

난임 해결 모델 레시피

난임 과정에서 발생하는 전략적 모델 분리를 곁들인

Team PoBu

Contents

- Backgrounds
- Preprocess
- Data Partition
- Modeling
- Use Case



배경

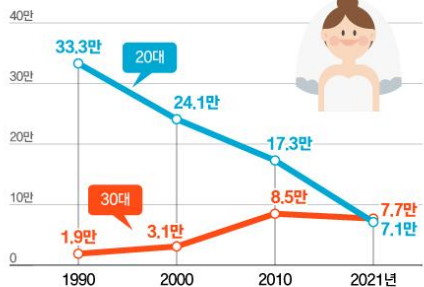
증가하고 있는 난임 문제

초혼 나이의 증가

해를 거듭할 수록
결혼 나이가 늦어지는 현상

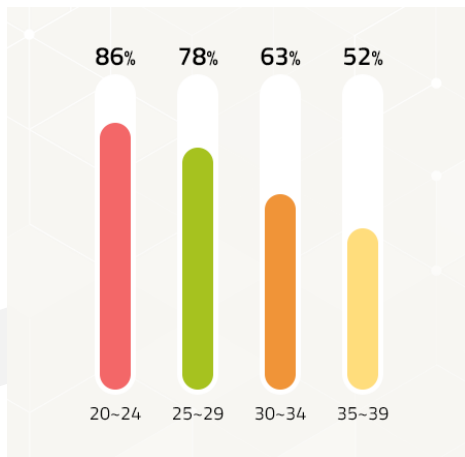
20~30대 여성 초혼 건수 추이

단위: 건



자료: 통계청 국가통계포털(KOSIS) The JoongAng

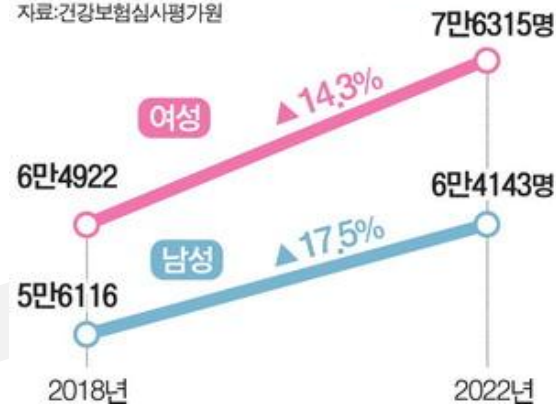
나이별 임신 가능성



연령이 높아질 수록
기하급수적으로 작아지는 확률

최근 5년간 난임 시술 진료 환자 수

자료: 건강보험심사평가원



난임 환자의 증가 추세

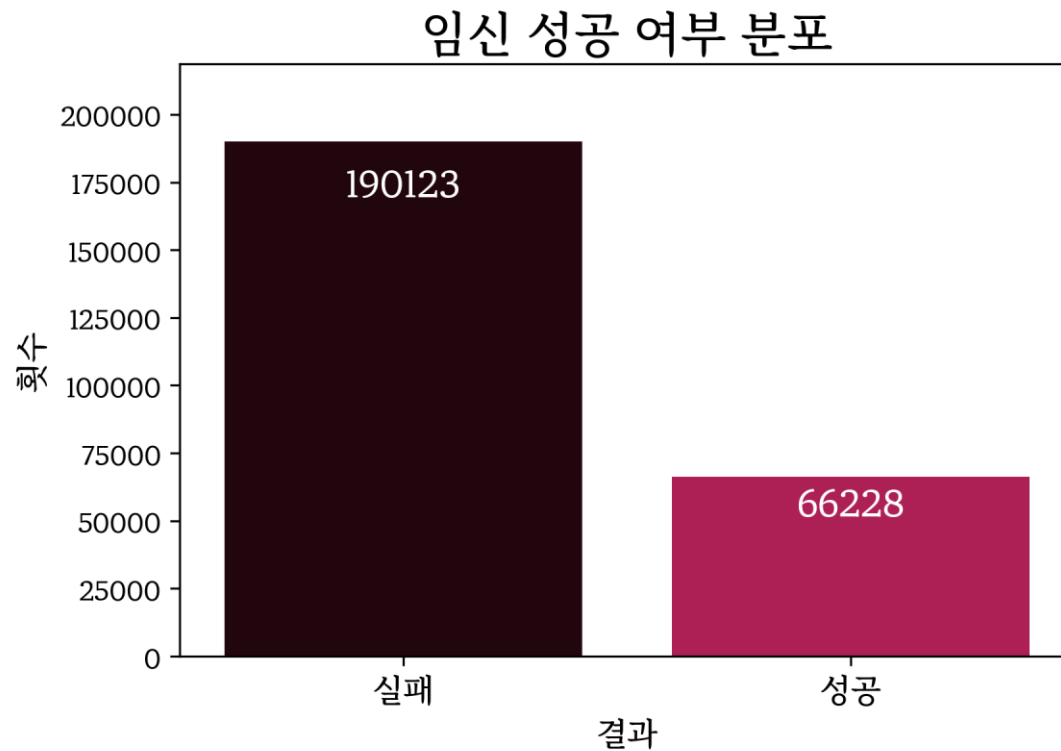
난임을 겪고 있는 부부의 수
지속적인 증가

난임을 미리 예측 할 수 있는 AI모델 개발 필요



배경

반응변수 특징



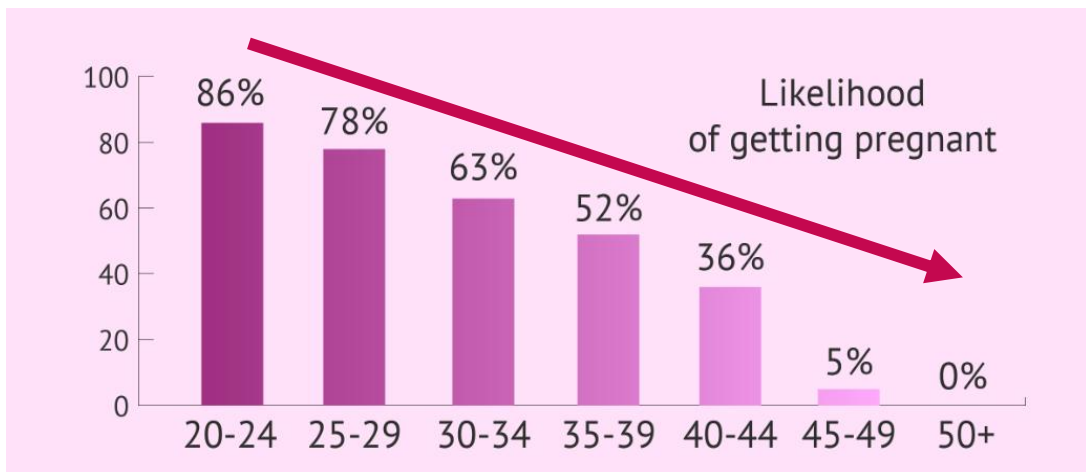
불균형이 약 3:1 정도 존재하지만
평가 지표 **AUC가 불균형에 강건한 특징**을 가지고 있기 때문에 **가중치 없이** 그대로 사용



전처리

시술 당시 나이 매핑

나이에 따른 출산율의 변화



