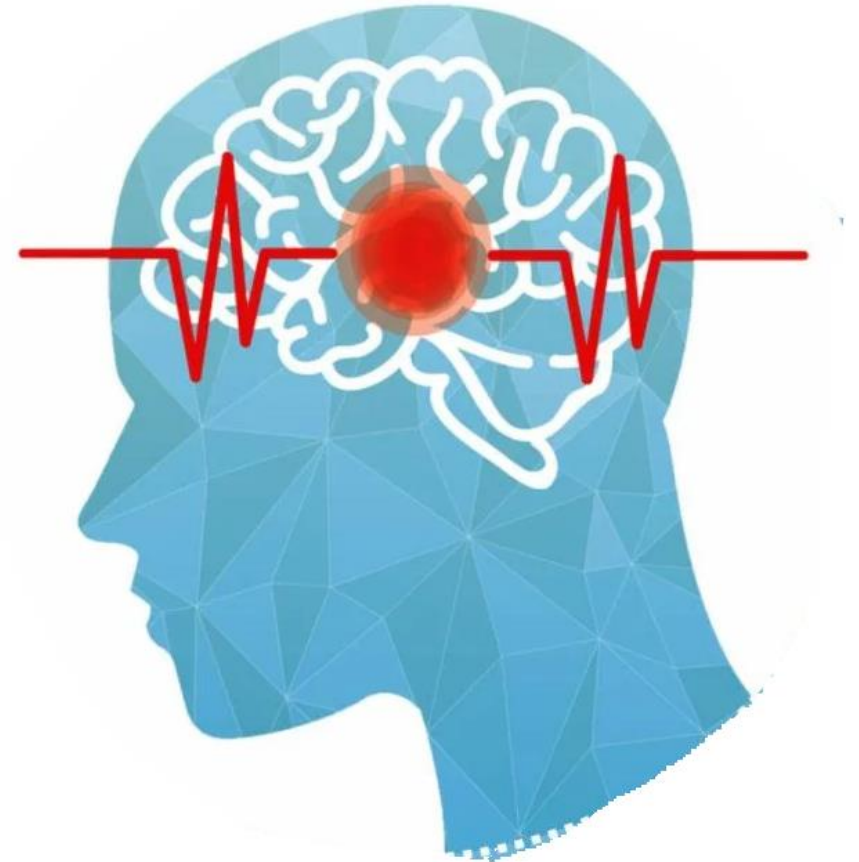
- **질병 진단 및 치료에서 예방 및 관리로 트렌드 변화**

- ✓ 국민 생활 수준 상향 평준화
- ✓ 의료 기술 발전
- ✓ COVID-19

뇌졸중

- 뇌 혈관계 질병.
- 전 세계 두 번째 주요 사망 원인.
- 다른 질병과 상관관계 높음.

- ✓ 고혈압
- ✓ 콜레스테롤
- ✓ 당뇨병
- ✓ 심장병
- ✓ etc.



목표

◆ AI 모델 구축

- 기초 정보를 활용한 질병 예측.
- 발병 위험을 파악 및 정밀 검진 필요 환자 구분.
- 의료진의 의사 결정 지원.
- 발병에 영향을 미치는 요소 파악 및 관리 필요성 제시.

데이터

◆ Kaggle

ID	고유 식별 번호	Int
성별	남자/여자/그 외	Object
나이	0.08 ~ 82세	Int
고혈압	유/무	Int
심장병	유/무	Int
결혼	유/무	Int
근무유형	사기업/자영업자/공무원/미성년자/무직	Object
거주유형	지방/도시	Object
평균혈당	55.22 ~ 267.6	Float
BMI	10.3 ~ 80.1	Float
흡연	비흡연/금연/미상/흡연	Object
뇌졸중	유/무	Int

] 'target'

Raw data

(train.csv (15304, 12)
test.csv (10204, 11))

