

Deep Learning Final Team Project

# 딥러닝 팀프로젝트 최종발표

4조 고준서 권혜현 김하림 전규리

# 발표순서

1

## 문제 정의

Quick summary

2

## 성능 향상

AutoEncoder  
Grid Search

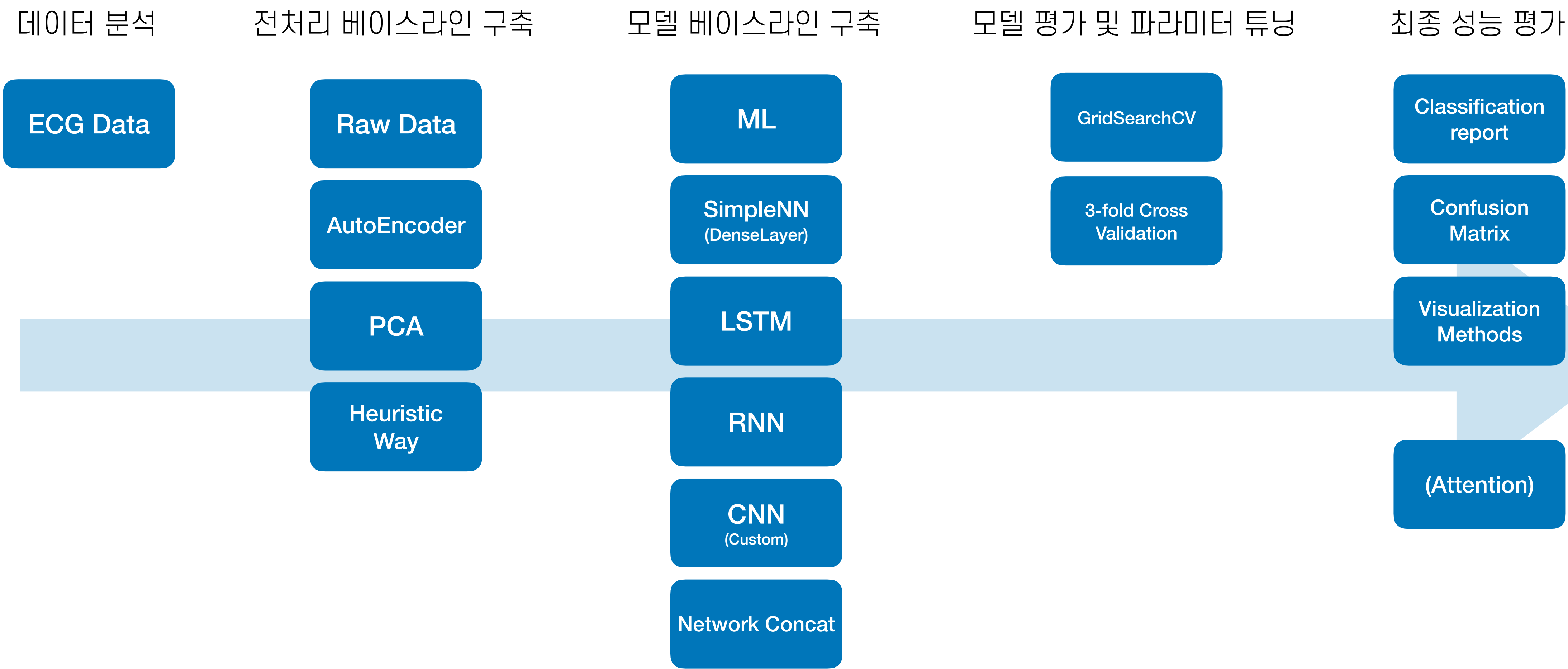
3

## 모델

Machine Learning  
Simple One Layer  
Neural Network  
LSTM

# Schedule For Modeling

문제 해결을 위한 여러가지 베이스 라인 검토



# Schedule For Modeling

문제 해결을 위한 여러가지 베이스 라인 검토

데이터 분석

전처리 베이스라인 구축

모델 베이스라인 구축

모델 평가 및 파라미터 튜닝

최종 성능 평가

ECG Data

Raw Data

AutoEncoder

PCA

Heuristic Way

ML

SimpleNN  
(DenseLayer)

LSTM

RNN

CNN  
(Custom)