나이가 클 수록 출산율이 급격하게 떨어지는 현상을 보임

매핑 숫자 편성도

나이	일반적인 매핑	출산율 반영 매핑
만18-34세	1	1024
만35-37세	2	512
만38-39세	3	256
만40-42세	4	128
만43-44세	5	32
만45-50세	6	4
알 수 없음	0 or -999	-999

젊은 나이일 수록 높은 숫자를 부여
기하적인 감소의 의미를 부여하기 위해 2의 제곱수 활용



전처리

시술 시기 코드 매핑

P: 구간별 임신 성공 비율

시술 당시 나이	만18-34세	만35-37세	만38-39세	만40-42세	만43-44세	만45-50세	알 수 없음
시술 시기 코드							
TRXQMD	0.3988	0.2333	0.1632	0.1402	0.0410	0.0226	0.0006
TRZKPL	0.4091	0.2243	0.1598	0.1410	0.0429	0.0221	0.0007
TRVNRV	0.4130	0.2227	0.1560	0.1395	0.0440	0.0237	0.0009
TRJXFG	0.4097	0.2193	0.1518	0.1459	0.0446	0.0266	0.0016
TRYBLT	0.4003	0.2257	0.1493	0.1465	0.0493	0.0271	0.0016
TRCMWS	0.3924	0.2267	0.1433	0.1502	0.0541	0.0314	0.0014
TRDQAZ	0.3769	0.2257	0.1494	0.1549	0.0570	0.0340	0.0019

