

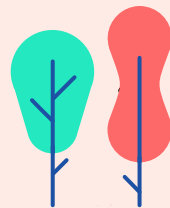


재원일수 예측모델

다중 분류 모형으로 효과적으로
재원일수 관리하기



Table of Contents



01

개요

재원일수란?
문제 정의

02

데이터 탐색 및 전처리

데이터 탐색 및 전처리
가설 설정

03

머신러닝 모델 적용

모델 학습 및 성능 비교
모델 평가

04

모델 해석 및 인사이트

모델 해석
인사이트 및 한계점





재원일수란?

재원일수란 병원에 입원한 환자가
병상을 점유한 일수 (OECD, 2013)

평균 재원일수는 진료결과와
의료서비스 질적 변이를 간접적으로
진단하는 가장 유용한 도구

재원일수의 영향



01

환자 측면

- 환자의 병원 내 감염에 대한 노출 또는 합병증 증가
- 의료비 부담 증가

02

병원 시스템 측면

- 급성기 환자 치료 기회 감소
- 병상 회전을 감소로 수익 감소

03

재원일수 관리 서비스

- 머신러닝 예측모형으로 효과적으로 재원일수 관리하여 의료 서비스의 효율성 증대

02

데이터 탐색 및 전처리

- 데이터 개요
- 타겟의 분포
- 타겟과 특성 간의 관계
- 데이터 전처리
- 가설 설정



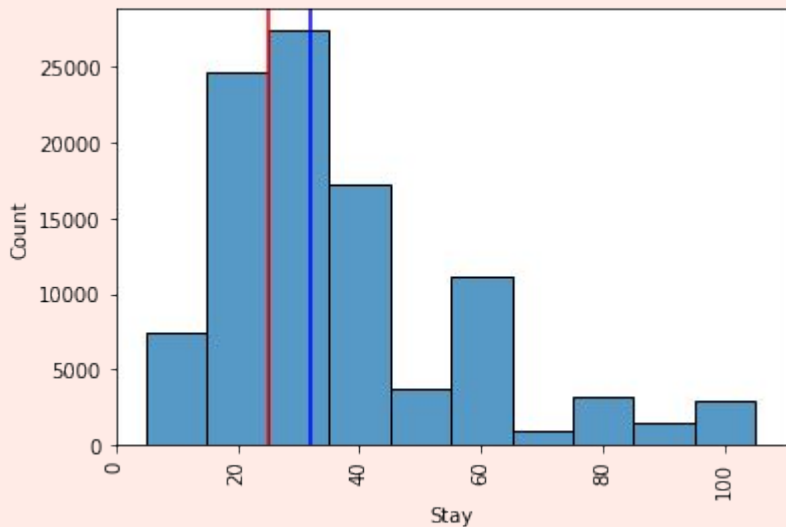
데이터 개요

샘플 수 318438

특성 수 18

Column		
0	case_id	케이스 ID
1	Hospital_code	병원 코드
2	Hospital_type_code	병원 유형 코드
3	City_Code_Hospital	병원 도시 코드
4	Hospital_region_code	병원 지역 코드
5	Available Extra Rooms in Hospital	사용할 수 있는 추가 병실 수
6	Department	담당 부서
7	Ward_Type	병동 유형
8	Ward_Facility_Code	병동 시설 코드
9	Bed Grade	침대 등급
10	patientid	환자 ID
11	City_Code_Patient	환자 도시 코드
12	Type of Admission	입원 유형
13	Severity of Illness	중등도
14	Visitors with Patient	방문객 수
15	Age	나이
16	Admission_Deposit	입원 보증금
17	Stay	재원일수

타겟의 분포



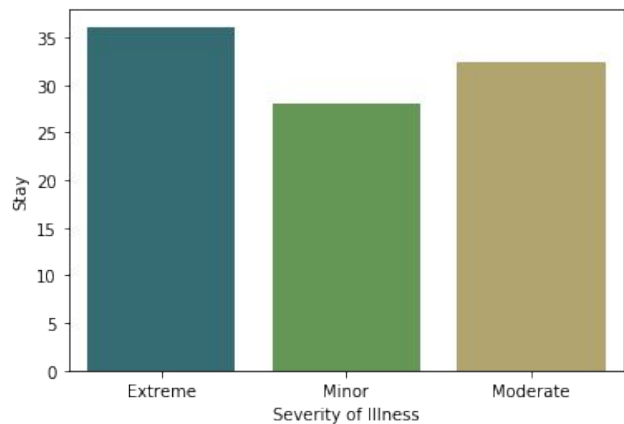
타겟의 분포

OECD 평균 재원일수 8.1
한국 평균 재원일수 19.1

count	100000.000000
mean	31.917700
std	21.449686
min	5.000000
25%	15.000000
50%	25.000000
75%	35.000000
max	105.000000
Name: Stay, dtype: float64	

