

---

# XAI기반 심장질환 예측

(XAI-Based Heart Disease Prediction)

---

4조	[조정우 임태경 이민서 이지훈]
교과목명	오픈소스SW프로젝트
교수	정유선 교수님
제출일	2023년 1월 19일 목요일

## 목 차

---

초록

---

키워드

---

Abstract

---

Keywords

---

### I 서론

---

1.배경 소개

---

2.기존 시스템 한계점

---

3.연구 목표

---

4.XAI 설명

---

### II 본론

---

1. 데이터 준비

---

2. 머신러닝 모델 구조

---

3. SHAP

---

### III 결과

---

### IV 결론

---

<참고문헌>

---

---

# XAI기반 심장질환 예측

## (XAI-Based Heart Disease Prediction)

4조

조정우 임태경 이민서 이지훈

---

### 초 록

이 보고서에서, 저희는 심장병 데이터셋을 사용하여 심장 질환을 예측하기 위한 새로운 XAI 기반 접근법을 제안합니다. 이 모델은 나이, 성별, 가슴통증유형, 휴식 혈압, 콜레스테롤, 공복 상태의 혈당, 안정된 상태의 심전도 결과, 달성한 최대 심박수, 운동으로 발생한 협심증, 운동으로 발생하는 ST분절 저하, ST분절 기울기를 포함한 환자 정보의 데이터 세트에 대해 훈련합니다. 저희는 특징 선택과 해석 가능한 머신러닝 기술을 결합하여 심장 질환의 가장 중요한 예측 변수를 식별합니다. 저희의 결과는 XAI 기반 모델이 정확성과 해석 가능성 측면에서 기존의 기계 학습 방법을 능가한다는 것을 보여줍니다. 또한 SHAP기법이 모델의 예측을 시각화하고 이해할 수 있도록 제시합니다. 이러한 접근 방식은 심장 질환의 조기 진단과 관리를 개선하는 동시에 의사 결정 과정에 대한 신뢰를 높일 수 있는 잠재력을 가지고 있습니다.

### Abstract

In this report, we propose a novel XAI-based approach for predicting heart disease using heart disease datasets. This model trains a dataset of patient information including age, gender, type of chest pain, resting blood pressure, cholesterol, fasting blood sugar, stable ECG results, maximum heart rate achieved, angina caused by exercise, decreased ST segmentation caused by exercise, and ST segment slope. We combine feature selection with interpretable machine learning techniques to identify the most important predictors of heart disease. Our results show that XAI-based models outperform conventional machine learning methods in terms of accuracy and interpretability. We also present SHAP techniques to visualize and understand the predictions of the model. This approach has the potential to improve early diagnosis and management of heart disease while increasing confidence in the decision-making process.

**키워드 :** 설명 가능한 인공지능, 기계 학습, 예측 모델링, 심장병 예측

XAI (Explainable Artificial Intelligence), Machine learning, Predictive modeling,  
Heart disease prediction

# I 서론

## 1. 배경

심장병은 전 세계적으로 사망의 주요 원인입니다.

심장과 혈관에 영향을 미치는 조건으로는 관상동맥질환, 심부전, 부정맥 등이 있습니다.

심장병의 조기 발견과 진단은 합병증과 사망의 위험을 줄이는데 큰 도움이 있으므로, 조기 진단의 어려움을 해결하고자, 또한 진단과 치료의 난이도 상승과 오진율 증가, 의료비용 증가에 따라 의사또는 전문의를 대체, 또는 보조하는 의료 인공지능 기술의 수요가 증가하고 있습니다.

그러기 때문에 저희는 정확하며, 기존 인공지능의 단점을 해결한 해석 가능한 모델이 필요하다 느껴졌습니다.

## 2. 기존 시스템 한계점

기존 ai 머신러닝에는 "블랙박스" 라는 AI 시스템의 주요 문제 중 하나인

"그들이 어떻게 결정을 내리는지 이해하기 어려울 수 있다는 것"이 있습니다.

이러한 시스템의 내부 작동은 쉽게 접근할 수 없기 때문에, 주어진 결정이 내려진 이유를 결정하거나 의사 결정 과정에서 오류를 식별하고 수정하는 것이 어려울 수 있습니다.

또한 이러한 블랙박스 특성상 해석 가능성이 부족하기 때문에, 이해관계자나 관련 종사자에게 모델이 안전하게 작동하고 있음을 확신시키기 어려움이 있었습니다.

전반적으로 AI는 여전히 개발 및 연구 분야이며 새로운 돌파구는 이러한 한계를 극복하는데 도움이 될 것 같습니다.