위에서부터 아래로 1 ~ 7의 숫자로 매핑 적용



전처리

주요 생성 및 처리 변수

01

해당주기 난자주인 나이

생성 목적

시술에 사용하는
난자의 출처가
기증, 본인인지에 따라
난자 주인의 나이를 기입

방법

범주를 맞추기 위해
어릴수록 높은 숫자 매핑 적용

02

이식률

생성 목적

총 생성된 배아 중
실제로 이식된 배아의 비율을
퍼센트로 나타낸 것

방법

$\frac{\text{이식된 배아 수}}{\text{총 생성 배아 수}}$

03

나이 x 이식 수

생성 목적

나이가 어리고
이식된 배아수가 많을수록
임신 가능성이 높도록
반영한 변수

방법

시술당시 나이 X 이식된 배아 수

04

특정 시술 유형

생성 목적

각 시술과 동일 시술 내 세부
조합의 의미를
포함하여 만든 변수

방법

어떤 시술을 받았는지에 따라
One-Hot Encoding 진행



전처리

그 외 생성 및 전 처리

01

난자 수, 배아 수 보정

- 난자, 배아의 수요와 공급 수 일치하도록 조절
- 폐기 및 잉여 배아 수 등 개수에 관련된 파생변수 생성

02

정자 출처 보정

- 파트너 정자와 혼합된 난자와 기증된 정자와 혼합된 난자가 동시에 존재하는 경우 '배우자 및 기증 제공'으로 전환

03

과거력 관련 변수

- 시술, 임신, 출산 경험 횟수를 횟수 순서대로 매핑하나 6회 이상의 경우 20으로 매핑
- 실패 횟수, 유산횟수 계산

04

배란자극 X 이식 수

- 배란자극 유무 X 이식 수

05

불임 원인

- 불임 원인 모음
- 남성, 여성 모음
- 여성요인 제외 (단일 값 칼럼)

06

경과일 수

- 합산 경과일
- 결측치는 0으로 대체

07

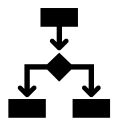
유전 검사/진단

- 검사/진단 유무와 해당 주기에 사용되었는지를 통합

08

결측 처리

- -999로 결측 치환



분할

기술 유형

DI

.....

E	F	AN	AO
기술 유	특정 시	총 생성	미세주
DI	IUI		
DI	IUI		
DI	IUI		
DI	IUI		
DI	IUI		

IVF

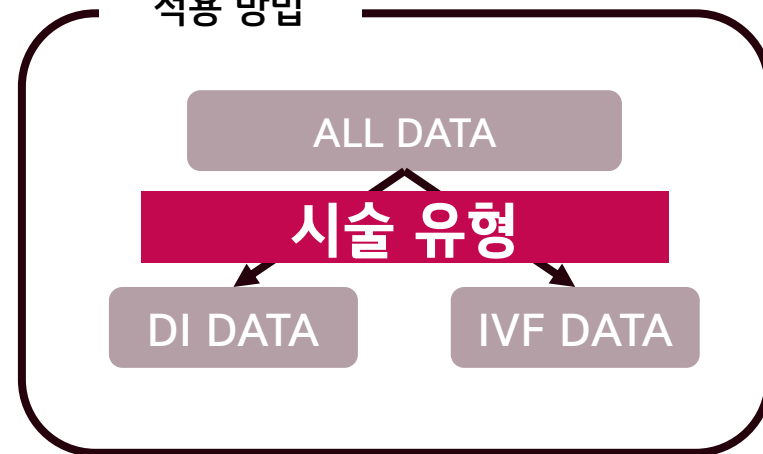
.....

E	F	AN	AO
기술 유	특정 시	총 생성	미세주
IVF	ICSI	4	5
IVF	ICSI	0	1
IVF	IVF	5	0
IVF	ICSI	0	4
IVF	ICSI	6	6
IVF	ICSI	3	5

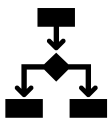
데이터 분할 결정

.....