데이터

◆ Kaggle

ID	고유 식별 번호	Int	제거
성별	남자/여자/그 외	Object	binary 타입 변경
나이	0.08 ~ 82세	Int	
고혈압	유/무	Int	
심장병	유/무	Int	
결혼	유/무	Int	
근무유형	사기업/자영업자/공무원/미성년자/무직	Object	
거주유형	지방/도시	Object	
평균혈당	55.22 ~ 267.6	Float	
BMI	10.3 ~ 80.1	Float	
흡연	비흡연/금연/미상/흡연	Object	
뇌졸중	유/무	Int	



train

```
Female  9446
Male    5857
Other    1
```

test

```
Female  6338
Male    3865
Other    1
```

‘그 외’ sample이 1개이므로 다른 데이터에 편입

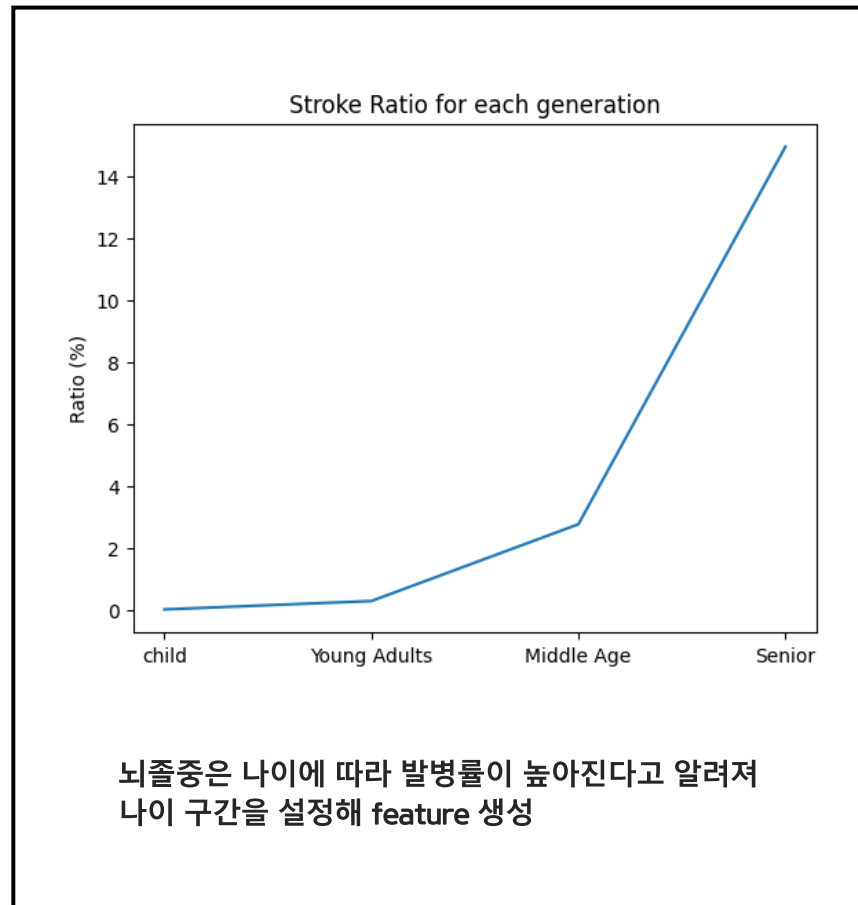
] ‘target’

데이터

◆ Kaggle

ID	고유 식별 번호	Int	제거
성별	남자/여자/그 외	Object	binary 타입 변경
나이	0.08 ~ 82세	Int	feature 추가
고혈압	유/무	Int	
심장병	유/무	Int	
결혼	유/무	Int	
근무유형	사기업/자영업자/공무원/미성년자/무직	Object	
거주유형	지방/도시	Object	
평균혈당	55.22 ~ 267.6	Float	
BMI	10.3 ~ 80.1	Float	
흡연	비흡연/금연/미상/흡연	Object	
뇌졸중	유/무	Int	

] 'target'