## 3. 연구 목표

기존 머신러닝 알고리즘 모델은 어떻게 특정 결정이나 예측에 도달하였는지 이해하기 어려우며, 모델의 출력을 신뢰하기 어려웠습니다.

그러기 때문에, 저희는 안전한 AI모델 개발이 필요했습니다.

본 연구의 목적은 심장질환의 존재를 정확하게 예측할 수 있는 효율적이고 확장 가능한 그레이디언트 부스팅 라이브러리인 XGBoost를 이용한 모델을 이용하여 심장질환 예측한 출력을 XAI를 이용하여 모델의 출력이 해석이 가능하도록 신뢰할수 있게 만드는것이 목표입니다.

## 4. XAI란

XAI는 머신러닝 모델을 인간에게 보다 해석 가능하고 투명하게 만드는 데 초점을 맞춘 AI 연구 분야입니다.

XAI분야의 최종 목표는 인간이 이해하기 쉬운 방식으로 예측과 의사결정 과정을 설명할 수 있는 모델을 개발하는 것 입니다.

XAI를 개발한 이유는 머신러닝 모델의 신뢰도와 책임성을 높이기 위함입니다.

머신 러닝 모델은 종종 의료 및 금융과 같은 중요한 의사 결정 시나리오에서 사용되며, 여기서 모델이 어떻게 예측하는지 이해하는 것이 중요합니다. XAI 기법은 이러한 모델의 투명성과 해석 가능성을 높여 의사 결정 과정을 더욱 쉽게 이해하고 잠재적인 편향이나 오류를 식별하는데 도움이 됩니다.

쉽게 말해, XAI는 인간이 AI를 이해할 수 있고 명확하게 만드는 방법입니다.

## II 본론

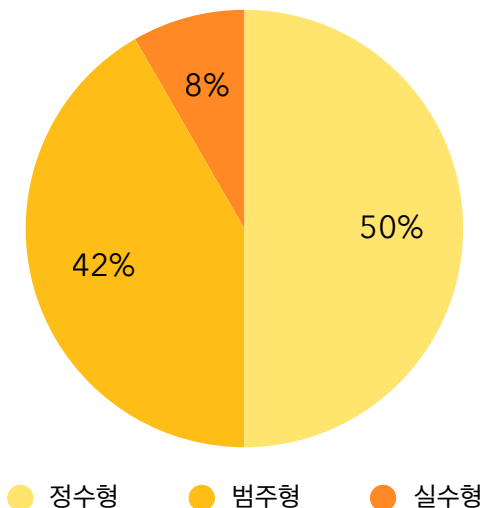
### 1. 데이터 준비

저희가 가져온 데이터 세트는 Your Machine Learning and Data Science Community이며 세부 정보는 12개의 속성을 갖고 있으며 데이터 개수는 918개 입니다. 속성 특성은 정수형, 실수형, 범주형입니다.

저희가 가져온 데이터 세트 타입은

Age : 나이 ,  
Sex : 성별 (남자,여자),  
ChestPainType : 가슴통증유형,  
RestingBP : 휴식 혈압,  
Cholesterol : 콜레스테롤,  
FastingBS : 공복 상태의 혈당,  
RestingECG : 안정된 상태의 심전도 결과 ,  
MaxHR : 달성한 최대 심박수,  
ExerciseAngina : 운동으로 발생한 협심증,  
Oldpeak : 운동으로 발생하는 ST분절 저하,  
ST\_Slope : ST분절 기울기  
로 구성이 되어있습니다.

데이터 세트 속성 특성 비율



### 2. 모델 구조

#### 전처리

가져온 데이터 세트를 모델링 하기전에 전처리 과정을 거쳐주어야 합니다.