적용 방법



전체 칼럼 약 130개 중 78개 칼럼에 대해 DI에서 결측 또는 단일 칼럼
모델을 나눠 적합하기로 결정



분할

현재 시술용

배아 생성 주요 이유

기증용

난자 저장용

배아 저장용

연구용

현재 시술용

데이터 확인

현재 시술용에 따른 임신 성공 여부 확률



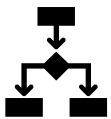
ID	시술 시	시술 당	임신 시	시술 유형	배아 생성 주요 이유	임신 성공 여부
TRAIN_035	TRCMWS	만40-42세		IVF	배아 저장용	1
TRAIN_045	TRDQAZ	만18-34세		IVF	배아 저장용	1
TRAIN_055	TRYBLT	만38-39세		IVF	배아 저장용	1
TRAIN_077	TRDQAZ	만43-44세		IVF	배아 저장용	1
TRAIN_107	TRYBLT	만18-34세		IVF	배아 저장용	1
TRAIN_116	TRDQAZ	만18-34세		IVF	배아 저장용	1
TRAIN_212	TRDQAZ	만43-44세		IVF	배아 저장용	1
TRAIN_227	TRCMWS	만18-34세		IVF	배아 저장용	1

특수 케이스를 감안하여 현재 시술용이 없는 경우
임신을 원하지 않는 상황으로 판단하여 학습 배제

데이터 분할 결정

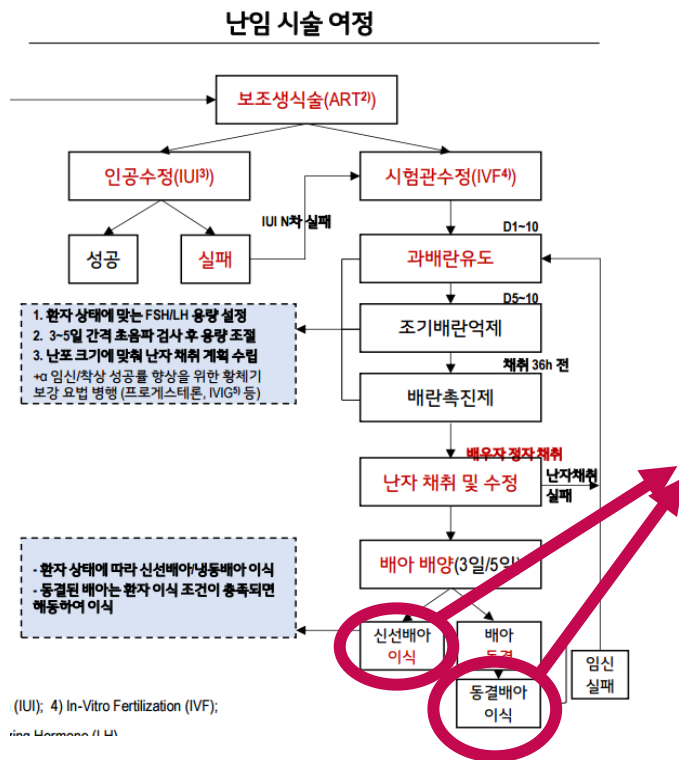
적용 방법





분할

이식된 배아 수



데이터 확인

이식된 배아 수	0	1
0.0	1.00	0.00
1.0	0.69	0.31
2.0	0.69	0.31