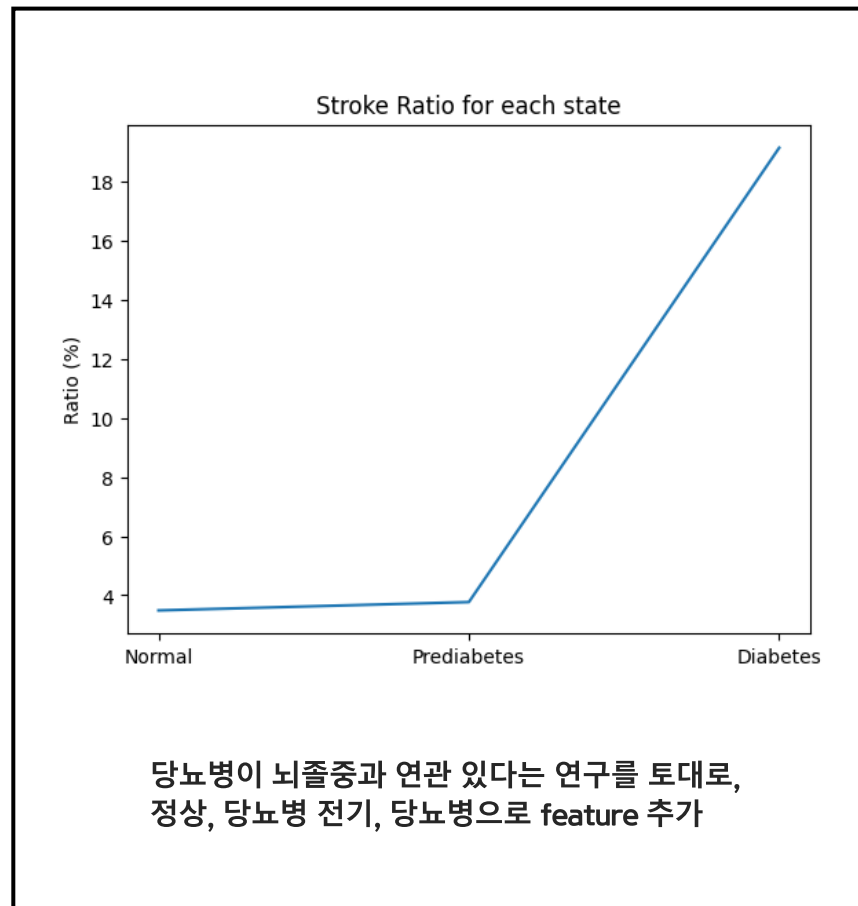


데이터

◆ Kaggle

ID	고유 식별 번호	Int	제거
성별	남자/여자/그 외	Object	binary 타입 변경
나이	0.08 ~ 82세	Int	feature 추가
고혈압	유/무	Int	
심장병	유/무	Int	
결혼	유/무	Int	
근무유형	사기업/자영업자/공무원/미성년자/무직	Object	Encoding
거주유형	지방/도시	Object	binary 타입 변경
평균혈당	55.22 ~ 267.6	Float	feature 추가
BMI	10.3 ~ 80.1	Float	
흡연	비흡연/금연/미상/흡연	Object	
뇌졸중	유/무	Int	

] 'target'

