

PREDICTING ICU Mortality

A Deep Learning Approach

2023 Fall - Data Analytics

정보융합학부 김주혁 장원재





TABLE OF CONTENTS

01. INTRODUCTION

배경 및 프로젝트 주제

02. DATA PREPROCESSING

데이터 전처리 과정

03. PROJECT PIPELINE

프로젝트 과정

04. FORWARD PLAN

향후 일정

01.

INTRODUCTION

배경 및 프로젝트 주제 소개

COVID-19

코로나로 드러난 '참담한' 중환자의료 현실...학회, 각개전투 시작

청년의사 & 송수연 기자 | ⌚ 입력 2022.11.30 05:07 | 💬 댓글 0

코로나19 중환자병상 부족에 “무의미한 중환자 줄여야”

청년의사 & 김은영 기자 | ⌚ 입력 2021.12.08 13:23 | 💬 댓글 0

오미크론 변이 유행 시 한국만 초과 사망 증가...이유는?

청년의사 & 송수연 기자 | ⌚ 입력 2022.09.06 06:00 | ⌚ 수정 2022.09.06 07:05 | 💬 댓글 0

COVID-19

“코로나19 팬데믹이 ‘후진국 수준’인 우리나라 중환자 의료체계를 수면 위로 끌어올렸다. 하지만 그뿐이었다. 위중증 환자 급증으로 **중환자 병상 부족 문제**가 반복될 때마다 중환자 의료체계를 개선해야 한다는 목소리도 커졌지만 행정명령으로 병상만 확보하면 그만이었다.”

“임 회장은 ‘정부는 중환자 병상을 늘리기 위해 노력하고 있지만 이와 함께 **무의미한 중환자실 입실**도 줄일 필요가 있다’며 ‘중환자실을 일반병상 만들 듯 만들어낼 수는 없다. 또 만들어진다고 한들 의료인력을 쉽게 늘릴 수 있는 것도 아니다’라고 말했다. 임 회장은 ‘지난 2014년 상급종합병원 연구에 따르면 내과계 중환자실 환자의 10%가 입실 당일 이미 **무의미한 입원**인 것으로 조사됐다’면서 “**무의미한 중환자실 입원**을 줄이면 상당수 중환자 병실을 만들어내는 효과가 있을 것”이라고 했다.”

“김영삼 교수 연구결과, 올해 3월 초과사망 1만8000명 발생 폐렴 등 입원 환자 줄어 ‘비코로나 환자 의료접근성 떨어졌다’ ‘**필요 인원의 58% 인력으로 중환자실 운영, 사망률 높아**’”

→ 코로나19와 같은 전염병이 다시 재발한다면, 현재의 의료 체계로 감당하기에 어려운 부분이 많음

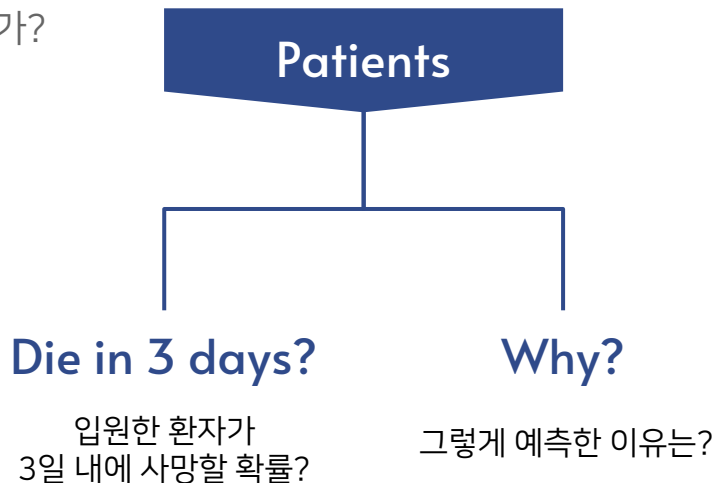
→ 중환자 의료체계 관련 지원이 필요하며, 병실, 인력 등 제한된 자원을 적절히 분배할 수 있도록 **환자의 중증도**를 정확히 추정해야 함

프로젝트 소개 - Predicting ICU Mortality

ICU : Intensive Care Unit, 중환자실

딥러닝 기반 중환자 중증도 추정

- 중환자실에 입원한 환자가 3일 내 사망할 확률은 어떻게 되는가?
- 위중한 환자에 대한 자원 집중
- 적정 간호인력 수요 산정
- 사망 영향 요인 식별



02.

