

# 표 형식 데이터에 대한 딥러닝을 활용한 암 발병 예측

**소속** 정보컴퓨터공학부

분과 A

팀명 **esc**

### 참여학생

최지광, 송민재, 서진욱

지도교수

송길태

## 과제 개요

## 과제 목표

딥러닝은 이미지, 텍스트와 같은 데이터에서 높은 성능을 보이지만 표 형식 데이터에 대해선 머신러닝보다 낮은 성능을 보인다. 이 문제를 해결하기 위해 표 형식 데이터에 대한 딥러닝 모델의 성능을 향상시키고 표 형식 데이터에 대한 딥러닝 결과를 **Explainable AI**를 통해 해석하는 것이 목표이다.

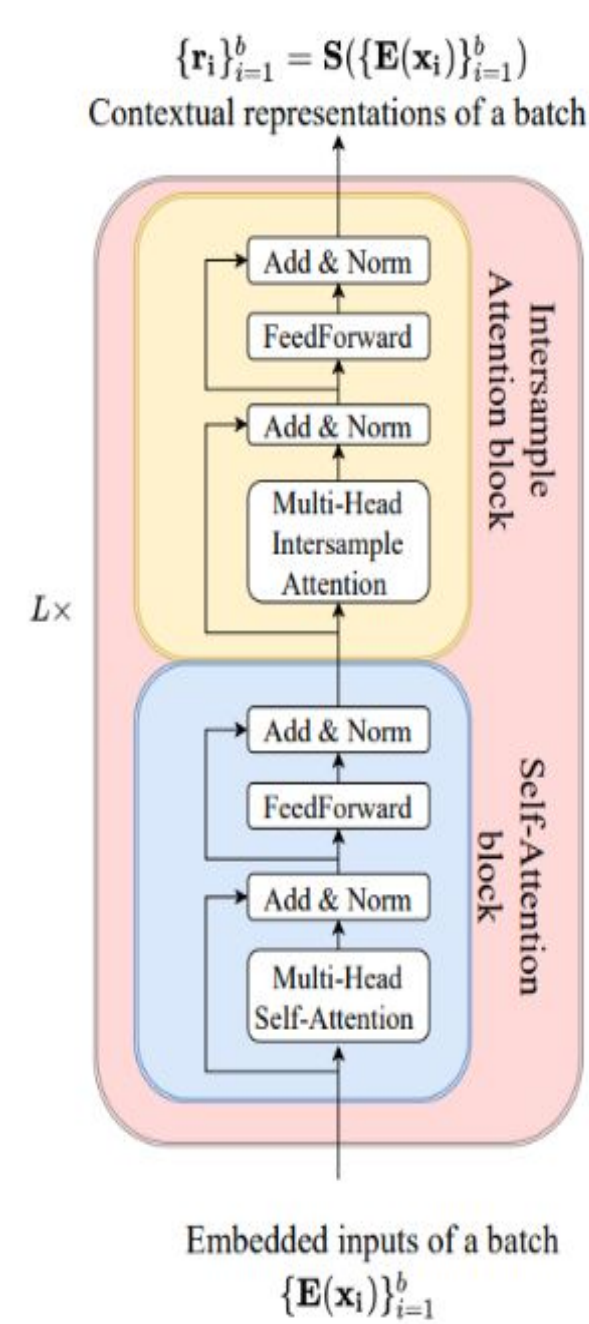
## 과제 내용

- 여러 모델을 비교해 성능이 가장 좋은 딥러닝 모델 선정
- 암의 발병에 있어 어떤 **feature**가 영향을 많이 주는지를 비교할 수 있는 SHAP 모델 구현
- **feature**에 대한 개인의 입력값을 받았을 때 암의 발병 확률과 예방방안을 알려주는 웹서비스 개발