Network Concat

GridSearchCV

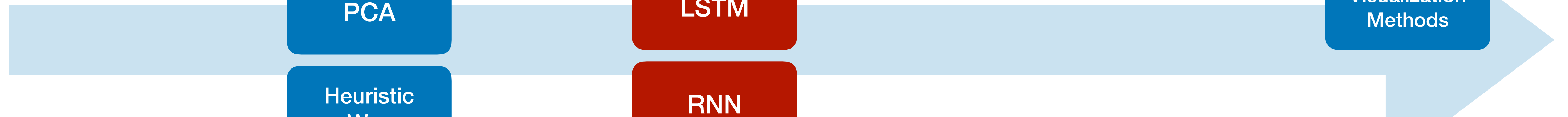
3-fold Cross  
Validation

Classification  
report

Confusion  
Matrix

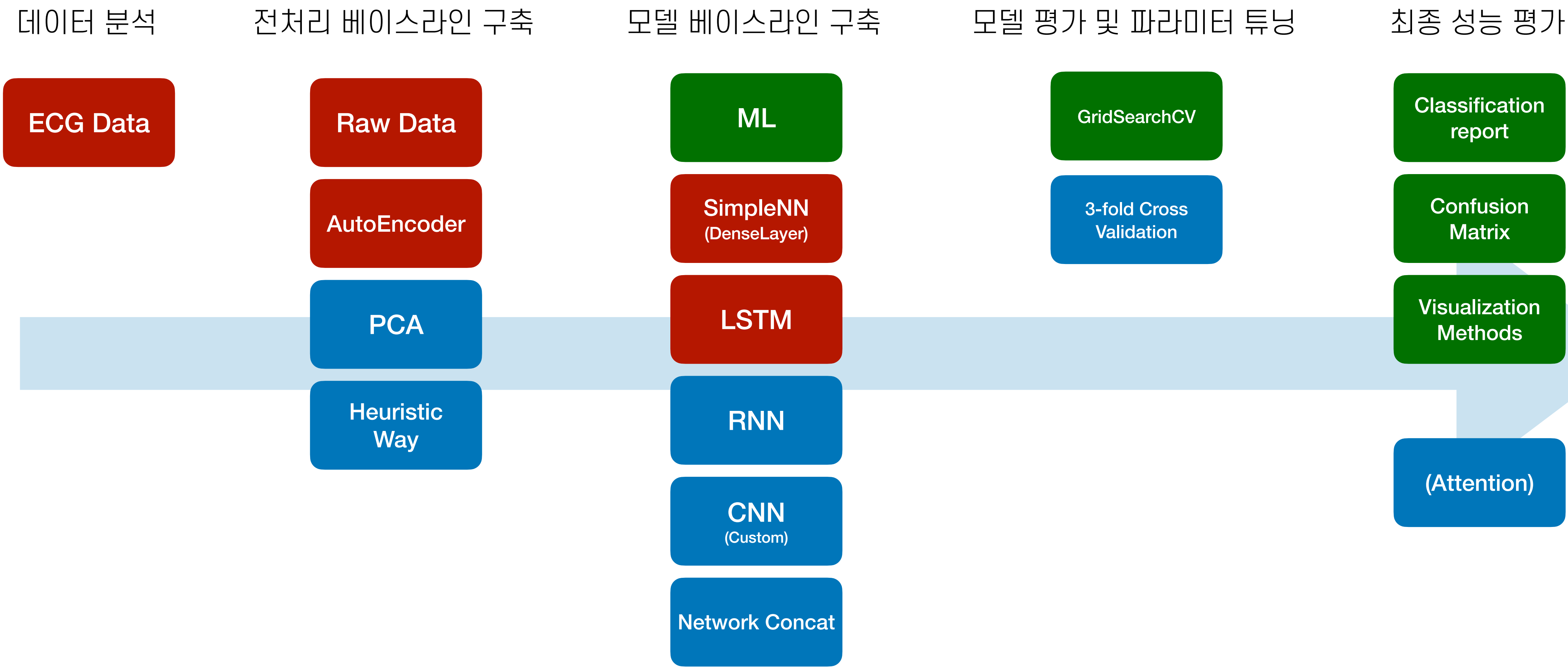
Visualization  
Methods

(Attention)



# Schedule For Modeling

문제 해결을 위한 여러가지 베이스 라인 검토



1

# 문제 정의

Problem Definition

ECG 데이터를 이용한 <sup>Normal repolarization</sup> 정상 재분극/<sup>Abnormal repolarization</sup> 비정상 재분극 분류