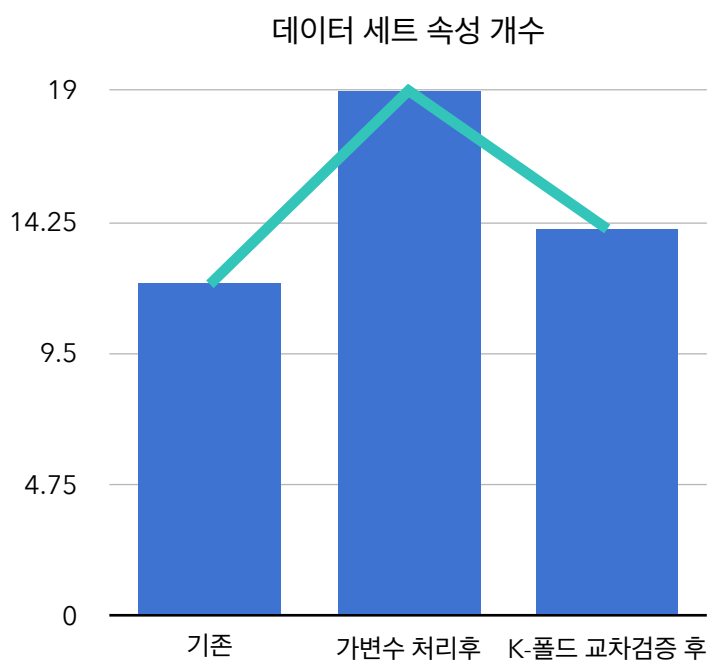
머신 러닝 전처리는 머신 러닝 모델에서 사용하기 위해 데이터를 준비하는 과정입니다. 전처리의 목표는 모델이 데이터를 최대한 사용할 수 있도록 하여 효과적으로 학습하고 정확한 예측을 할 수 있도록 하는 것입니다. 머신 러닝 모델의 성능은 입력 데이터의 품질과 형식에 크게 좌우되기 때문에 이는 중요합니다. 전처리는 데이터가 모델이 효과적으로 이해하고 사용할 수 있는 형태로 유지되도록 하는 데 도움이 됩니다.

#### 가변수 처리

저희는 먼저 데이터 가변수 처리를 해주었습니다. 가변수(Dummy variable) 처리는 범주형 변수를 0과 1의 값을 가지는 변수로 변환해 주는 것을 뜻합니다. 가변수를 만드는 이유는, 범주형 변수는 사용할 수 없고 연속형 변수만 사용 가능한 분석 기법을 사용하기 위함입니다.

원래 12개의 속성을 가지고 있었으나 가변수 처리를 통해 속성이 늘어나 19개가 되었습니다. 여기서 저희는 목표값을 항상 시키기 위해 변수 간의 상관 관계를 K 폴드 교차검증을 사용해 영향력이 없는 속성을 확인후 지웠습니다.

K-폴드 교차 검증은 머신 러닝에서 널리 사용되는 기술로, 모델의 하이퍼 파라미터를 선택하고 모델을 조정하고 모델의 성능을 평가하는데 널리 사용됩니다.



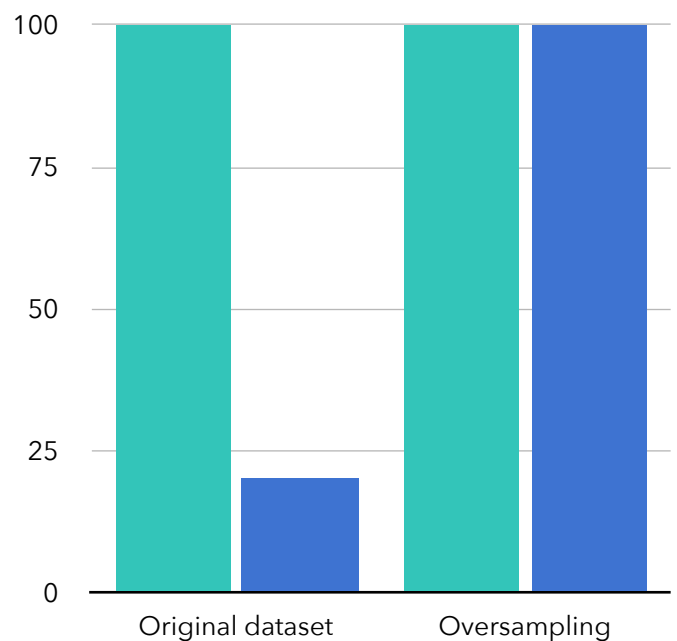
## 오버 샘플링

오버샘플링은 데이터 세트에서 클래스의 분포의 균형을 맞추기 위해 기계 학습에 사용되는 기술입니다. 데이터 세트의 한 클래스가 다른 클래스보다 샘플 수가 현저히 적을 때 사용됩니다. 이는 기계 학습 모델에 문제를 일으킬 수 있는데, 이는 소수 클래스의 예제가 충분하지 않아 다수 클래스에 대한 편향을 초래할 수 있기 때문입니다.

오버샘플링은 데이터 세트에서 소수 클래스의 대표성을 높이기 위해 기존 샘플을 복제하는 것을 포함합니다. 이는 모델이 소수 계층으로부터 더 효과적으로 학습할 수 있도록 도와서 편향을 줄이고 모델의 전반적인 성능을 향상시킬 수 있습니다.

또한, 오버샘플링은 의료 진단과 같은 영역에서 일반적인 문제인 불균형 데이터 세트를 처리하는 데에도 유용할 수 있습니다.