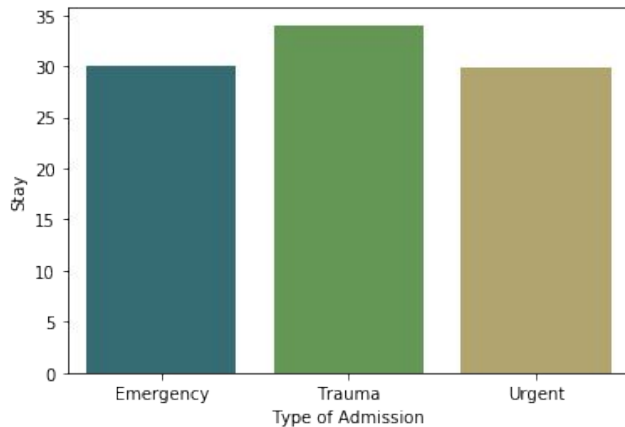
타겟의 평균

타겟과 특성 간의 관계

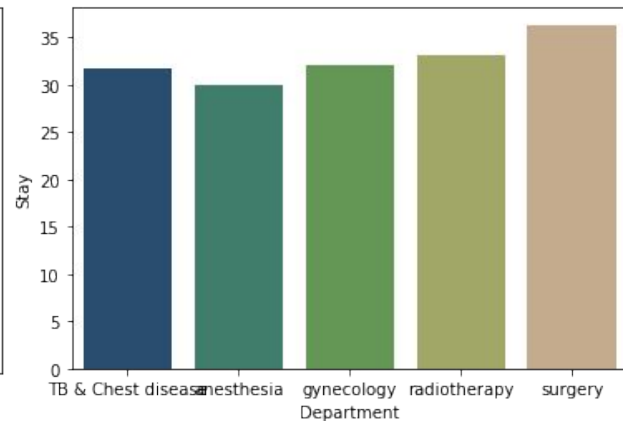


○

중증도



입원 유형

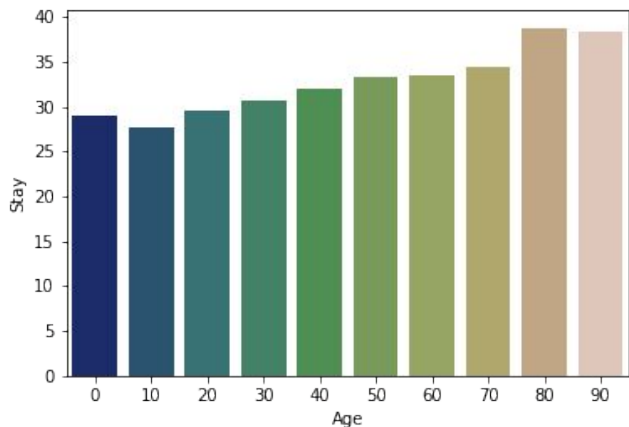


입원 부서

○

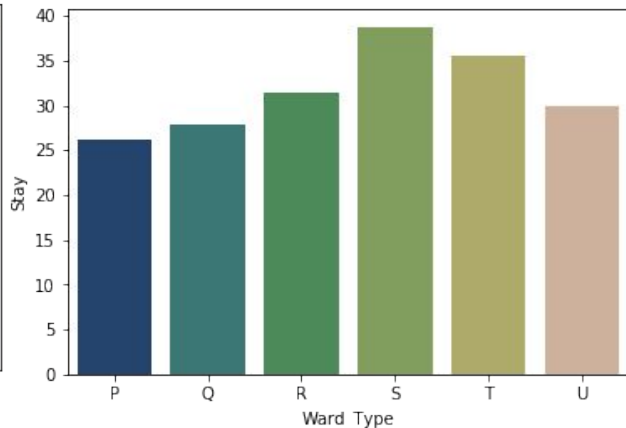
○

타겟과 특성 간의 관계

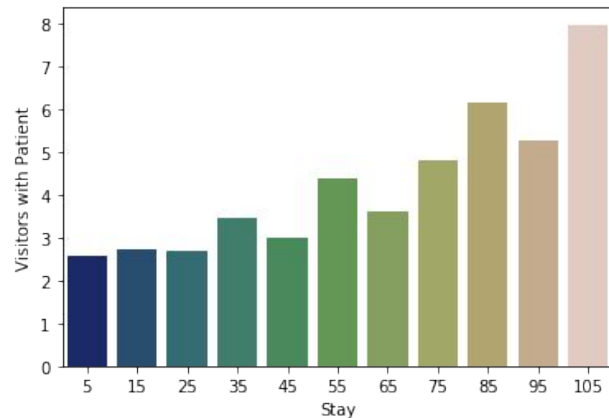


○

나이



병동 유형



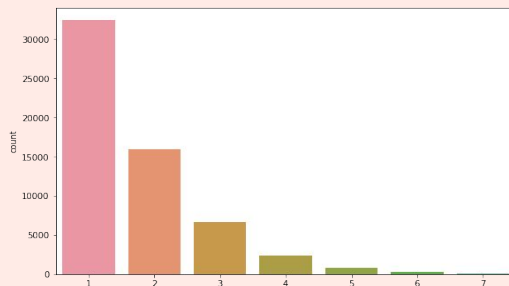
방문객 수는 퇴원 후에 알 수 있는
특성이므로 데이터 누수 방지를 위해
방문객 수 특성 제거 ○