## 1 질병 예측

전문가들의 질병 진단에 보조적 역할

## 2 비정상 재분극

흔하게 나타나는 현상

심각한 결과 초래



# 문제의 라벨 설정

0

비정상 재분극 없음

1

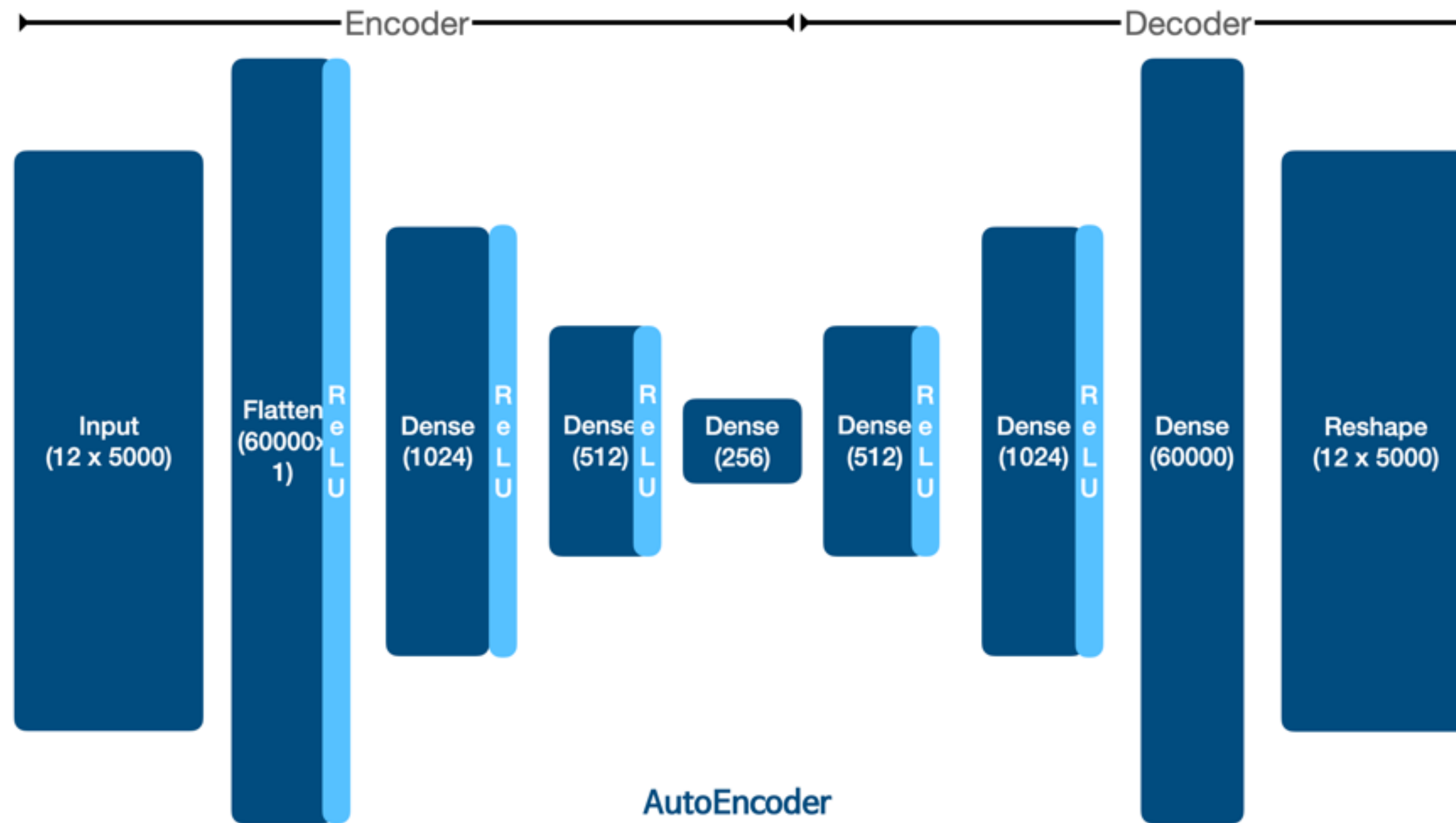
비정상 재분극 존재

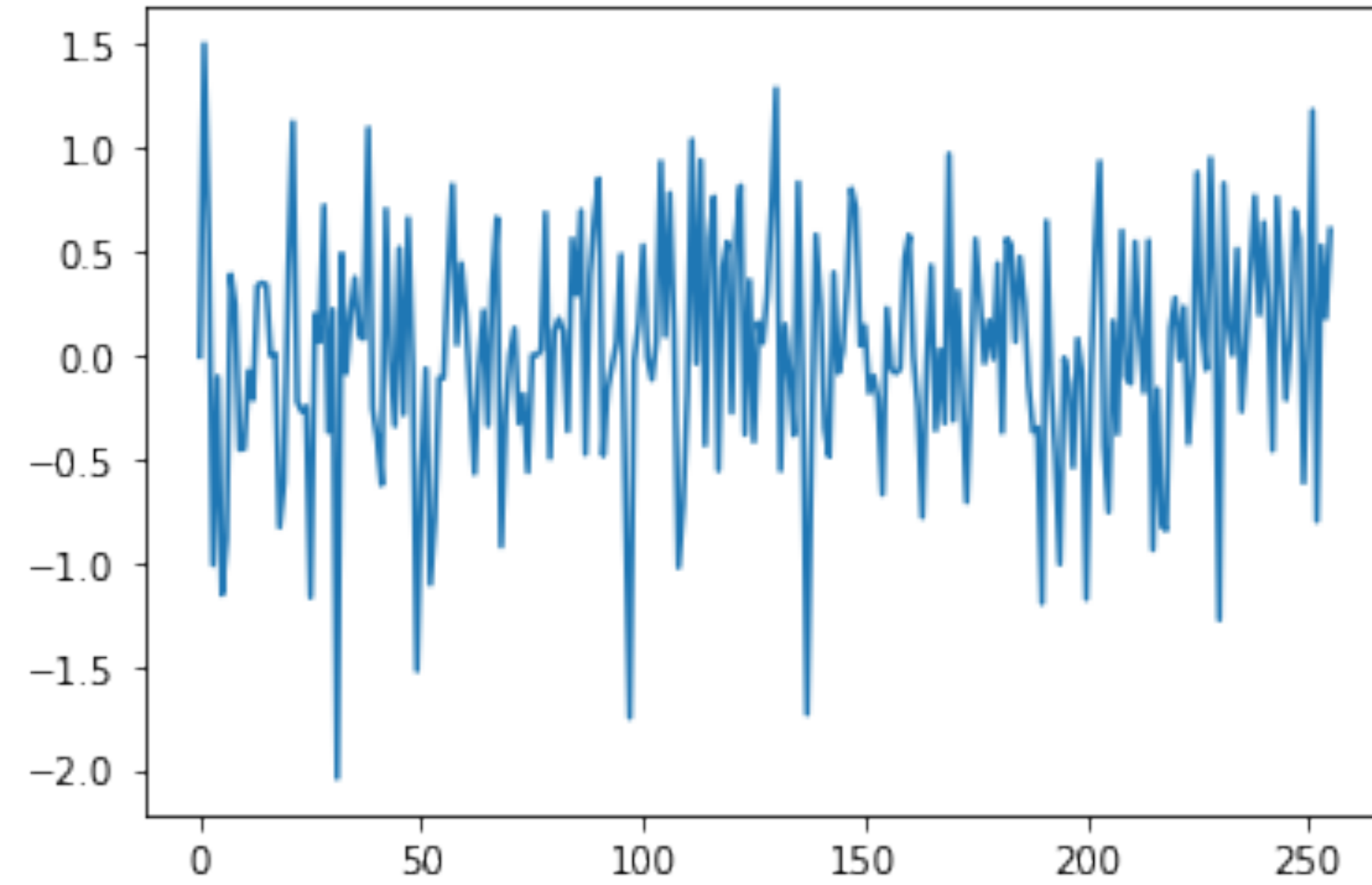
2

성능 향상