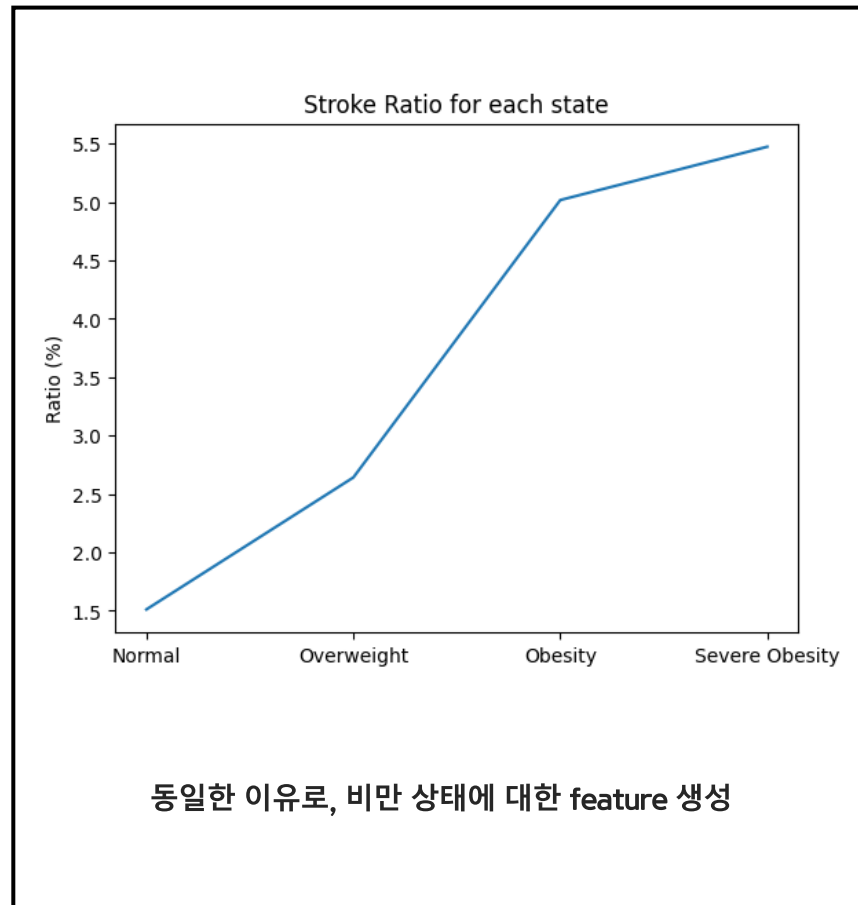


데이터

◆ Kaggle

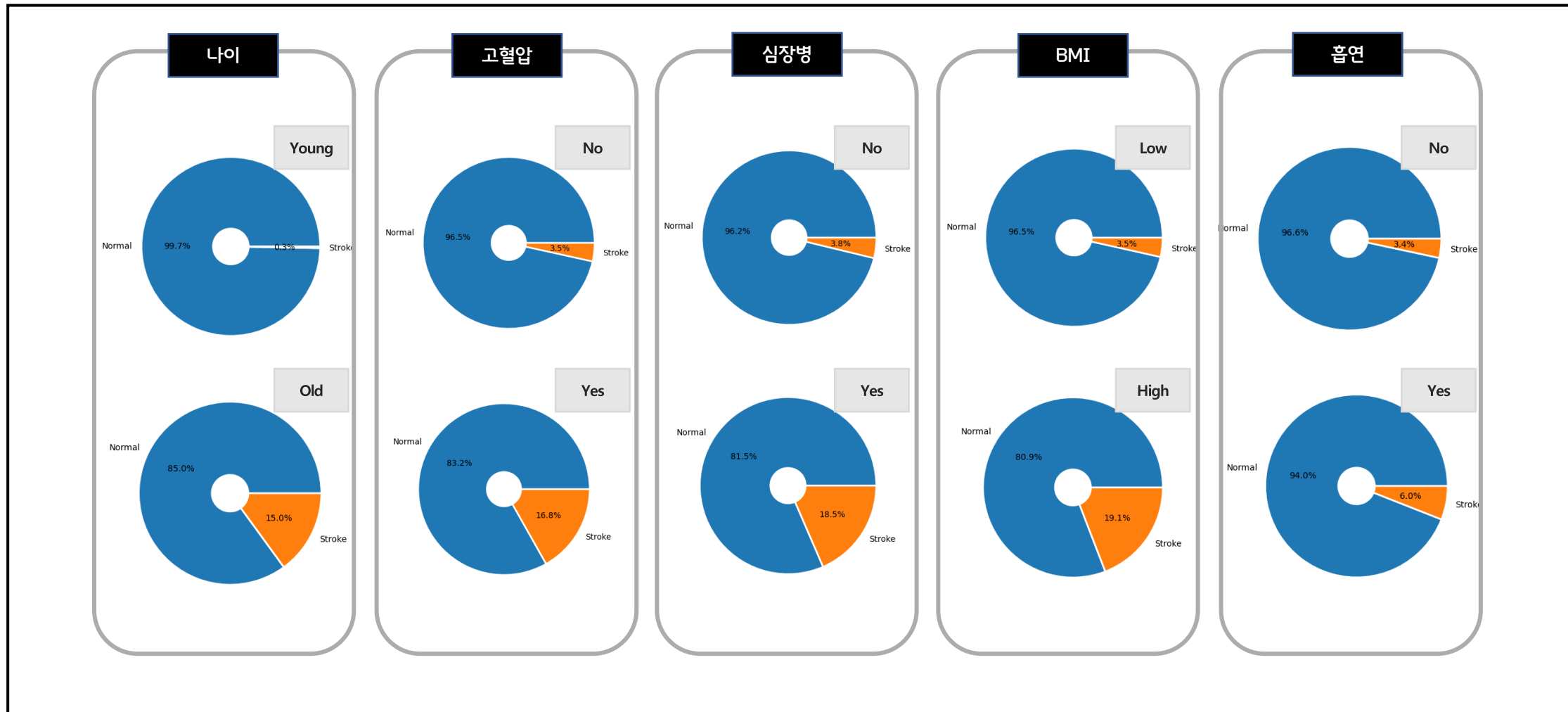
ID	고유 식별 번호	Int	제거
성별	남자/여자/그 외	Object	binary 타입 변경
나이	0.08 ~ 82세	Int	feature 추가
고혈압	유/무	Int	
심장병	유/무	Int	
결혼	유/무	Int	
근무유형	사기업/자영업자/공무원/미성년자/무직	Object	Encoding
거주유형	지방/도시	Object	binary 타입 변경
평균혈당	55.22 ~ 267.6	Float	feature 추가
BMI	10.3 ~ 80.1	Float	feature 추가
흡연	비흡연/금연/미상/흡연	Object	Encoding
뇌졸중	유/무	Int	