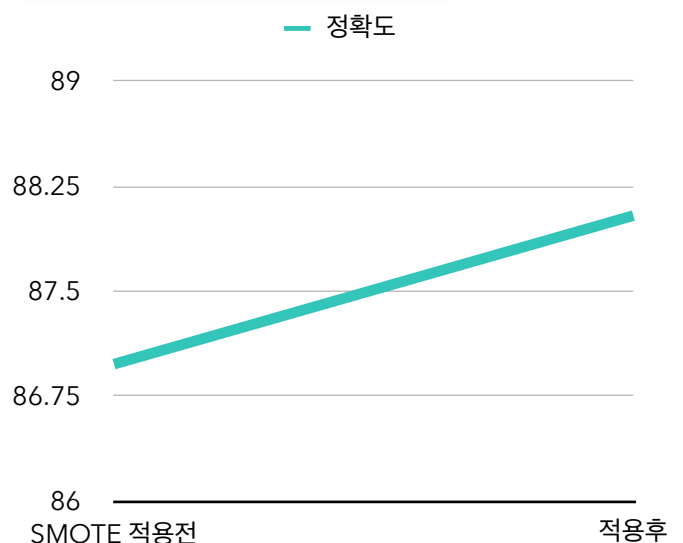
오버샘플링은 불균형 데이터 세트를 처리하는 간단하고 효과적인 기술이며, 모델의 성능을 향상시킬 수 있으며, 모델을 보이지 않는 데이터로 일반화하는 데 도움이 될 수 있습니다.



오버 샘플링 기법을 이용하여 저희는 학습용 검증 데이터 세트 개수 68개를 증가 시켜 데이터 불균형을 해결하였으며, 또한 SMOTE를 사용하여 정확도를 향상하였습니다.

## SMOTE

SMOTE의 기본 아이디어는 소수 클래스 샘플과 k-가장 가까운 소수 클래스 이웃을 선택한 다음 선택한 샘플과 이웃을 연결하는 선을 따라 합성 샘플을 생성하는 것입니다.



## 모델링

다음은 저희가 머신러닝에 사용된 모델을 XGBoost 라이브러리에 속하는 XGB Classifier를 사용했습니다

XGBoost는 그레이디언트 부스팅을 과적합 방지를 위해 예측을 반복적으로 개선하는 의사 결정 트리 기반 모델입니다.

저희가 사용한 이유에는 많은것이있지만 그중 몇가지를 말씀 드리자면,

1. 속도와 성능: XGBoost는 속도와 성능으로 알려져 있어 빠른 실행이 필요한 대규모 데이터 및 작업에 적합합니다.
2. 누락된 데이터 처리: XGBoost는 누락된 데이터를 처리하고 누락된 값을 귀속시킬 필요가 없이 모델을 구성할 수 있습니다.
3. 정규화 : XGBoost에는 과적합을 방지하기 위한 정규화 기능이 내장되어 있어 모델의 일반화를 개선할 수 있습니다.
4. 병렬 및 분산 컴퓨팅 지원 : XGBoost에는 병렬 및 분산 컴퓨팅 지원이 내장되어 있어 대규모 모델을 신속하게 교육하는 데 사용할 수 있습니다.

XGB Classifier는 XGBoost 라이브러리에 대한 인터페이스를 제공하는 Scikit-learn 라이브러리의 클래스 입니다.

친숙한 스킵킷 학습 프레임워크에서 XGBoost가 제공하는 강력한 그레이디언트 부스팅 알고리즘을 사용할 수 있습니다.

XGB Classifier는 특정 문제에 맞게 모델을 미세 조정할 수 있는 쉬운 방법을 제공합니다.

저희들은 목표 달성을 위해 XGBoost 모델을 훈련시키기 위해 Age : 나이 ,Sex : 성별 (남자,여자), ChestPainType\_ASY : 무증상, ChestPainType\_ATA : 비전형적 가슴통증, ChestPainType\_NAP : 비심인적 가슴통증 , Cholesterol : 콜레스테롤, FastingBS : 공복 상태의 혈당, MaxHR : 달성한 최대 심박수, ExerciseAngina : 운동으로 발생한 협심증, Oldpeak : 운동으로 발생하는 ST분절 저하, ST\_Slope\_Up : ST분절 기울기 상승, ST\_Slope\_Flat : ST분절 기울기 평평, ST\_Slope\_down : ST분절 기울기 하강, HeartDisease : 심장질환 판단을 포함한 환자 정보의 데이터 세트를 사용했습니다.

XAI사용을 위해 저희는 예측에 대한 기능의 기여를 나타내는 SHAP을 사용하여 머신 러닝 모델의 출력을 설명하기 위해 통합된 접근법인 SHAP(SHAPley Additive exPlanations)을 사용했습니다.