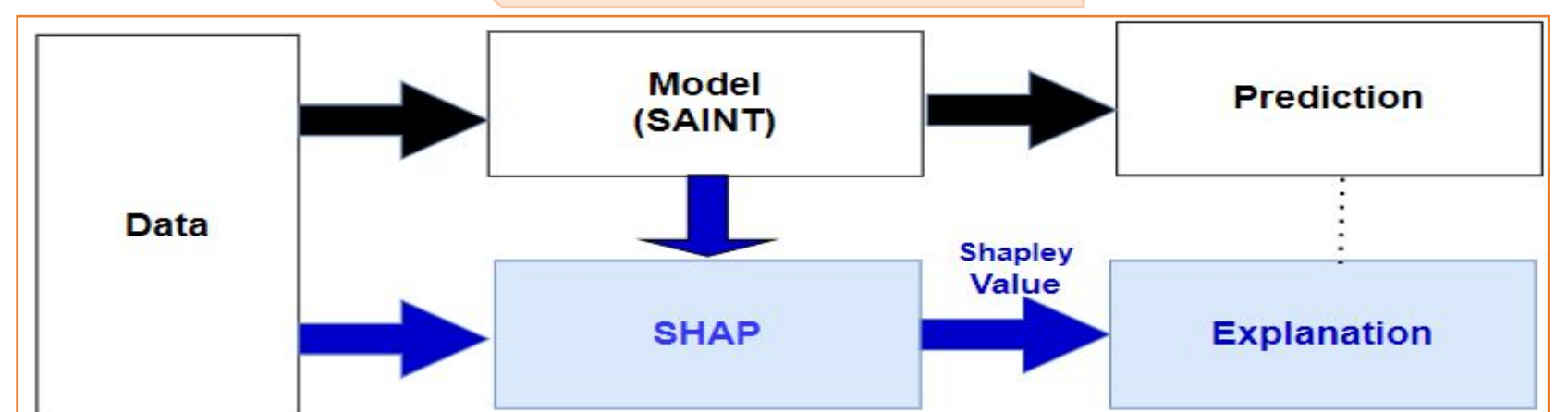
## 과제 구성

## 모델 비교 및 모델 선정

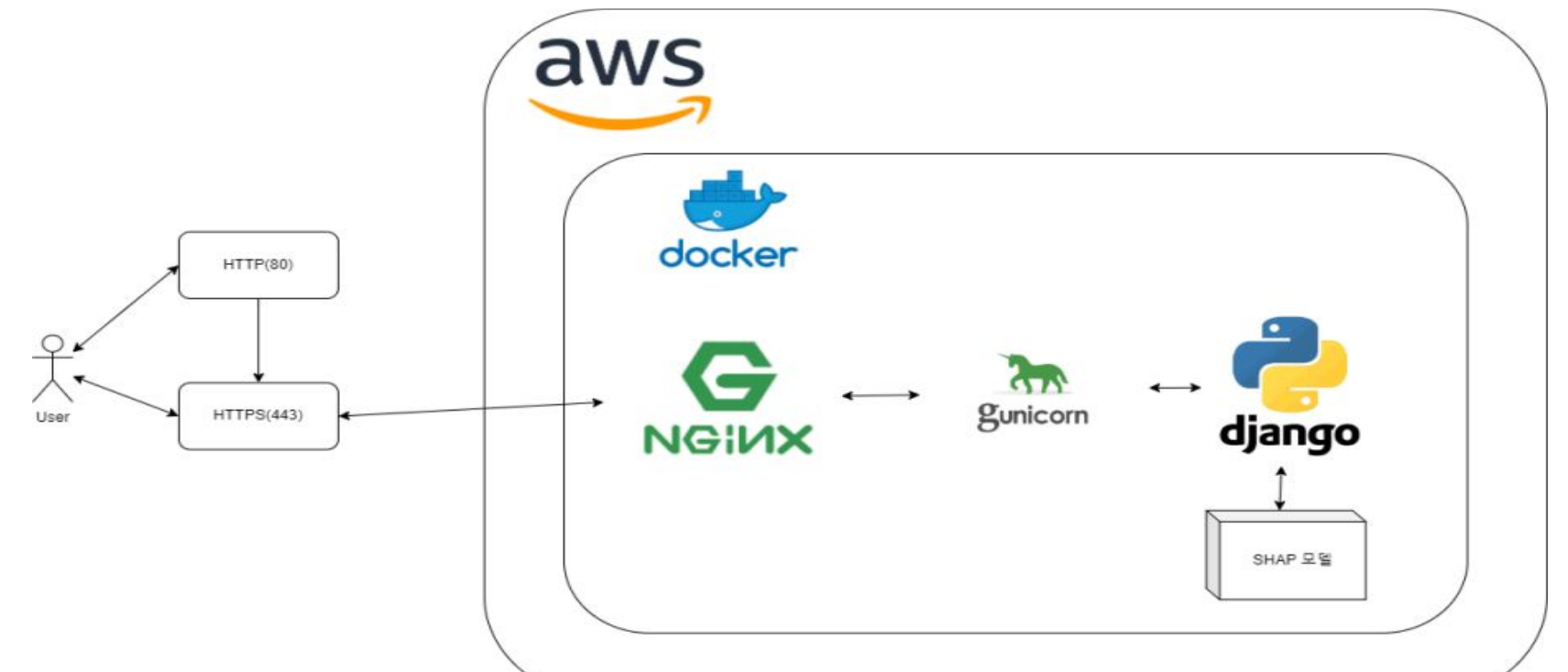
	f1 score	recall	AUC
Random Forest	0.9047	0.8995	0.8852
LightGBM	0.8860	0.8834	0.8834
MLP	0.8559	0.8655	0.8693
TabNet	0.9003	0.8976	0.8859
FT-Transformer	0.8862	0.8866	0.8843
GRANDE	0.9115	0.8879	0.9077
SAINT	0.9067	<b>0.9002</b>	0.8899