데이터 전처리

결측값 처리

전체 31만개 데이터 중 1.46%
해당하는 결측값 제거

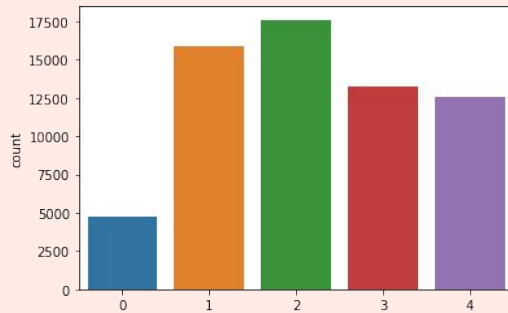
특성 공학



환자별 입원한 횟수
특성 추가

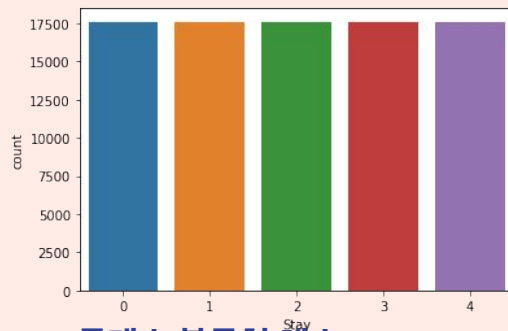


데이터 샘플링



데이터 10만개 샘플링

오버샘플링



클래스 불균형 해소

가설 설정

1

입원유형

입원 유형 **Urgent < Emergency < Trauma**
순으로 재원일수가
길어진다.

2

중증도

중증도 **Minor < Moderate < Extreme**
순으로 재원일수가
길어진다.

3

나이

나이가 많아질수록
재원일수가 길어진다.