## 3. SHAP

SHAP는 모델 해석 가능성을 위한 중요한 기술이며 모델의 특징 상호 작용과 편향을 식별하는 데 도움이 될 수 있으며, 이는 모델의 성능을 개선하거나 사람들에게 설명이 가능하도록 사용 될 수 있습니다.

SHAP는 머신러닝 모델의 출력을 설명하는 데 사용되는 기술입니다. 이 기술은 협동 게임 이론에서 나오는데요, 여기서 게임이론이란 우리가 아는 그런 게임을 말하는것이 아닙니다. 여러 주제가 서로 영향을 미치는 상황에서 서로가 어떤 의사결정이나 행동을 하는지에 대해 이론화한 것을 말합니다. 게임이론에서 나온 SHAPLY VALUE의 개념을 기반으로 합니다. SHAP 값은 모델 출력의 평균 차이로 해석이 가능합니다.

값이 양수이면 긍정적으로 기여하고, 음수이면 출력에 부정적으로 기여하고 있음을 나타내고 있습니다.

간단히 말해서, XAI(설명 가능한 AI)는 SHAP 값을 사용하여 모델이 어떻게 예측을 했는지 명확하고 해석 가능한 설명을 제공합니다. 최종 출력에 대한 각 피처의 기여도를 계산한 다음 기여도를 기준으로 각 피처에 점수를 할당합니다. 점수가 가장 높은 특징은 예측에 영향을 미치는 가장 중요한 요인으로 간주됩니다.

즉, SHAP 값은 각 피처가 최종 예측에 얼마나 기여하는지, 그리고 이러한 피처 값이 최종 출력을 생성하기 위해 서로 어떻게 상호 작용하는지, 그리고 인공지능 모델의 의사결정 과정을 이해하는 강력한 방법이며, 비전문가에게 모델이 내린 예측을 설명하는 방법입니다.

## III 결과

### 모델 성능 평가 결과

저희가 구축한 모델의 성능을 평가하기 위해 SMOTE 방법 적용 전 데이터세트, 적용 후 데이터세트를 K-Fold 교차검증을 수행하였습니다. K-Fold교차검증은 5개의 폴드로 나누었으며 데이터의 신뢰성을 얻기 위해 랜덤한 값으로 분리하여 검증을 실시하였습니다. 검증결과로 SMOTE를 적용하기 전 데이터세트의 검증정확도로 91%, 81%, 84%, 86%, 84%의 정확도를 얻었으며 85%의 평균 검증 정확도를 얻었습니다. SMOTE를 적용한 데이터세트의 검증정확도로 89%, 88%, 87%, 85%, 86%의 검증정확도를 얻었으며, 평균 검증 정확도는 87%로 SMOTE 적용하기 전인 85%에 비해 2% 향상한 결과를 도출하였습니다.

최종 결과로 85%의 성능이 87%로 2% 증가함을 확인하였으며, SMOTE를 적용한 모델 좋은 성능을 나타냄을 확인하였습니다.

그림1-1 121번째 학습데이터의 샐플리값 1차 평면도  
**SHAP 결과**

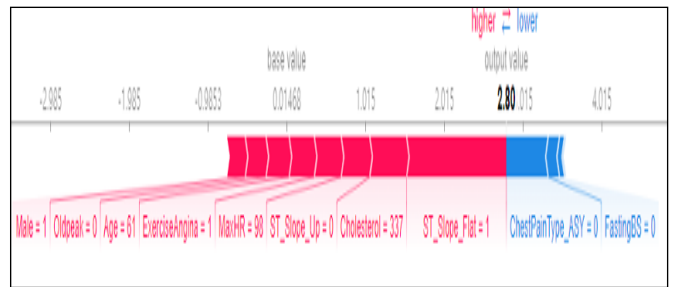


그림1-1은 Shap방법의 Shap.force\_plot 함수를 사용하여 특정 데이터에 대한 샐플리 값을 1차원 평면으로 나타낸 것입니다. 그래프를 해석하면 빨간색과 파란색의 영역으로 나뉘는 것을 확인할 수 있습니다. 모델 예측의 긍정적인 영향을 준 특징은 빨간색으로 구현되며 파란 부분은 모델 예측의 부정적인 영향을 나타냅니다. 이 데이터의 샐플리 값은 2.8로 양수를 띄며, 파란부분에 비해 빨간 부분의 영역이 비교적 넓은 영역을 차지해 심장질환을 가진 환자라는 것을 짐작할 수 있습니다. 특징의 기여도에 따라 영역별로 크기가 다르며 모델 예측에 가장 크게 기여한 요인은 ST\_Slope\_Flat로 이 환자의 심전도 중 ST분절의 기울기는 평평하다는 뜻을 나타냅니다. 두번째로 콜레스테롤이 모델 예측에 부정적인 영향을 나타낸다는 것을 위 그림을 통해 알 수 있습니다. 최종적으로 특정 데이터에 대해 어떤 요인이 긍정,부정적으로 작용하였는지 해석이 가능합니다.