DATA PREPROCESSING

활용 데이터 및 전처리 과정 소개

데이터 확보 과정

중환자실에 입원한 환자가 **3일 내 사망할 확률**은 어떻게 되는가?

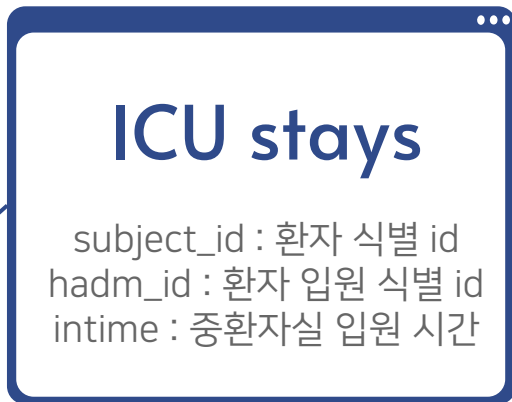
→ MIMIC-IV 데이터셋의 Chartevents, Inputevents, Outputevents, Labevents 데이터 사용

→ **입원 후 6시간 내에 발생한 이벤트**만을 고려

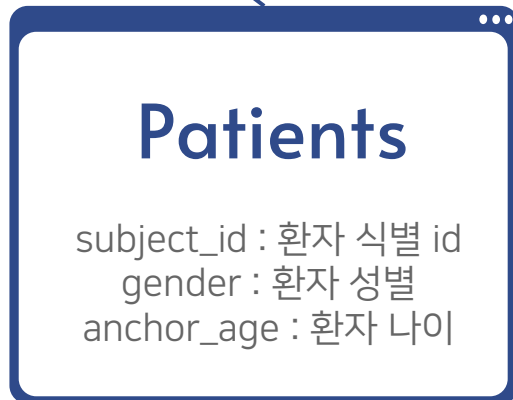
중환자실에 입원한 환자 한 명을 하나의 샘플(Row, Instance)로 생각하여 딥러닝 모델 학습

→ 샘플의 특성(Column, Feature)은 6시간 내 발생한 이벤트들의 측정치

TABLE MERGE



환자의 기본 인적 정보



중환자실 입원 환자가
사망까지 걸리는 시간 계산

3일 내에 사망 여부 확인

TABLE MERGE

storetime - intime 계산

입원 후 6시간 내 발생한
event만 사용

ICU stays

subject_id : 환자 식별 id
hadm_id : 환자 입원 식별 id
intime : 중환자실 입원 시간

Outputevents

hadm_id : 환자 입원 식별 id
storetime : 데이터 저장 시각
itemid : Event 식별 id
value : Event 측정 값

itemid 식별

d_items

item_id : Event 식별 id
label : Event 설명

TABLE MERGE

Outputevents

hadm_id : 환자 입원 식별 id
storetime : 데이터 저장 시각
itemid : Event 식별 id
value : Event 측정 값

6시간 이내의 데이터 추출

```
outputevents['intime'] = pd.to_datetime(outputevents['intime'])  
outputevents['storetime'] = pd.to_datetime(outputevents['storetime'])
```

```
outputevents['time_to_store'] = outputevents['storetime'] - outputevents['intime']  
outputevents['time_to_store'] = outputevents['time_to_store'].dt.total_seconds()
```

```
# 6시간 이내의 데이터
```

```
outputevents['time_to_store_in_day'] = (outputevents['time_to_store'] < 86400 / 4) & (outputevents['time_to_store'] > 0)
```

itemid 식별

```
outputevents_in_6hour = outputevents[outputevents.time_to_store_in_day  
outputevents_in_6hour = pd.merge(outputevents_in_6hour, d_items[['itemid', 'label']], on=['itemid'], how='left')
```

hadm_id로 group, sum 및 pivot

```
# 6시간 내 value의 합을 사용
```

```
tmp = outputevents_in_6hour.groupby(['hadm_id', 'label'])['value'].sum()  
tmp = pd.DataFrame(tmp).reset_index()
```

```
outputevents_in_6hour_pivot = tmp.pivot(index='hadm_id', columns='label', values='value')  
outputevents_in_6hour_pivot = outputevents_in_6hour_pivot.reset_index()  
outputevents_in_6hour_pivot.to_csv('outputevents_in_row.csv')
```