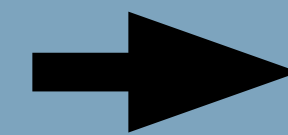
Performance Improvement

### 1 구조





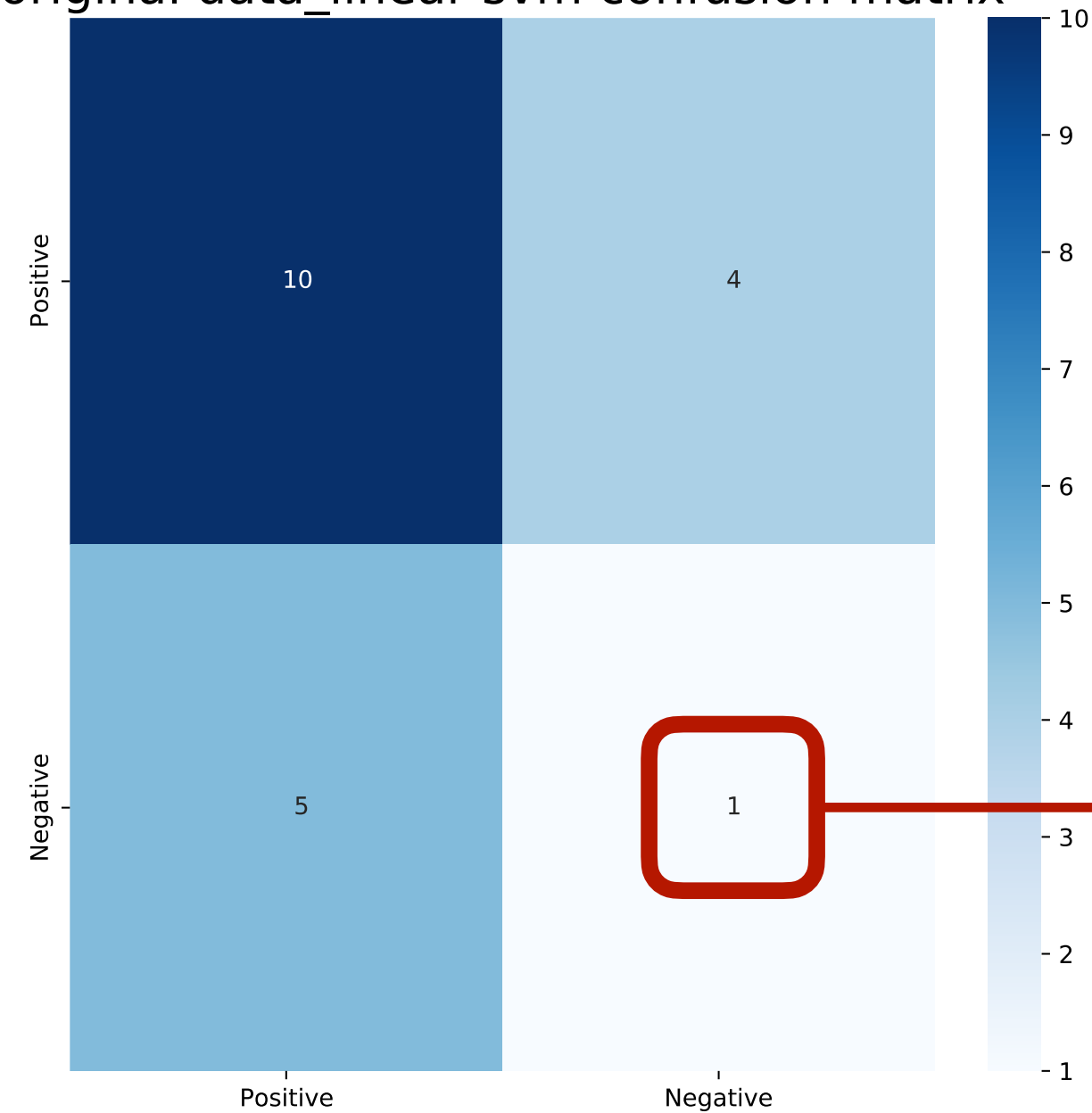
네트워크의 loss 최소화  
인코더 부분만을 활용



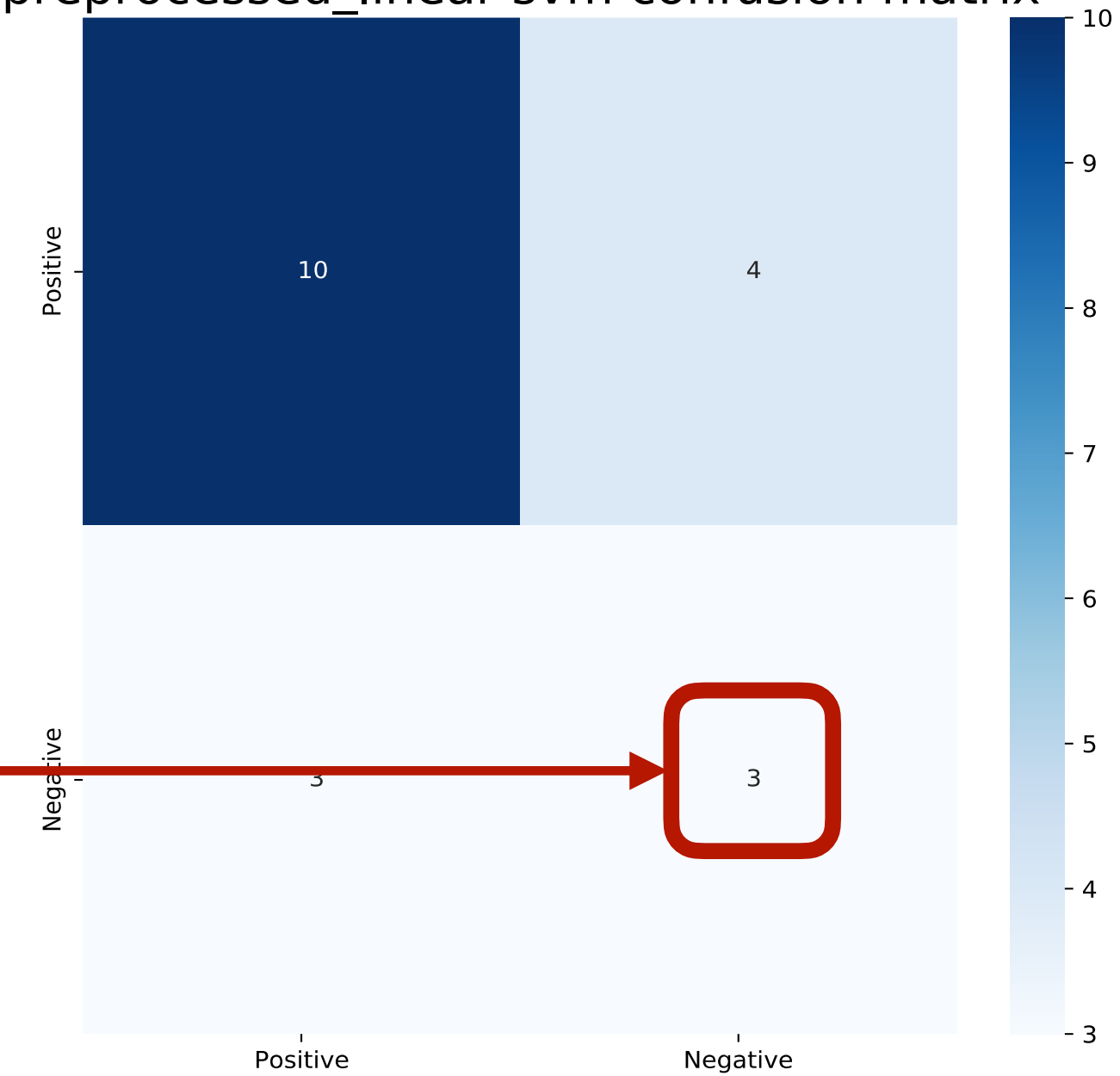
차원 축소  
[5000X12] -> [256X1]

### 2 성능 향상

original data\_linear svm confusion matrix



preprocessed\_linear svm confusion matrix



Negative를 negative로 예측하는 정확도 상승

### 1 설명

#### 하이퍼 파라미터 튜닝

여러 파라미터 지정 후 코드를 실행하면 모델에 가장 적합한 파라미터의 조합을 찾아 모델을 피팅하고 예측할 수 있다.