3.0	0.83	0.17

임신 성공 여부

데이터 분할 결정

적용 방법

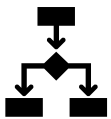
현재 시술용 IVF DATA

이식된 배아 수 > 0

No

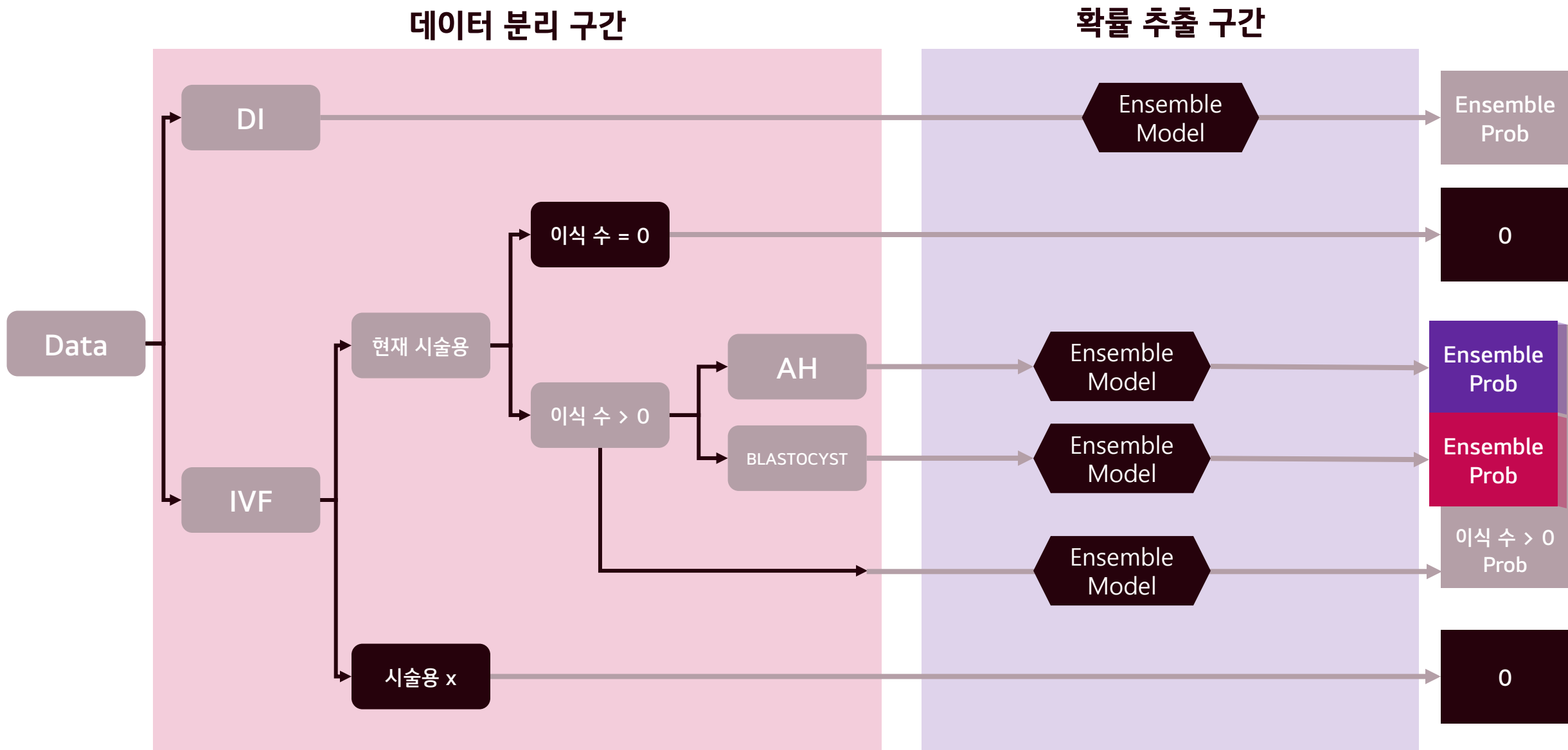
Yes

이식이 되지 않은 경우
시도 자체가 없는 것으로 판단



분할

모델 전체 파이프라인





모델링

주 사용 모델

Ensemble



CatBoost



LightGBM

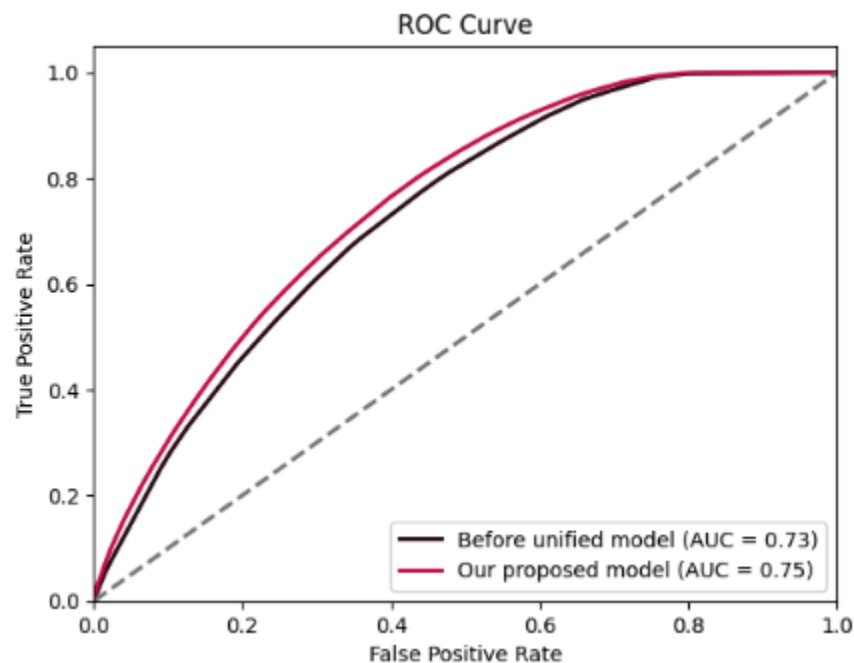
XGBoost



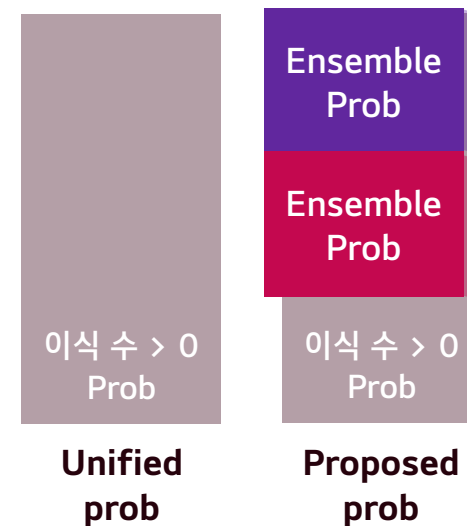
모델링

BLASTOCYST 분할 전후 Valid 비교