그림 1-2 학습용 데이터 ST\_Slope\_Flat와 HeartDisease 관계

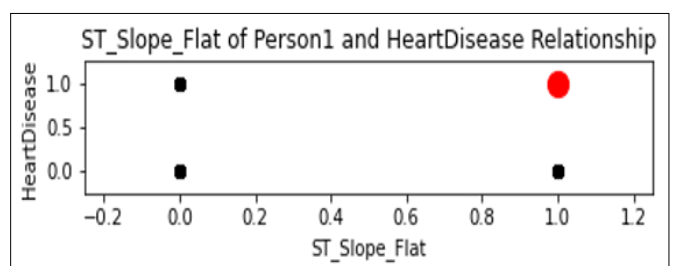




그림1-2은 특정 데이터에 가장 기여도가 높은 ST\_Slope\_Flat와 HeartDisease의 관계, 원본 데이터의 ST\_Slope\_Flat와 HeartDisease의 관계를 결합한 그림 입니다. HeartDisease의 특징으로 1값은 심장 질환을 가지고 있으며, 0값은 심장질환을 가지고 있지 않음을 나타냅니다. 그림을 보면 두 특징 모두 참(1)값을 나타내고있습니다. 이로 인해 특정데이터의 환자를 심장병이라고 예측한 모델의 결과가 맞음을 알 수 있습니다.

그림 2-1 790번째 학습데이터의 샐플리값 1차 평면도

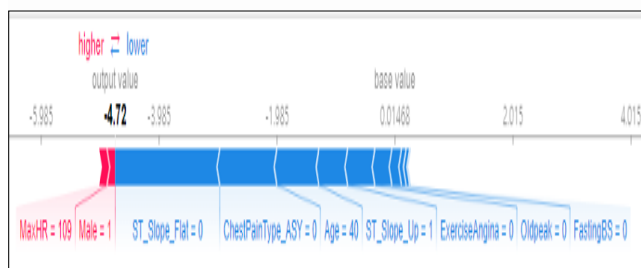


그림 2-1은 전에 소개했던 그림과 정 반대인 심장질환이 없는, 정상인 환자 데이터의 샐플리값 입니다. 샐플리값은 4.72로 양수형태를 띄고 있으며 정상의 예측에 대해 가장 크게 기여한 요인은 전과 동일한 ST\_Slope\_Flat 요인입니다. 위 환자는 ST분절의 기울기가 평평하지 않으며 평평했던 환자는 심장병에 걸릴 경향이 있다는 것을 짐작할 수 있습니다. 두번째로 가장 크게 작용한 요인은 ChestPainType\_ASY(무증상의 가슴통증)이며 값은 0을 나타내어 다른 유형의 가슴통증을 앓고있지만 심장질환은 아닌 것을 알 수 있습니다.

그림 2-2 학습용 데이터 ST\_Slope\_Flat와 HeartDisease 관계

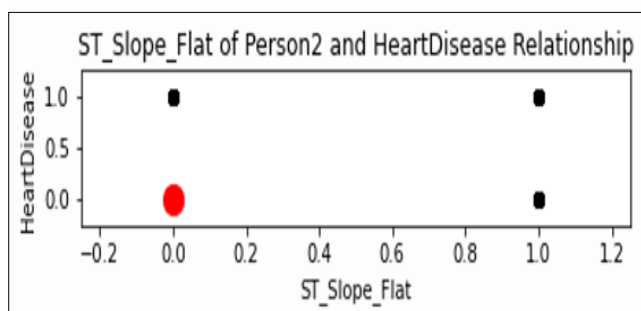


그림 2-2는 그림1-2와 똑같은 방식이며 원본 데이터와 특정데이터에 대한 ST\_Slope\_Flat와 HeartDisease관계를 연관지었을때 두 특징 모두 0값을 나타냈고 위의 환자는 심장질환을 가지고 있지 않은 것을 확인할 수 있습니다.