TABLE MERGE

Pivot 까지 진행 결과

```
outputevent = pd.read_csv('./outputevents_in_row.csv', index_col=0)
outputevent
```

✓ 0.1s

Python

	hadm_id	Anderson (gastric)	Blakemore	Cath Lab	Cerebral Subdural #1	Cerebral Subdural #2	Cerebral Ventricular #1	Cerebral Ventricular #2	Chest Tube #1	Chest Tube #2	...	Sump #1	Sump #2	Suprapubic	T Tube	TF Residual	TF Residual Output	Tap	Void	Wound Vac #1	Wound Vac #2
0	20000094	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	20000808	NaN	NaN	NaN	NaN	NaN	146.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	20001305	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	20001361	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	20001395	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	400.0	NaN	NaN
...
59742	29999186	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
59743	29999444	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
59744	29999498	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	10.0	NaN	NaN
59745	29999625	NaN	NaN	NaN	NaN	NaN	21.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
59746	29999828	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

59747 rows × 71 columns

TABLE MERGE

storetime - intime 계산

입원 후 6시간 내 발생한
event만 사용

ICU stays

subject_id : 환자 식별 id
hadm_id : 환자 입원 식별 id
intime : 중환자실 입원 시간

Inpatevents

hadm_id : 환자 입원 식별 id
storetime : 데이터 저장 시각
itemid : Event 식별 id
amount : Event 측정 값
patientweight : 환자 체중

itemid 식별

d_items

item_id : Event 식별 id
label : Event 설명

TABLE MERGE

storetime - intime 계산

입원 후 6시간 내 발생한
event만 사용

ICU stays

subject_id : 환자 식별 id
hadm_id : 환자 입원 식별 id
intime : 중환자실 입원 시간

Chartevents

hadm_id : 환자 입원 식별 id
storetime : 데이터 저장 시각
itemid : Event 식별 id
valuenum : Event 측정 값

itemid 식별

d_items

item_id : Event 식별 id
label : Event 설명

TABLE MERGE

storetime - intime 계산

입원 후 6시간 내 발생한
event만 사용

ICU stays

subject_id : 환자 식별 id
hadm_id : 환자 입원 식별 id
intime : 중환자실 입원 시간

Labevents

hadm_id : 환자 입원 식별 id
storetime : 데이터 저장 시각
itemid : Event 식별 id
valuenum : Event 측정 값

itemid 식별

d_labitems

item_id : Lab Event 식별 id
label : Lab Event 설명

```
data = pd.read_csv('./DA_data.csv', index_col=0)
data
```

✓ 22.0s

Python

	hadm_id	intime	gender	anchor_age	mortality_in_second	mortality_in_3days	Anderson (gastric)	Blakemore	Cath Lab	...	WBC_y	WBCApacheIIValue	WBCScore_ApacheIV	WBC_ApacheIV	Warming Device	Warming Device Status	WbcApacheIIIScore	Yawning	Yawning (COWS)
0	24528534	2154-03-03 04:11:00	M	25	NaN	False	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	28960964	2150-06-19 17:57:00	M	42	NaN	False	NaN	NaN	NaN	...	13.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	27385897	2138-02-05 18:54:00	M	70	835560.0	False	NaN	NaN	NaN	...	17.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	23483021	2123-10-25 10:35:00	M	87	NaN	False	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	20817525	2200-07-12 00:33:00	M	72	NaN	False	NaN	NaN	NaN	...	7.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
69180	21944963	2152-08-01 17:53:56	F	75	NaN	False	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
69181	27299174	2126-06-13 01:00:00	F	47	NaN	False	NaN	NaN	NaN	...	8.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
69182	28911582	2177-11-08 14:09:00	M	60	NaN	False	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
69183	22562812	2182-08-15 09:37:33	M	72	NaN	False	NaN	NaN	NaN	...	8.0	NaN	NaN	NaN	0.0	0.0	NaN	NaN	NaN
69184	22695803	2115-12-01 00:37:00	M	55	NaN	False	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