AUC 비교



통합 추론한 AUC보다
확률을 덮어 씌운 제안 모델의 AUC가 큰 것을 확인



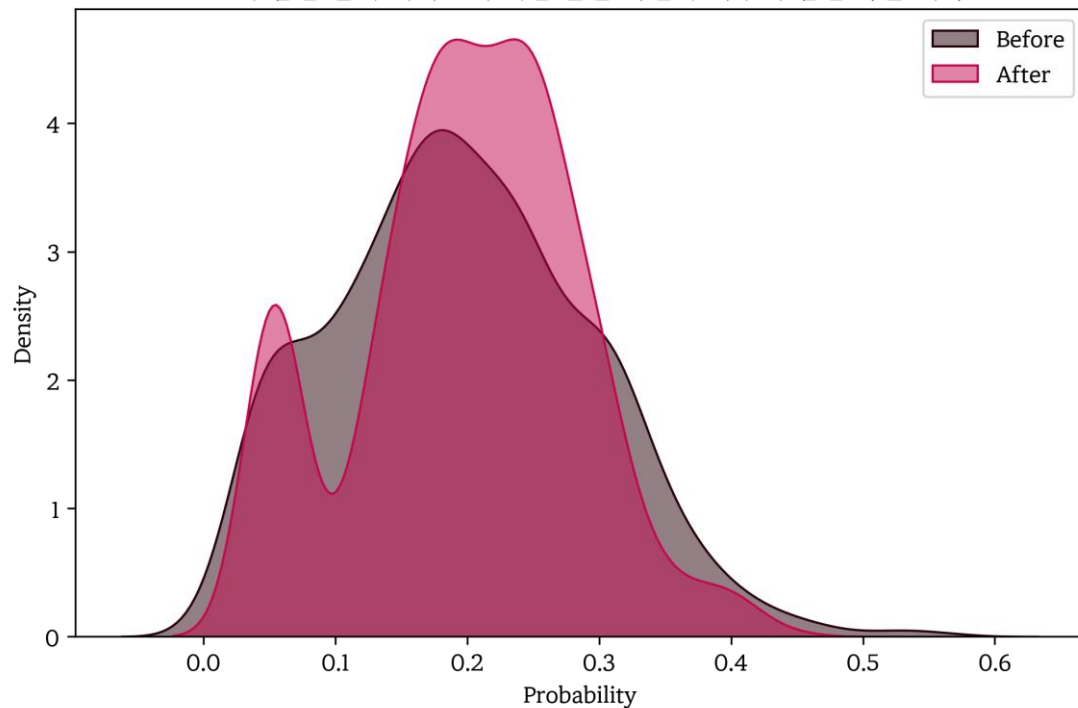


모델링

AH 분할 전후 Valid 비교

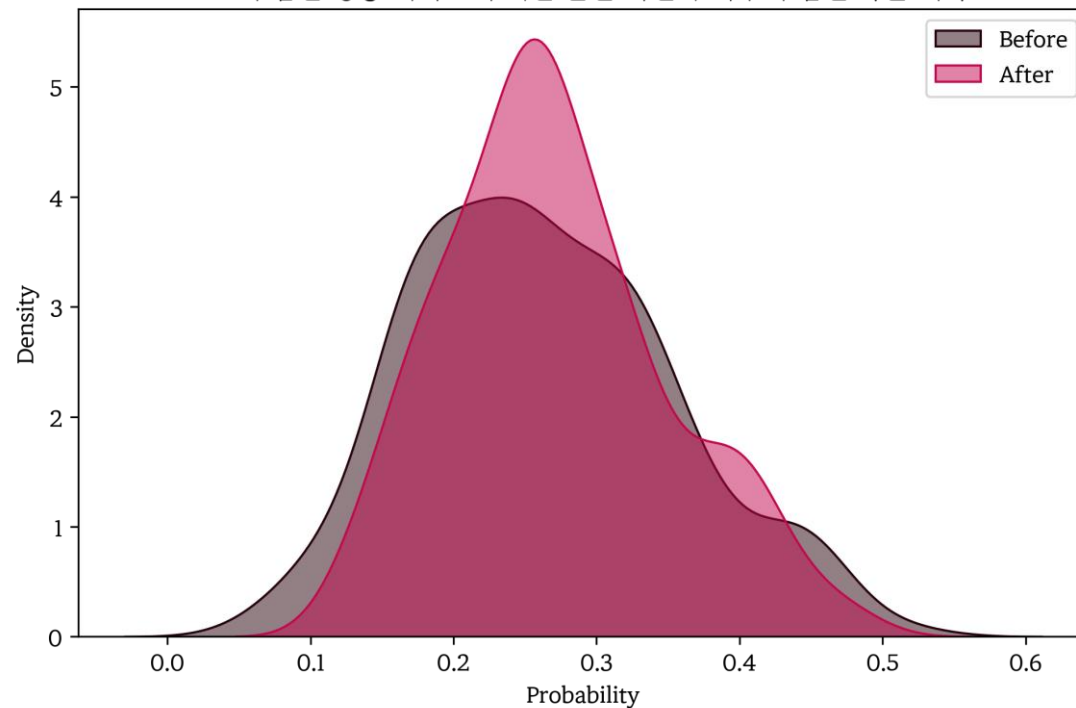
임신 실패 경향

AH의 임신 실패 케이스에 대한 분할 이전과 이후의 임신 확률 예측



임신 성공 경향

AH의 임신 성공 케이스에 대한 분할 이전과 이후의 임신 확률 예측



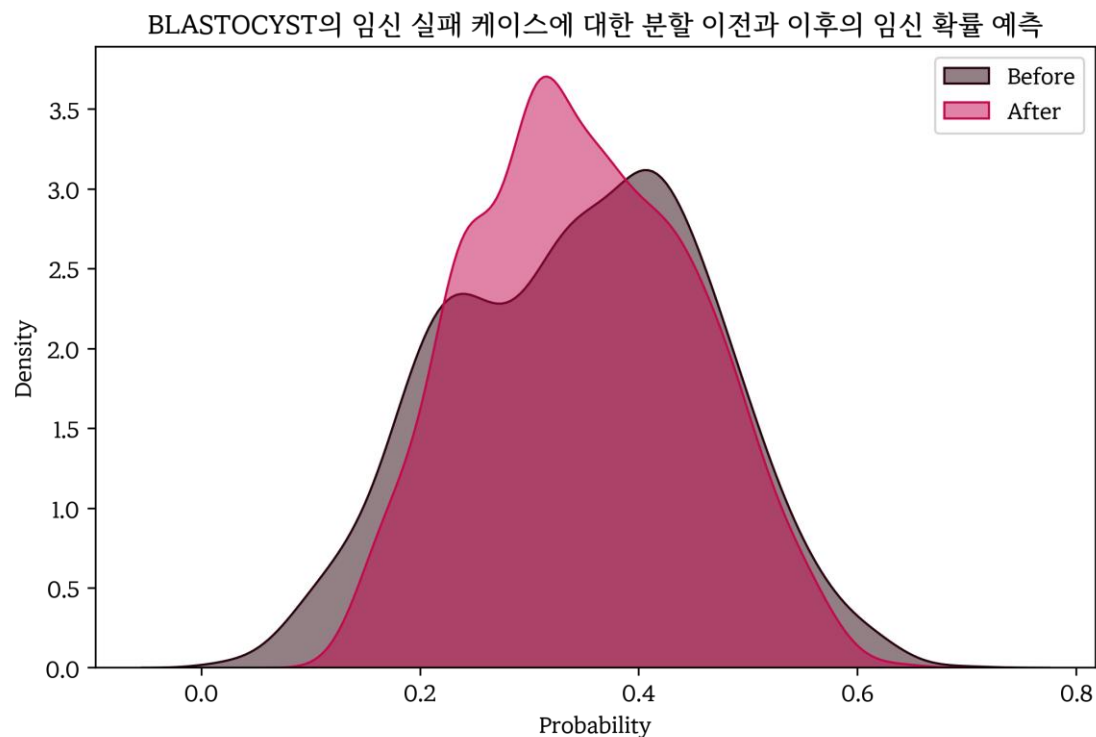
임신 실패에서 0.1 구간이 양쪽으로 퍼지는 형상을 보이지만
임신 성공 구간에서 미세하게 확률 증가 확인 가능