#### - 사용 예시

```
1 param_grid = {'C': [0.01, 0.1, 1, 10, 100],  
2               'penalty': ['l1', 'l2']}  
3 gs = GridSearchCV(estimator=lr, param_grid=param_grid, scoring='accuracy', cv=5)
```

3

# 모델과 성능

Models and Performance

## 1 Logistic Regression

독립변수의 선형 결합을 이용하여 발생 가능성을 예측하는 통계기법

- 모델 선정 이유

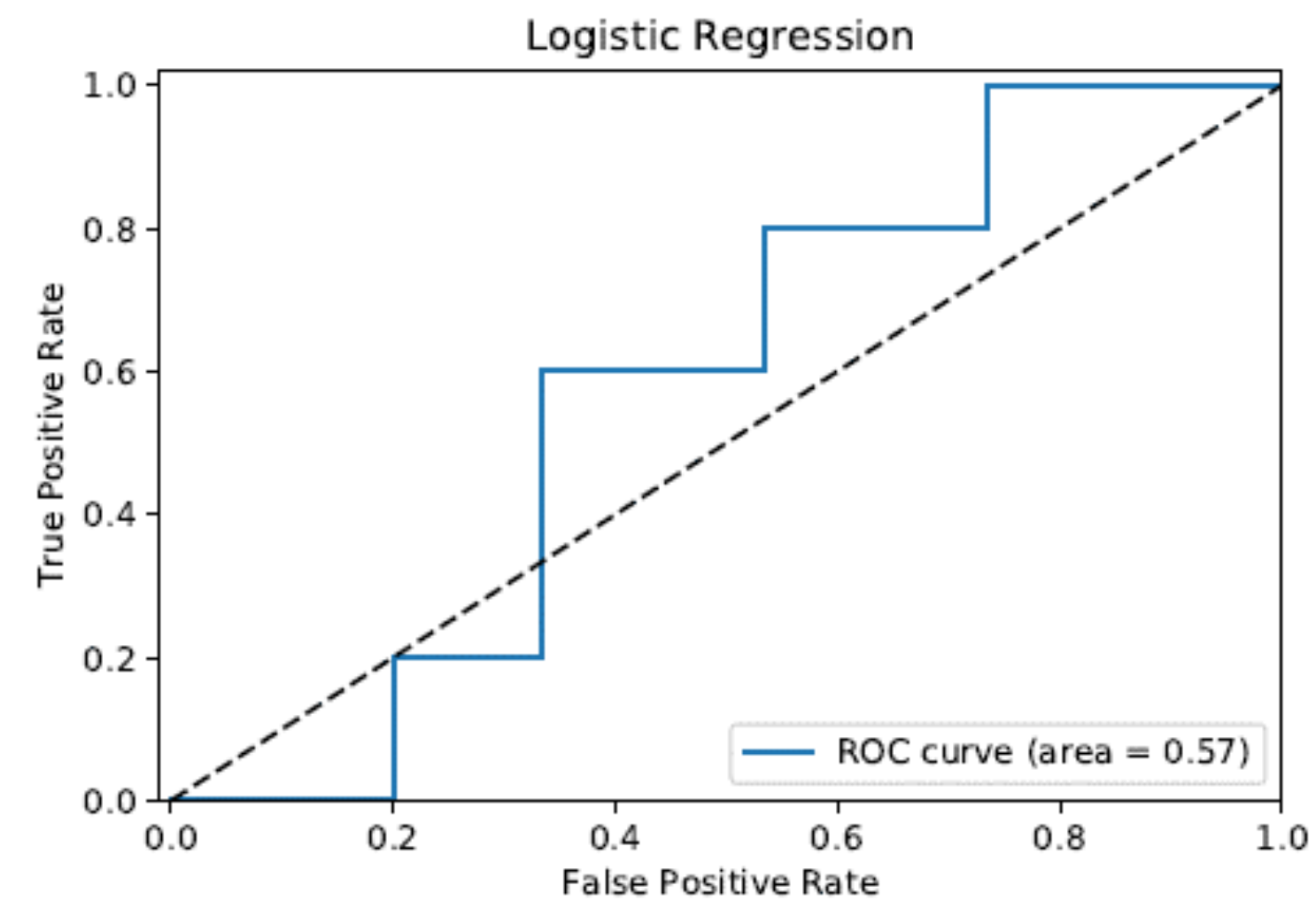
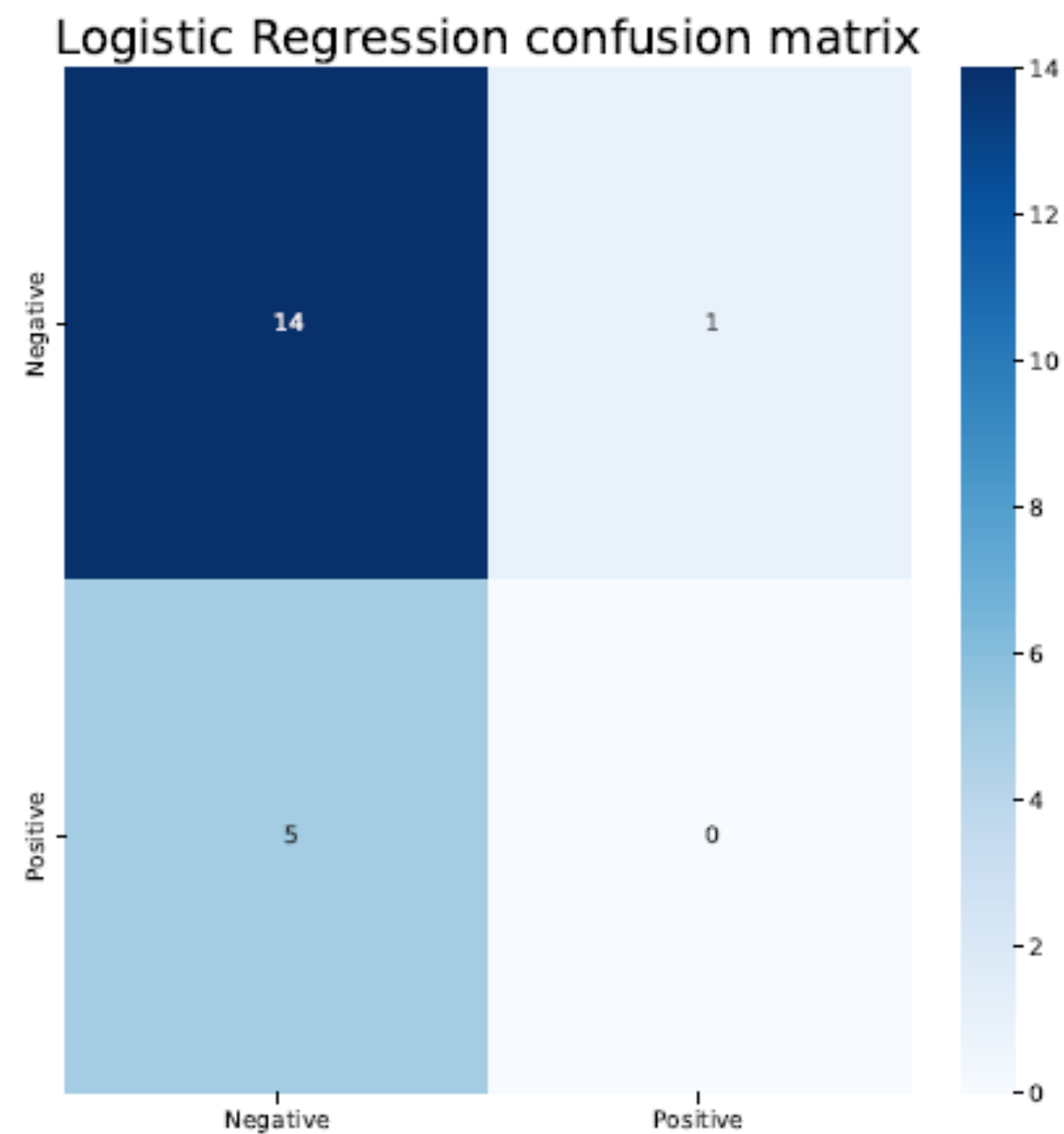
의료 분야 중 특히 질병의 위험인자나 요인 예측에 활용