그림 3 전체 데이터셋 개별 누적 샐플리 값

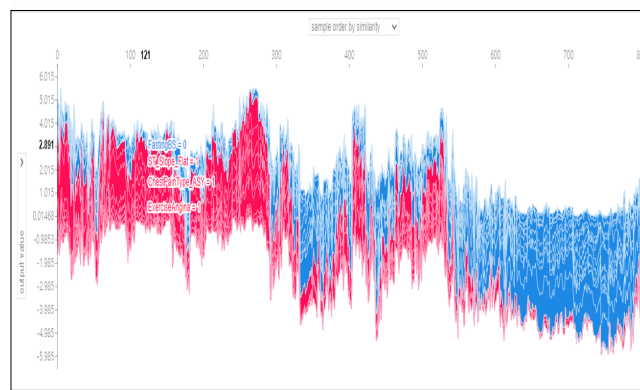


그림3은 SHAP의 shap.summary\_plot함수로 데이터셋의 개별 샐플리 값을 90도로 회전하여 전체데이터 수 만큼 누적된 그림 입니다. 즉, 그림1-1, 2-1의 그림을 90도로 회전하는 방식을 모든 데이터셋에 적용 한 것입니다. SHAP 라이브러리는 사용자의 움직임에 반응하기 위해 자바스크립트 툴킷을 사용합니다. 그래프에 마우스를 올리면 데이터 순서에대한 기여도를 볼 수 있습니다. 위 그림은 121번째의 데이터에 마우스를 올려둔 상태입니다. 복합적인 데이터 뿐만아니라 특징 하나에 대한 전체 누적 샐플리 영향도가 궁금하면 x축 상단의 드롭다운 메뉴를 열어서 특정 요인을 필터링하고 결과를 검토할 수 있습니다.

그림 3-1 전체 데이터 샐플리 값 Age-HeartDisease

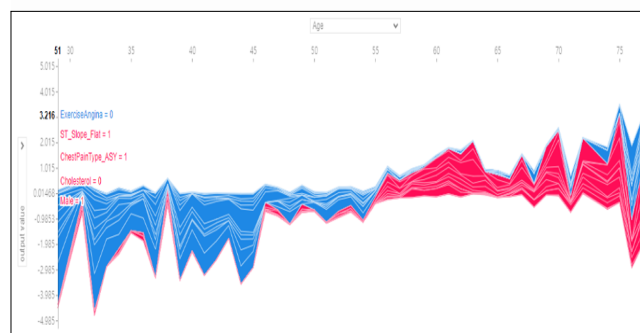


그림 3-1은 summary\_plot의 드롭박스를 이용해 전체 데이터에 대한 샐플리 영향도 중 Age(나이)와 심장 질환에 대해 상관관계를 필터링한 결과입니다. x축에서 우측으로 갈 수록 나이가 증가하는 것을 보이며 55세를 기준으로 55세 미만은 심장질환에 낮은 경향을 확인 할 수 있고, 55세 이상부터 심장질환에 높은 경향이 있음을 확인 할 수 있습니다.

그림 3-2 전체 데이터 샐플리 값 MaxHR-HeartDisease

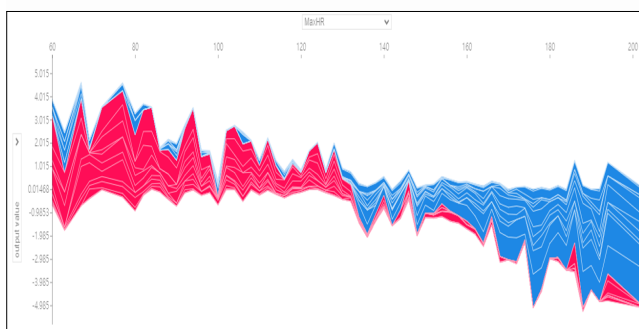


그림 3-2는 전체 데이터에 대한 샐플리 영향도 중 MaxHR(최대심박수)와 심장 질환에 대해 필터링한 결과입니다. 그림을 확인한 결과 최대 심박수 130수치 정도의 기준으로 최대심박수가 130미만인 데이터는 심장질환에 높은 경향을 확인할 수 있으며 130이상의 최대심박수를 가지는 데이터는 심장질환에 낮은 경향을 보이고 있는 것을 알 수 있습니다.