69185 rows × 2921 columns

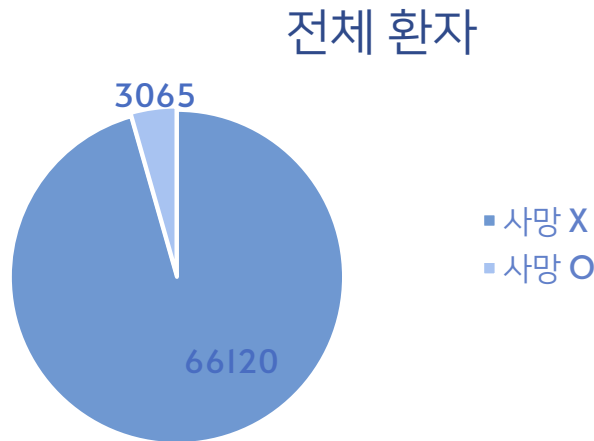
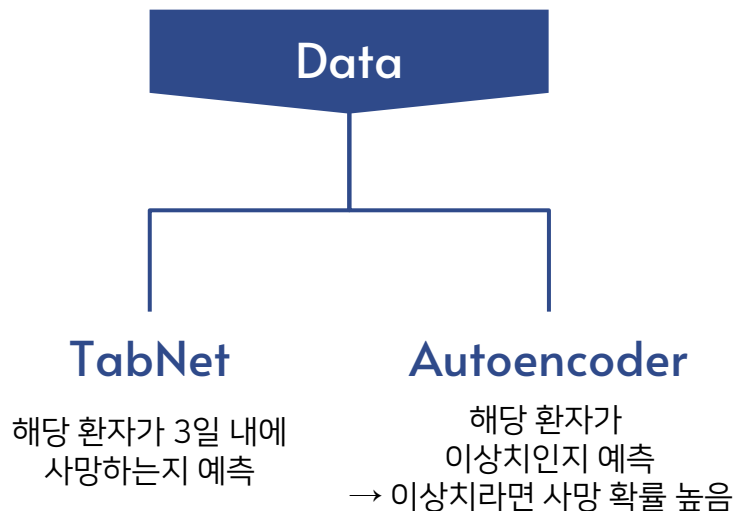
03.