# Machine Learning Methods

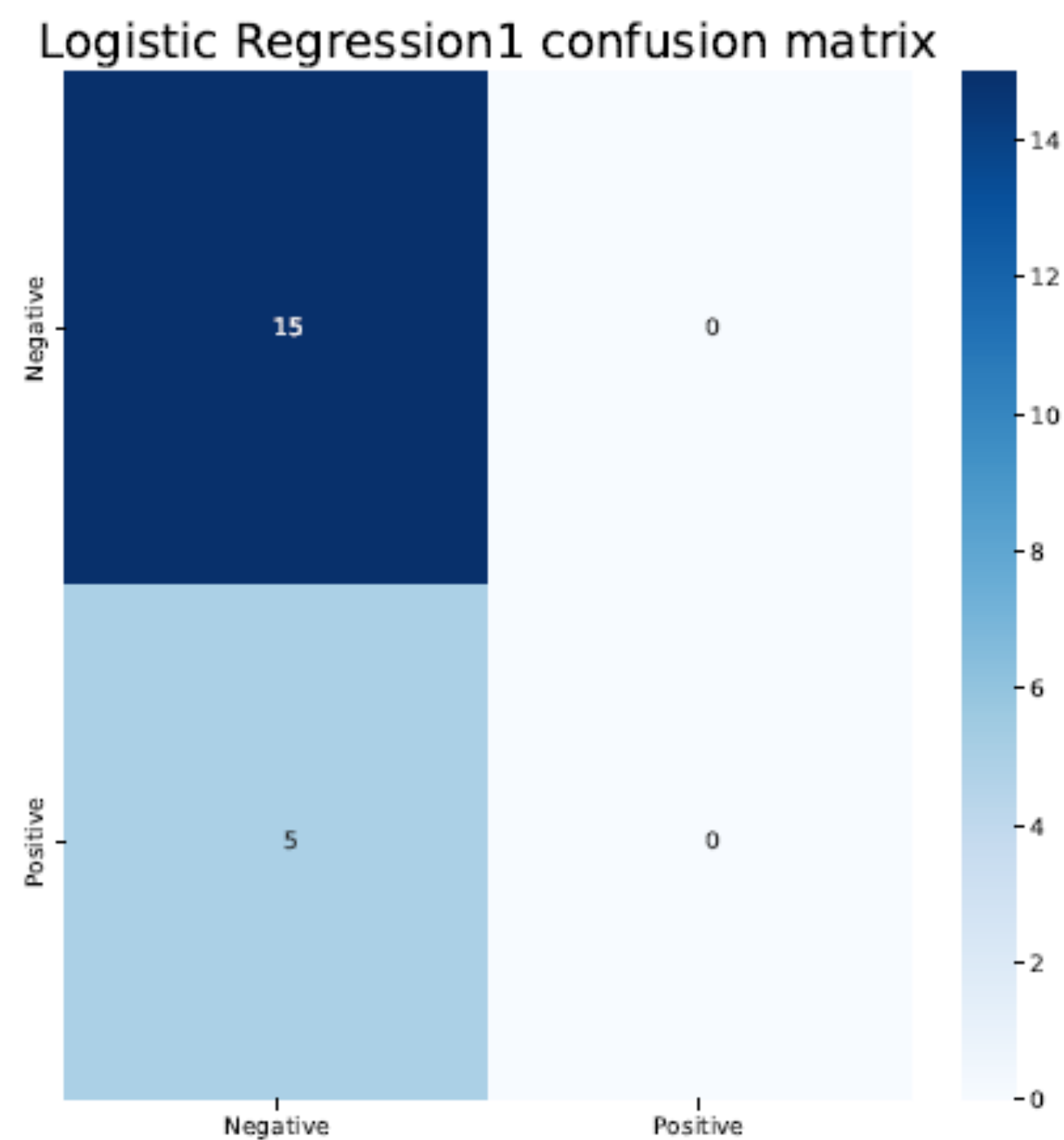
## 1 Logistic Regression

- 예측 결과 기존 데이터



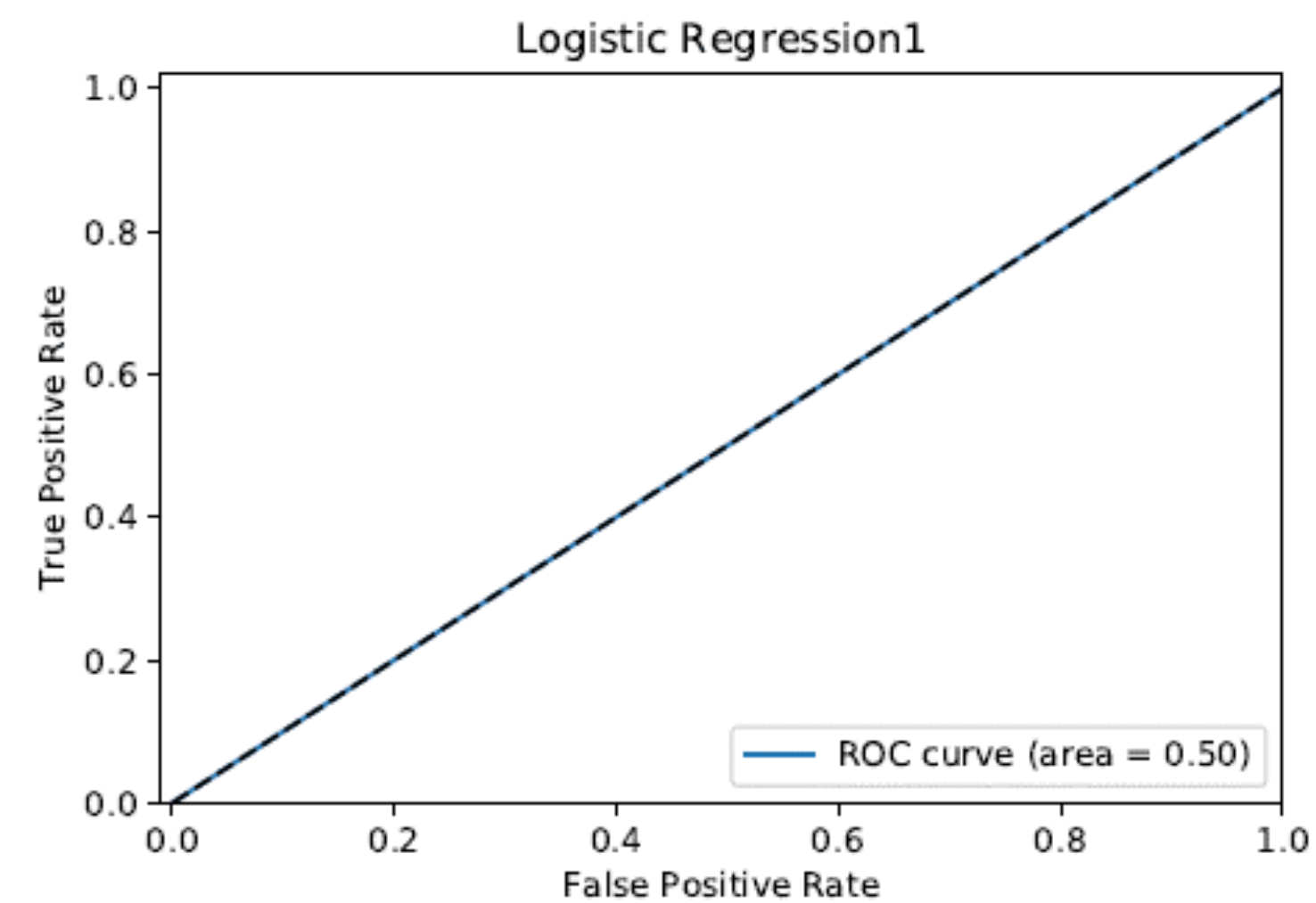
## 1 Logistic Regression

- 예측 결과 전처리 데이터



성능저하