03

머신러닝 모델 적용

- 모델링 과정
- 모델 성능 비교
- 모델 평가



모델링 과정

○○○

데이터 탐색

타겟 분포 시각화, 타겟과
특성 간의 관계 시각화



○○○

데이터 전처리

결측값 처리, 특성공학,
데이터 샘플링



○○○

다중 분류 모델 학습

결정 트리, 랜덤 포레스트,
XGB, LGBM 모델 Training

○○○

Grid Search CV

하이퍼파라미터 조정하여
모델 최적화



○○○

모델 성능 비교

Val dataset 으로 accuray,
f1 score 비교

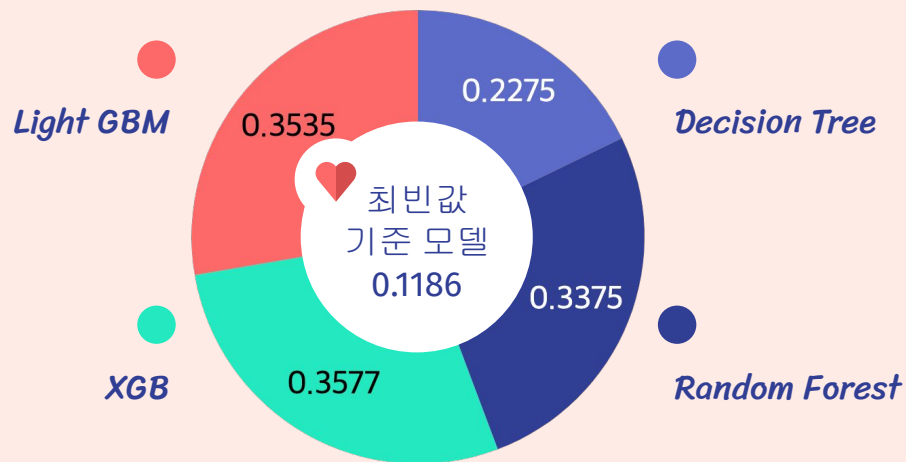


○○○

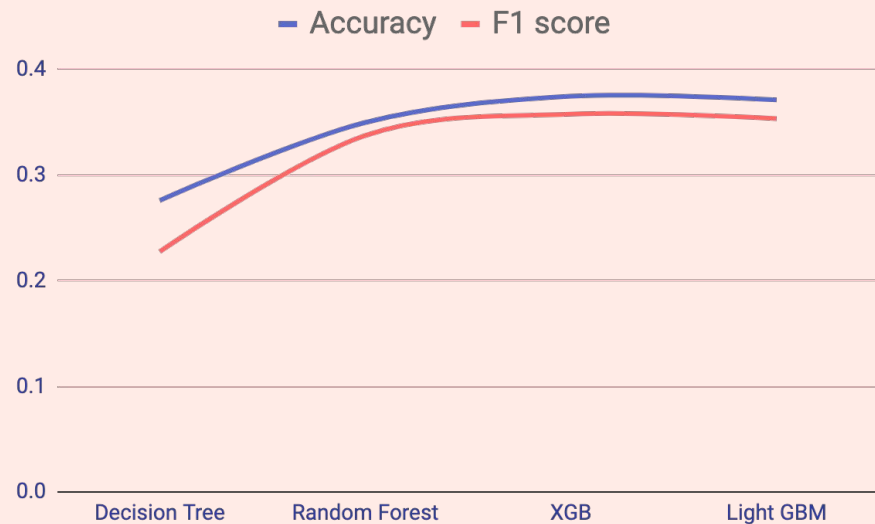
모델 평가

Test dataset 으로
accuray, f1 score, roc
auc score 평가

모델 성능 비교



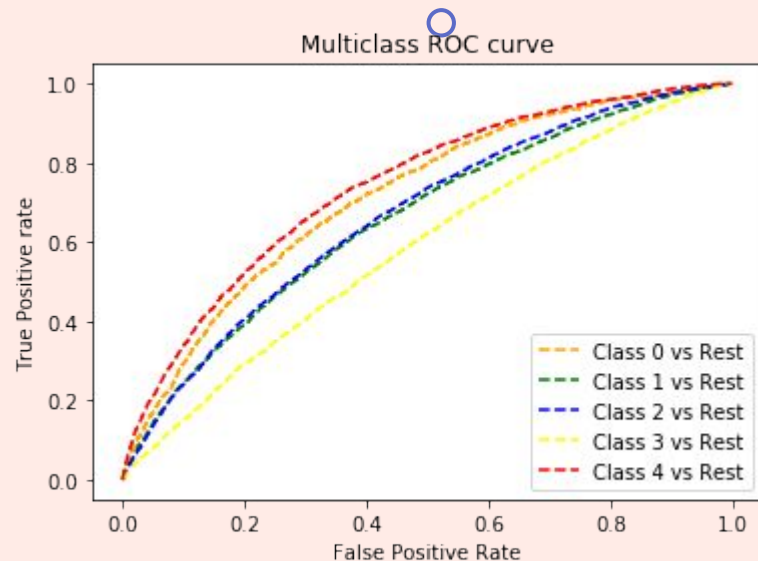
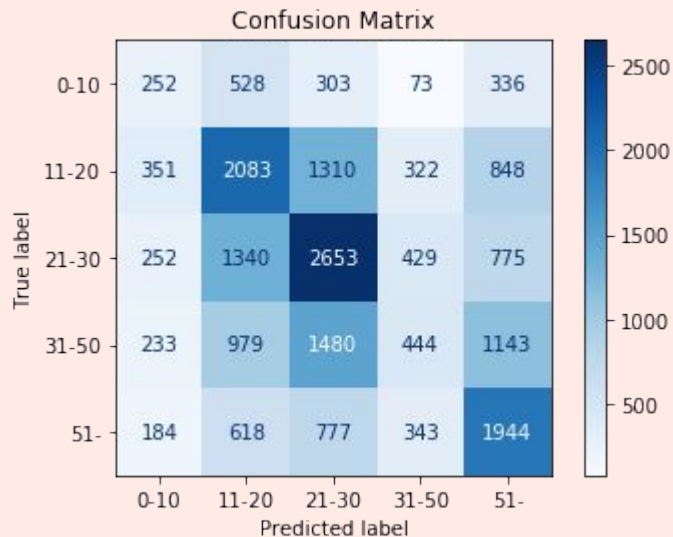
F1 score



Accuracy & F1 score

모델 평가

Accuracy 0.3714
f1 score 0.3541



Test Dataset

Confusion Matrix

ROC AUC score

04

모델 해석 및 인사이트 도출

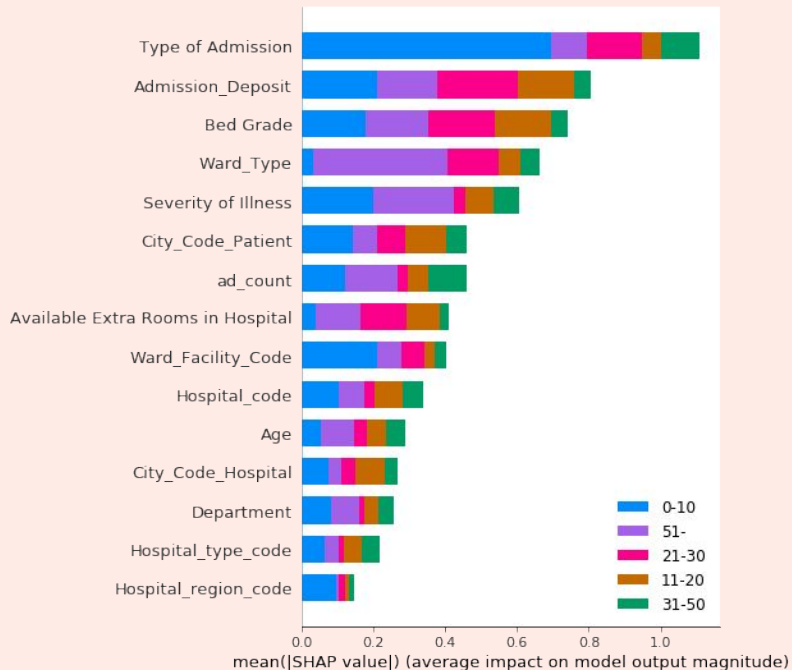
- 모델 해석
- 가설 검정
- 인사이트 도출
- 한계점 및 보완방안



모델 해석

Weight	Feature
0.0414 ± 0.0058	Ward_Type
0.0311 ± 0.0033	Bed Grade
0.0278 ± 0.0011	Admission_Deposit
0.0205 ± 0.0021	Severity of Illness
0.0163 ± 0.0046	Type of Admission
0.0145 ± 0.0031	City_Code_Hospital
0.0142 ± 0.0037	City_Code_Patient
0.0138 ± 0.0020	Ward_Facility_Code
0.0134 ± 0.0061	ad_count
0.0108 ± 0.0020	Hospital_code
0.0104 ± 0.0031	Available Extra Rooms in Hospital
0.0096 ± 0.0042	Hospital_type_code
0.0066 ± 0.0023	Hospital_region_code
0.0066 ± 0.0036	Age
0.0040 ± 0.0011	Department