PROJECT PIPELINE

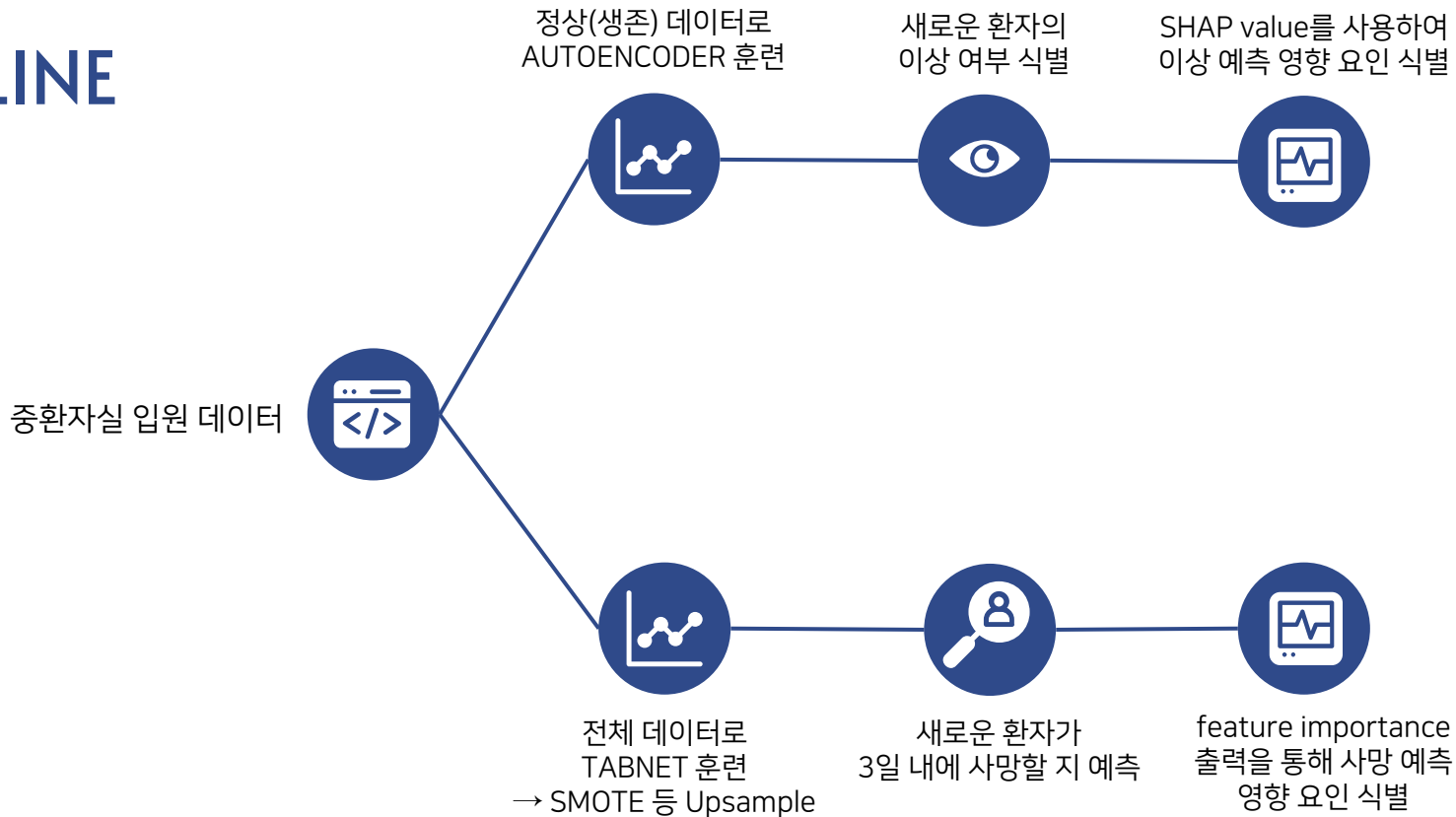
프로젝트 과정 소개

Deep Learning Approach

TabNet을 사용한 이진 분류 문제 vs. Autoencoder를 사용한 이상 탐지 문제



PIPELINE

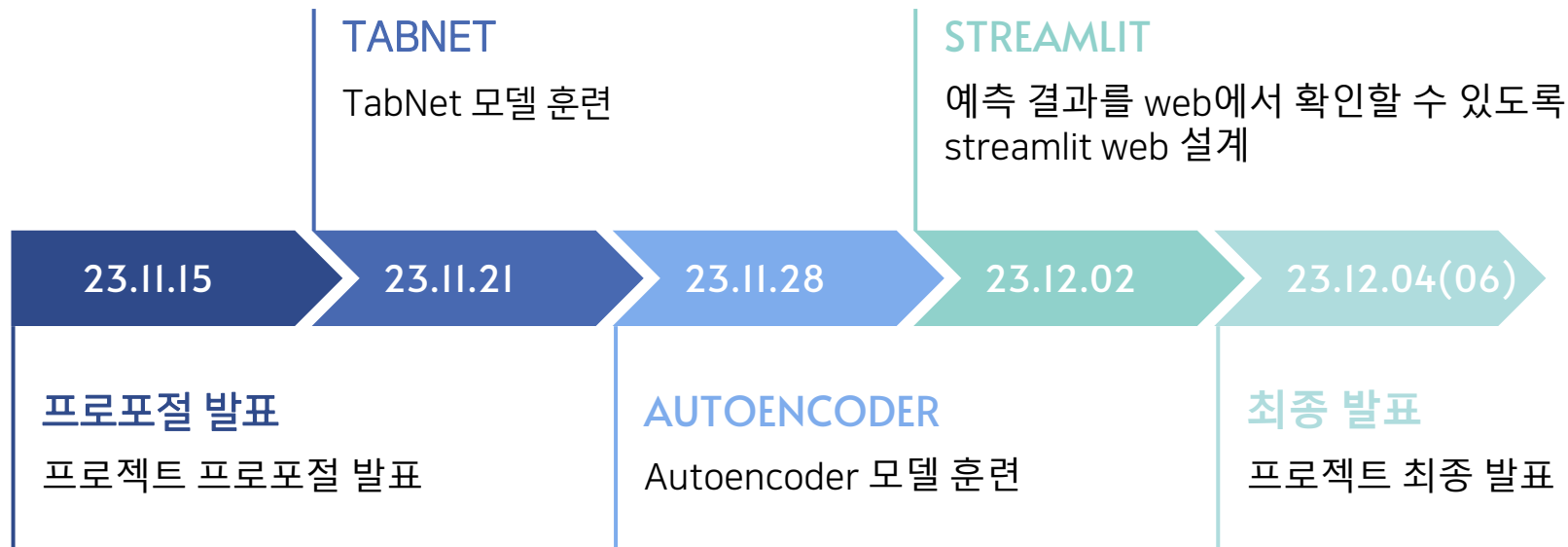


04.

FORWARD PLAN

향후 일정 소개

FORWARD PLANS





THANKS