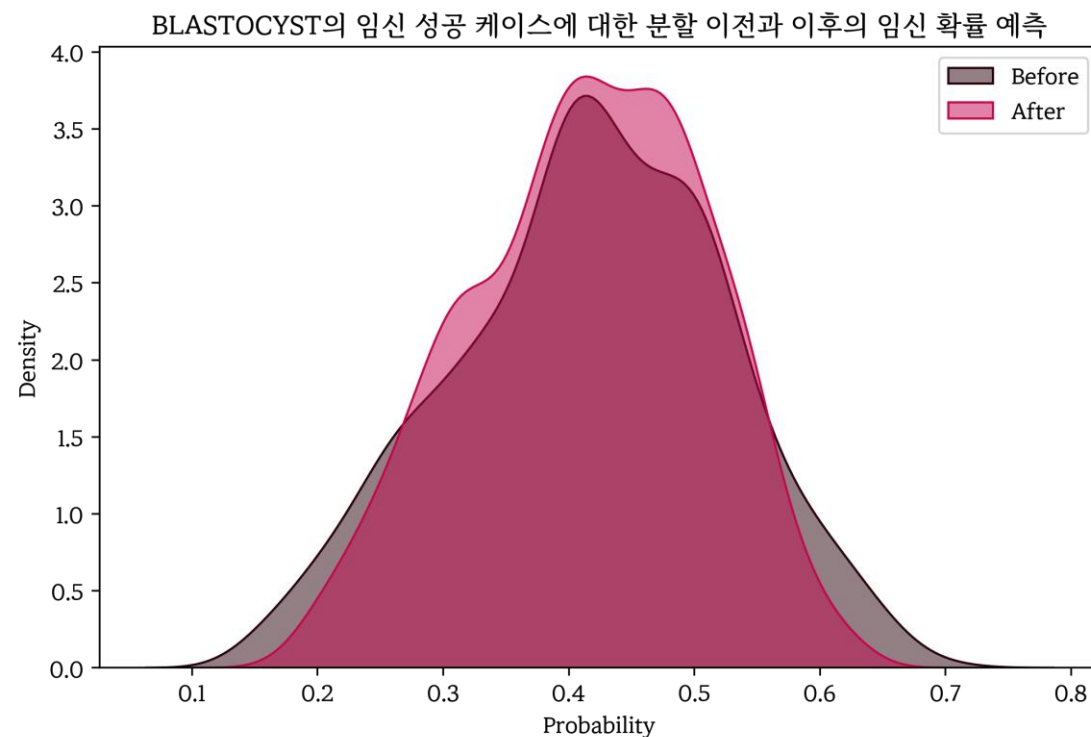
모델링

BLASTOCYST 분할 전후 Valid 비교

임신 실패 경향



임신 성공 경향



임신 실패에서 확률이 미세하게 감소하는 경향과
임신 성공 구간에서 미세하게 확률 증가하는 경향 확인 가능



모델 분기점	Public AUC
완전 통합 모델	0.7391664652
시술 유형 분할	0.7423765458
현재 시술용 반영	0.7425092228
이식된 배아 수 반영	0.7425750004
확률 OverWrite	0.7435055963

A photograph of a family walking away from the camera in a park during sunset. A man in a grey jacket and jeans, a young girl in a blue top and white tutu, and a woman in a checkered shirt and khaki pants are holding hands. The scene is bathed in warm, golden light from the setting sun, with fallen leaves on the grass.

Thank You