재원일수에 영향을 주는 요인
 병동 유형, 침대 등급, 입원 보증금 순



10일 이하 재원일수에 영향을 주는 요인
 입원 유형, 병동 시설 순
50일 이상 재원일수에 영향을 주는 요인
 병동유형, 중증도 순

모델 해석

입원 유형

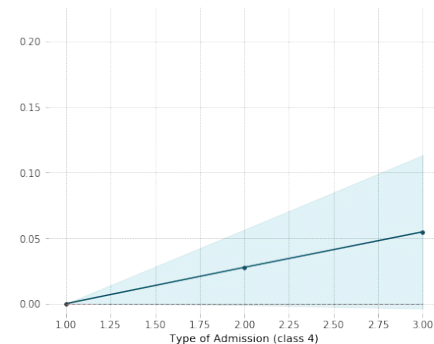
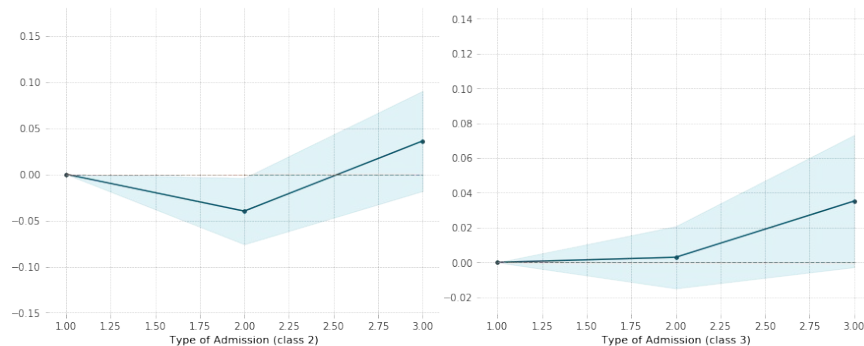
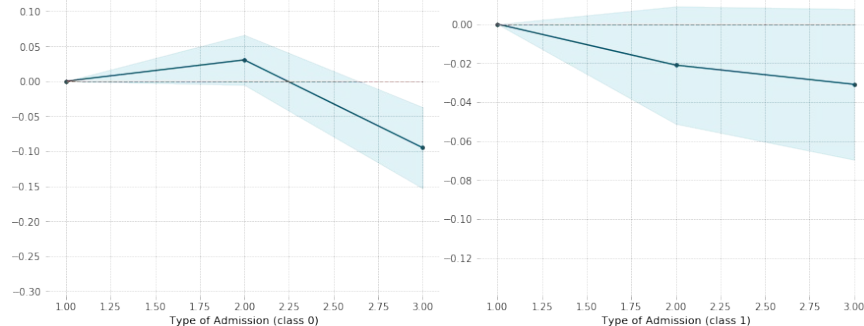
Urgent (긴급) : 걱정되는 의학적 상태나 질병이 있지만 생명을 위협하는 상태는 아닌 경우

Emergency (응급실) : 환자가 응급 질병과 부상을 입는 곳이며 그 중 일부는 생명을 위협할 수 있다.

Trauma (외상센터) : 외상성 부상의 예로는 자동차 사고, 총상 또는 추락으로 인한 부상이 있다.

Emergency - 10일 이하 또는 50일 이상 입원할 가능성이 높았지만 상대적으로 10 - 50일 입원할 가능성이 적었다.

Trauma - 20일 이상 입원할 가능성이 높았다.

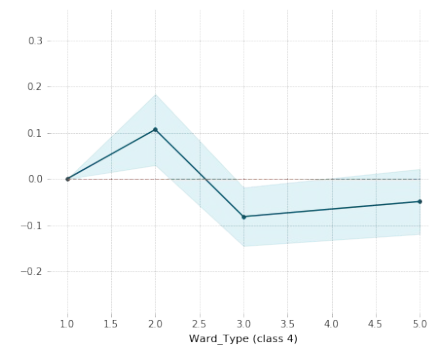
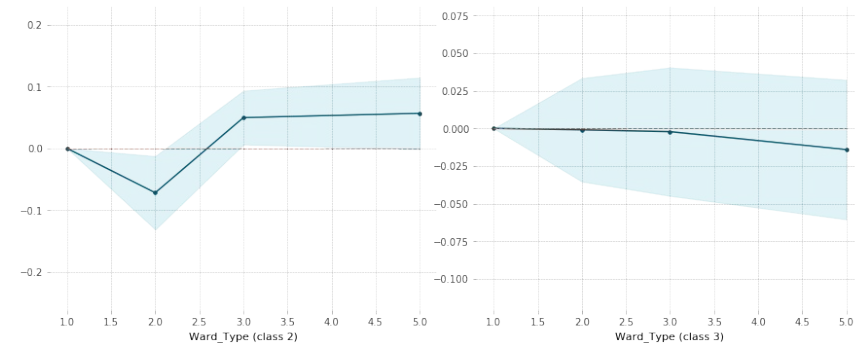
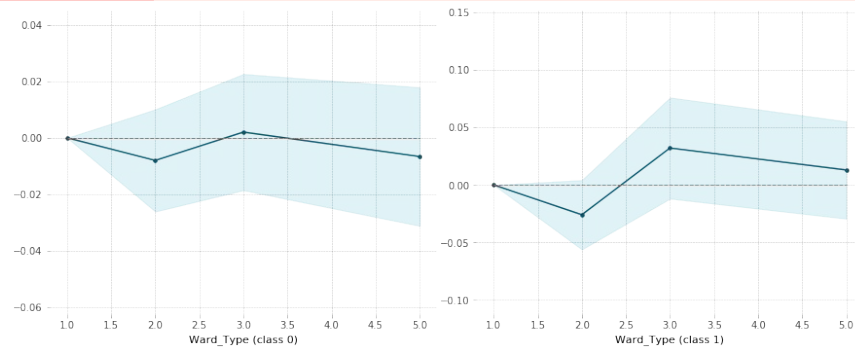


모델 해석

병동 유형

Q 병동 - 30일 이하 입원할 가능성이 높았다.