**SAINT**

## SHAP 구조도

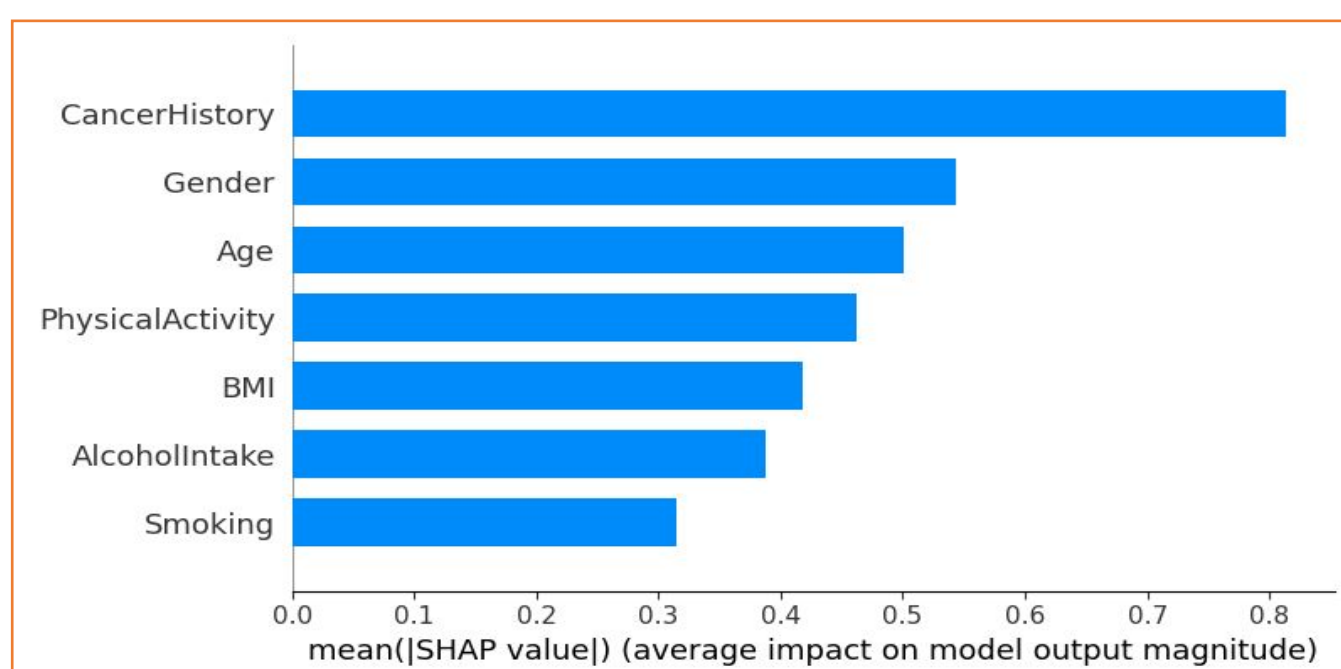


## 서비스 구조도

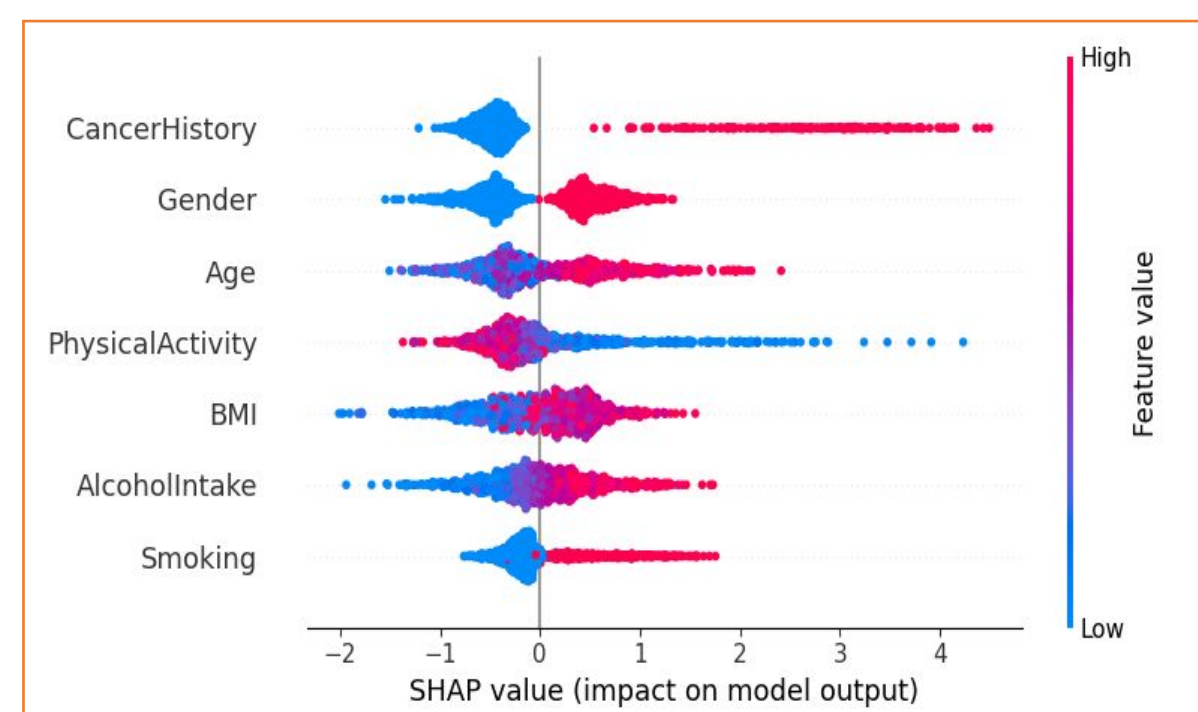


## 시각

## Bar 그래프



## 점 그래프



X축은 SHAP값으로 양수이면 feature가 예측을 더 높이는 방향으로 기여했다는 것을 뜻한다. 점의 색은 해당 feature의 값 크기로 빨간색은 feature값이 높을 때 파란색은 feature 값이 낮을 때를 말한다.

## 암 발병 예측 웹 서비스