] 'target'



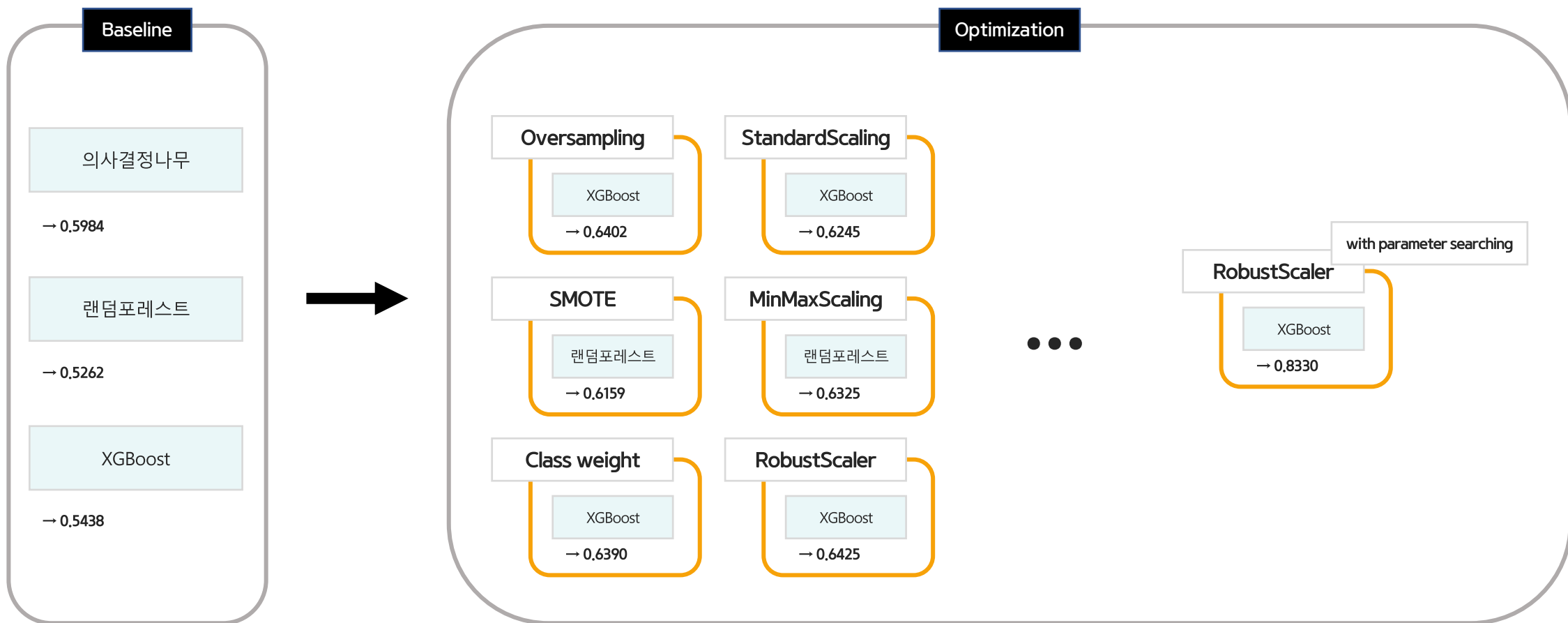
데이터

“5가지 위험 지표를 반영하는 feature 생성”