S 병동 - 50일 이상 입원할 가능성이 높았다.

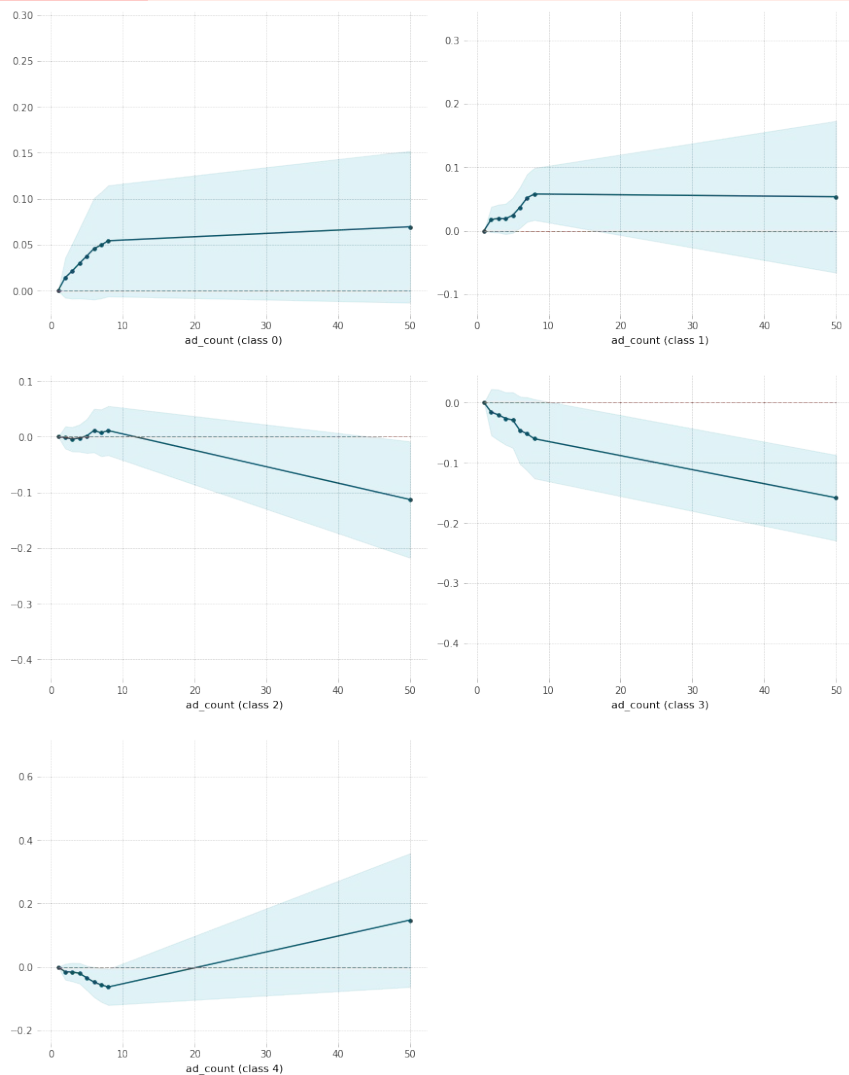


모델 해석

입원 횟수

입원한 횟수 8회 이하인 환자는
재원일수가 20일 이하일 가능성이 높았고,

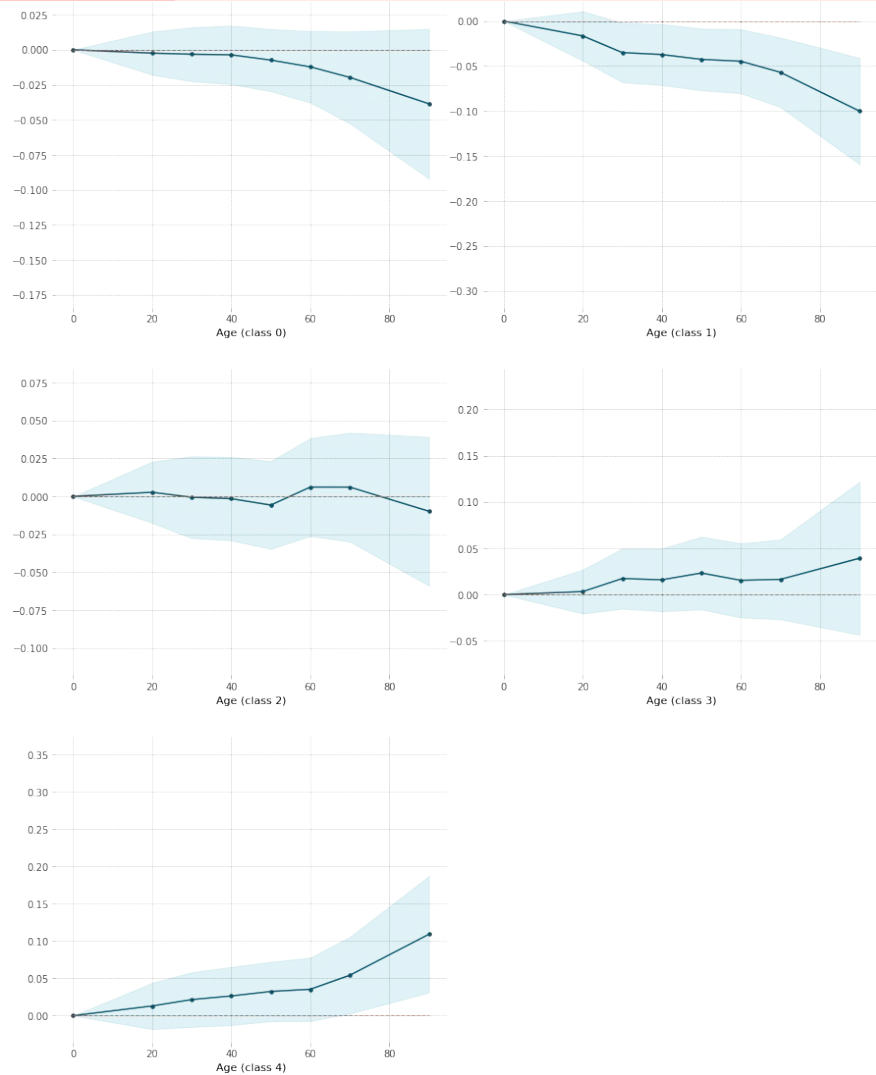
8회 이상인 환자는
재원일수가 50일 이상일 가능성이 높았다.



모델 해석

나이