그림 4-1 데이터, 특징 별로 샐플리 값의 절댓값을 계산한 막대 플롯

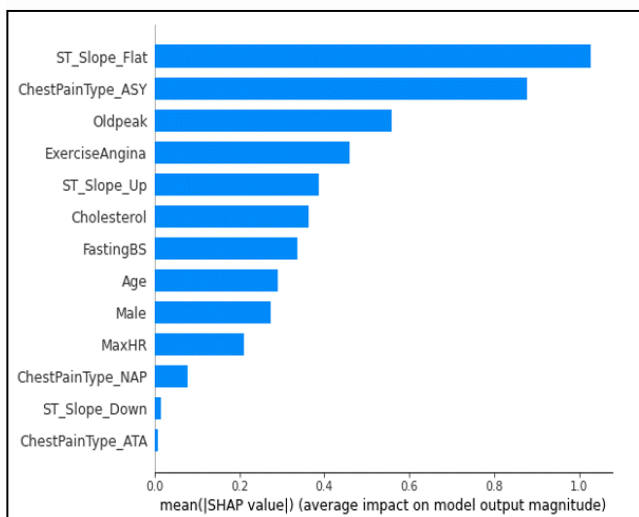


그림 4-2 전체 특징이 샐플리 값 분포의 기여도를 시각화

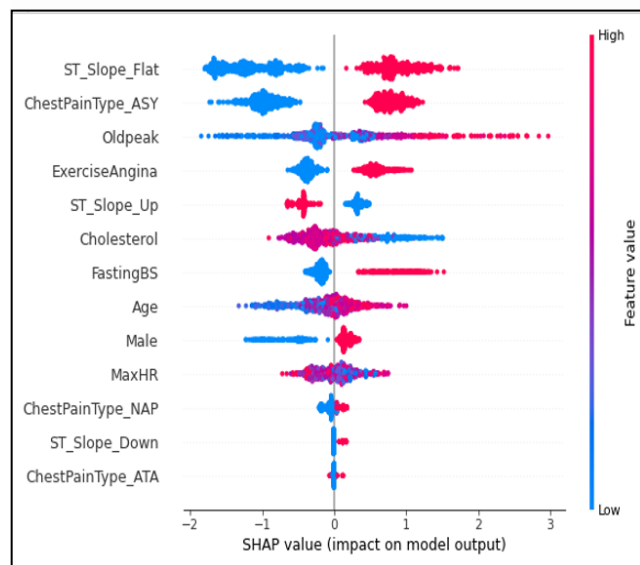


그림4-1, 4-2는 동시 비교에 효율성이 좋습니다. 그림4-1은 각 특징이 모델에 미치는 절대 영향도를 나타낸 결과입니다. y축의 순서로는 모델 예측의 절대 영향도가 큰 순서입니다. 상위 4개의 특징으로는 ST\_Slope\_Flat, ChestPainType-ASY, Oldpeak, ExerciseAngina이 있습니다. 그림 4-2는 4-1의 순서와 동일하고, 각 특징의 기여도가 샐플리 값에 얼마나 미치는지 시각화 한 그림입니다. x축은 샐플리 값, y축은 특징, 우측의 색깔의 정도는 특징 값에 대한 축입니다. 상위 4개의 특징은 그림4-1과 같으며 이를 통해 상위 4개의 특징이 심장질환을 결정하는데 큰 역할을 할 수 있음을 알 수 있습니다. 이 과정을 통해 SHAP은 모델 예측에 개별, 전체적인 요인들의 기여도를 알 수 있으며 설명 가능한 방법입니다. 또한 모델의 특성을 이용하지 않기 때문에 모델에 상관 없이 적용 가능한 Model-agnostic특징을 가지고 있으며 적용할 수 있는 SHAP종류로는 KernelShap, DeepShap, TreeShap이 있습니다. 저희는 특징의 개수가 많을 뿐더러 특징 전체를 통해 심장질환을 예측하는 모델이기 때문에 가장 효과적인 TreeShap을 사용하였습니다.

---

## IV결론

이번 프로젝트는 심장병 예측을 위한 XGBoost의 유용성을 입증하였고, 이 진단과 가장 관련이 있는 요인에 대한 통찰력을 제공하였습니다.

심장 질환 발병에 많은 영향을 끼쳤던 대표적인 특성으로는 ST분절의 평평한 기울기와 가슴 통증 유형중 무증상과 운동중 유발된 ST분절 하강과 운동으로 유발된 협심증이 있었습니다. 위와 같은 요인들을 미리 확인하여 모델에 적용해 예측하면 현재 본인의 심장 건강 상태에 관해 1차적으로 예측할 수 있고 심장질환 발병요인들을 제거 함으로써 의사와 기타 의료 전문가가 심장 질환의 조기 발견 및 진단에 도움을 주며 질병에 예방할 수 있을것으로 기대됩니다.

## <참고 문헌>

### 데이터 세트 출처

출처 : <https://www.kaggle.com/code/gkitchen/heart-disease-prediction-with-xgboost/notebook>

### 논문 출처

A Unified Approach to Interpreting Model Predictions

저자 - [Scott Lundberg](#), Su-In Lee

### 도서 출처

XAI 설명 가능한 인공지능, 인공지능을 해부하다

저자 - 안재현 지음