모델 구축

AUC 점수를 평가 지표로 학습 수행



결과

Kaggle 제출 결과

Submissions

You selected 0 of 2 submissions to be evaluated for your final leaderboard score. Since you selected less than 2 submission, Kaggle auto-selected up to 2 submissions from among your public best-scoring unselected submissions for evaluation. The evaluated submission with the best Private Score is used for your final score.

☒ Submissions evaluated for final score

All Successful Selected Errors

Submission and Description

Private Score ⓘ



submission1.csv

Complete (after deadline) · 1d ago

0.89381

approximately 80% of the test data.
board reflects the final standings.

Members

Score

Entries

Last

Lisle



0.90085

22

3mo



0.90075

35

3mo



0.90037

7

3mo



0.90003

10

3mo



0.90002

17

3mo



0.89945

32

3mo



0.89943

10

3mo



0.89921

1

3mo



0.89916

5

3mo



0.89916

17

3mo

4

▲ 352

Akahachi



5

▲ 36

Jose Cáliz



6

▲ 162

Viktor Taran



7

▲ 153

Liam Presland



8

▲ 220

k0-takahashi



9

▲ 152

movingon



10

▲ 262

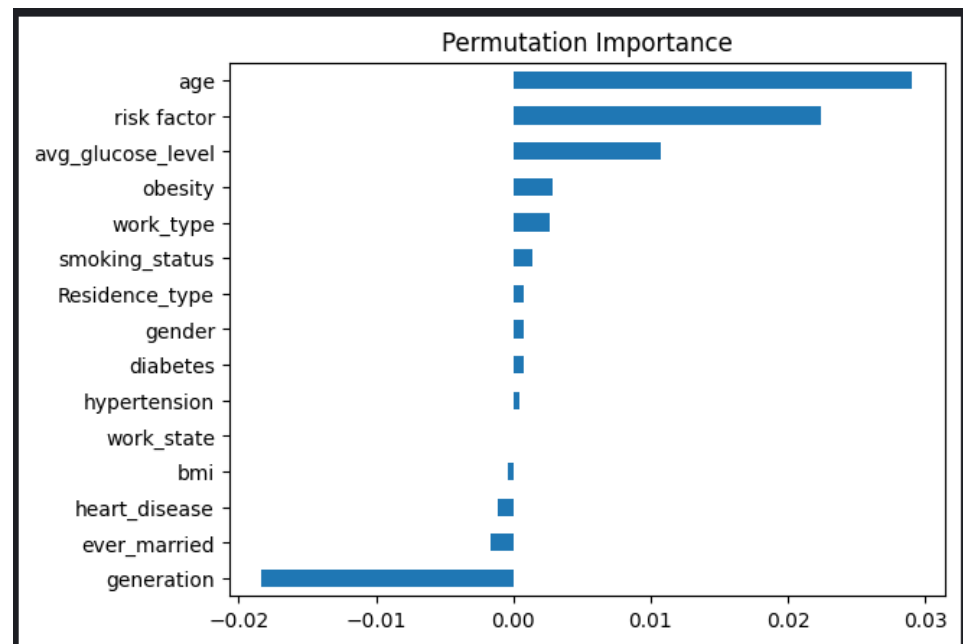
Bernhard Klinger



결과

◆ 결과 및 개선 방안

- Baseline과 비교해 성능 향상.
- 건강 문진표 수준의 데이터로도 양호한 성능의 모델 개발 가능.
- Feature importance를 참고해 추가 탐색 진행.
- 기존의 모델 뿐만 아니라 새로운 모델 테스트 필요.
- 자체적으로 새로운 데이터 수집 시도.





Thank you.

2023. 04. 12
이진재