나이가 많을수록 재원일수가 길어진다.

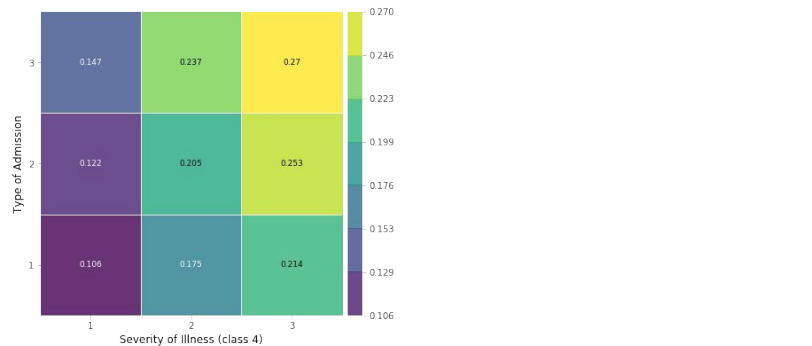
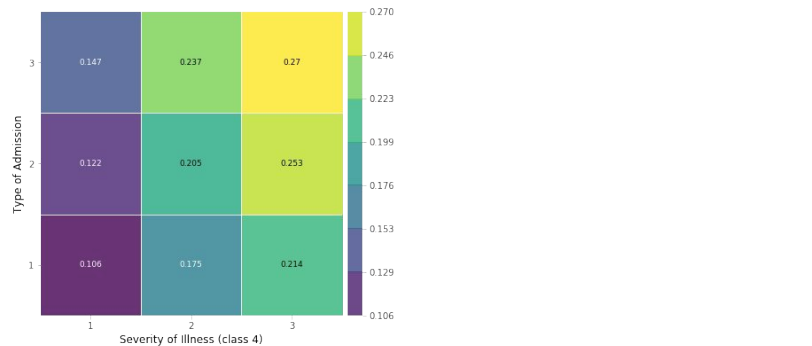
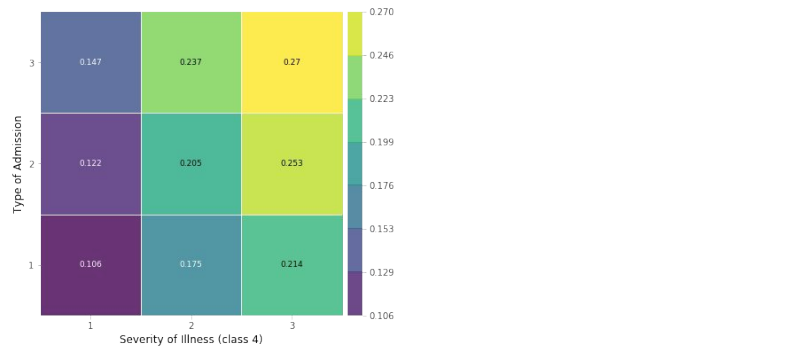
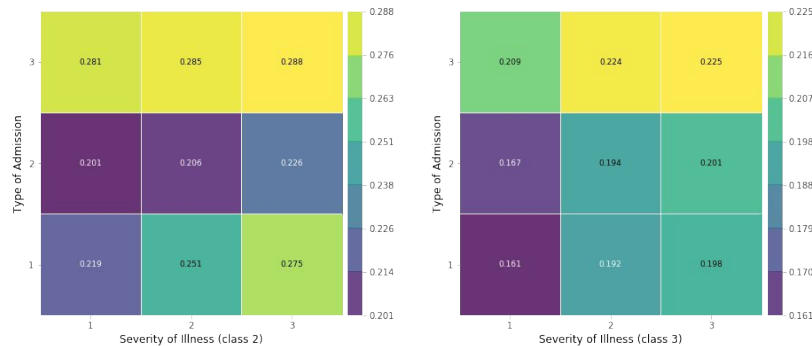
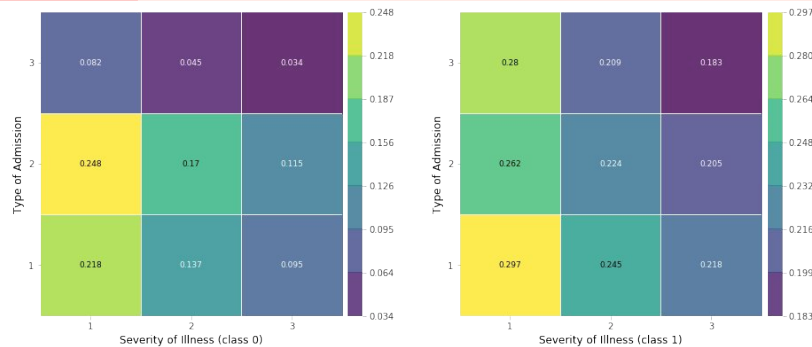


모델 해석

중증도와 입원유형

중증도가 낮고, 입원유형이 Urgent 일수록
재원일수가 짧다.

중증도가 높고, 입원유형이 Trauma 일수록
재원일수가 길다.



가설 검증

1

입원유형

입원 유형 Urgent <
Emergency < Trauma
순으로 재원일수가
길어졌다.

2

중증도

중증도 Minor <
Moderate < Extreme
순으로 재원일수가
길어졌다.

3

나이

나이가 많아질수록
재원일수가 길어졌다.

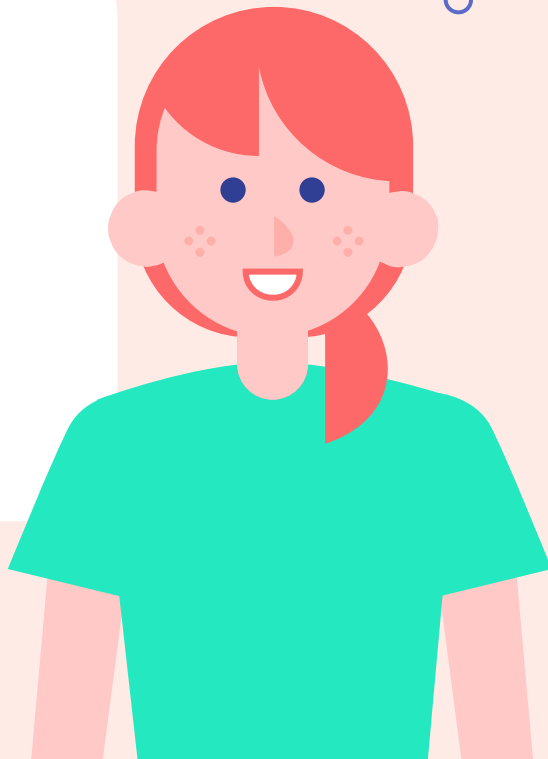
인사이트 도출

“

재원일수에 영향을 미치는 주요 요인들은 **병동유형, 침대등급, 입원보증금, 중증도, 입원유형** 등이 있습니다.

재원일수를 효과적으로 관리할 경우 **병상가동률 증가, 환자 당 입원비용 감소, 환자 진료 기회 증가** 등의 효율성을 증대시킬 수 있습니다.

맞춤형 재원일수 관리서비스는 장기입원 예측군, 집중관리군을 대상으로 **입원 시점에 장기입원 알람서비스, 재원일수 초과 시점에 재원일수 초과 알람서비스**로 진행될 수 있습니다.



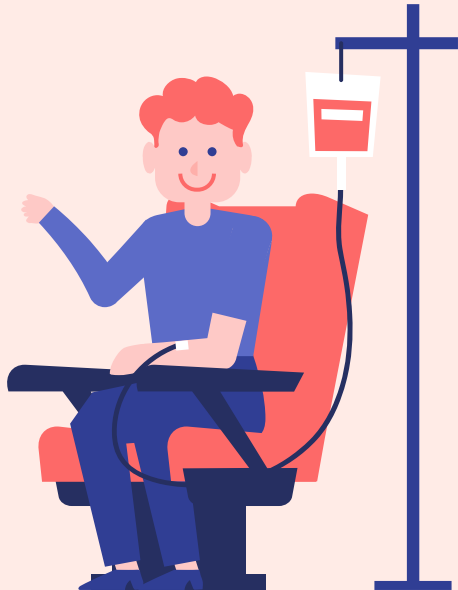
한계점 및 보완방안

데이터의 한계

대부분의 특성이 코드화 되어 있어
구체적인 모델 해석에 어려움이 있다.

모델의 한계

다중분류모델에서 정확도가 낮은 범주가 있다.
⇒ 이중분류모델로 전처리하여 장기입원을 예측할
수 있다.





Thanks!

감사합니다!