비선형 데이터에 약한 LR  
비선형성이 뚜렷해진 전처리 데이터



## 2 XGBoost Classifier

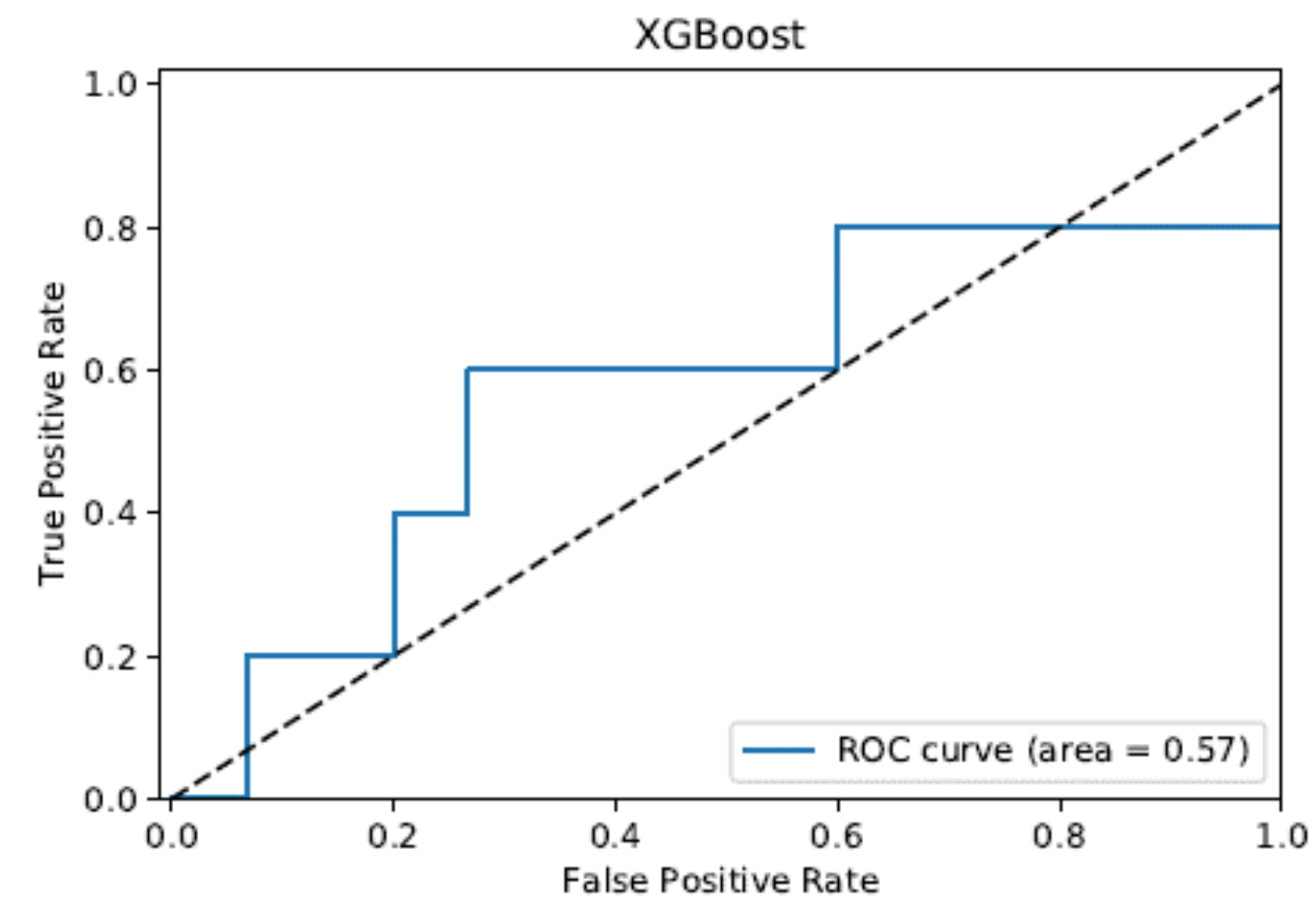
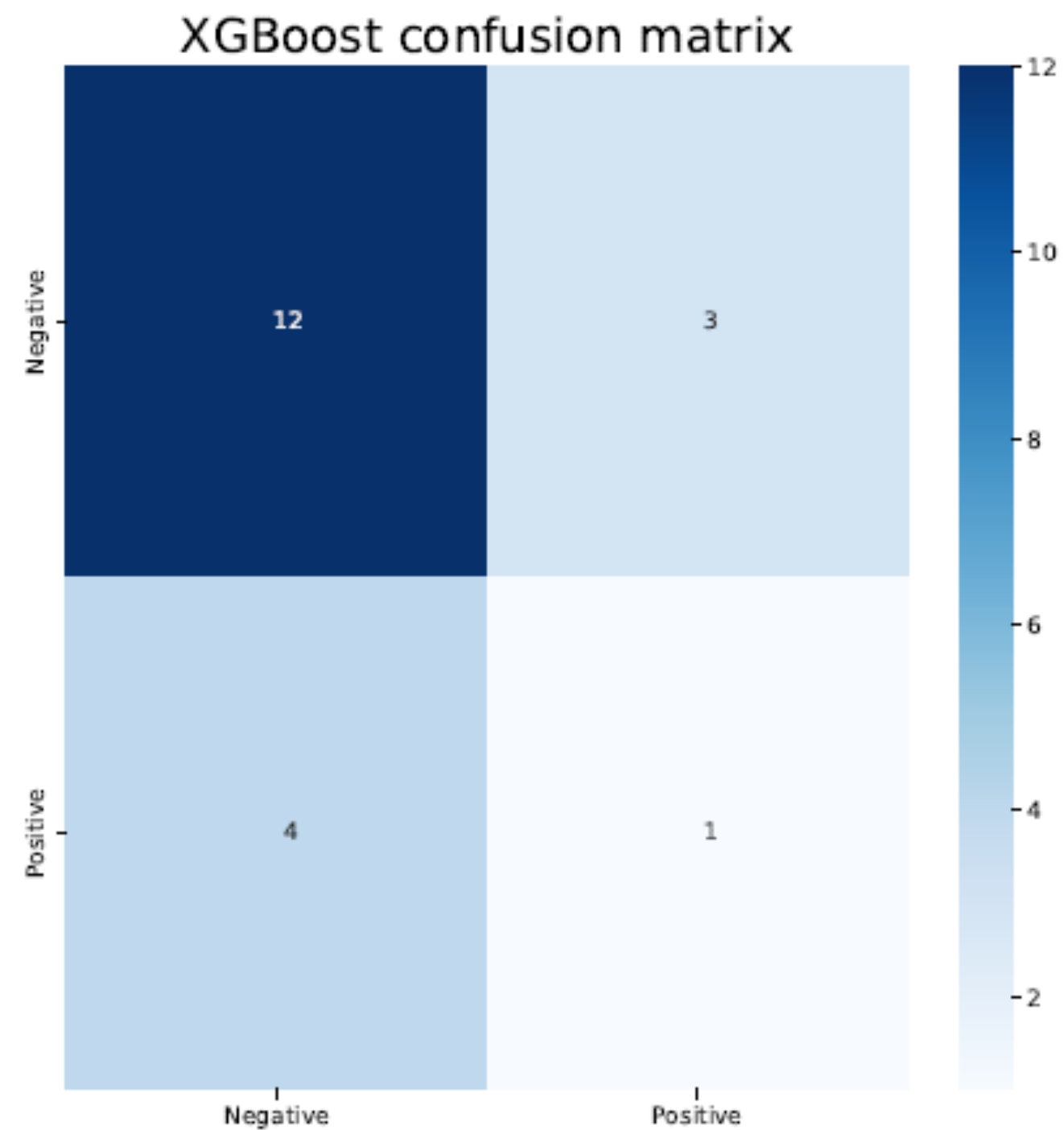
의사결정 기반의 앙상블 머신러닝 알고리즘

- 모델 선정 이유

빠른 학습과 분류로 인한 빠른 속도  
분류 문제에서의 높은 성능

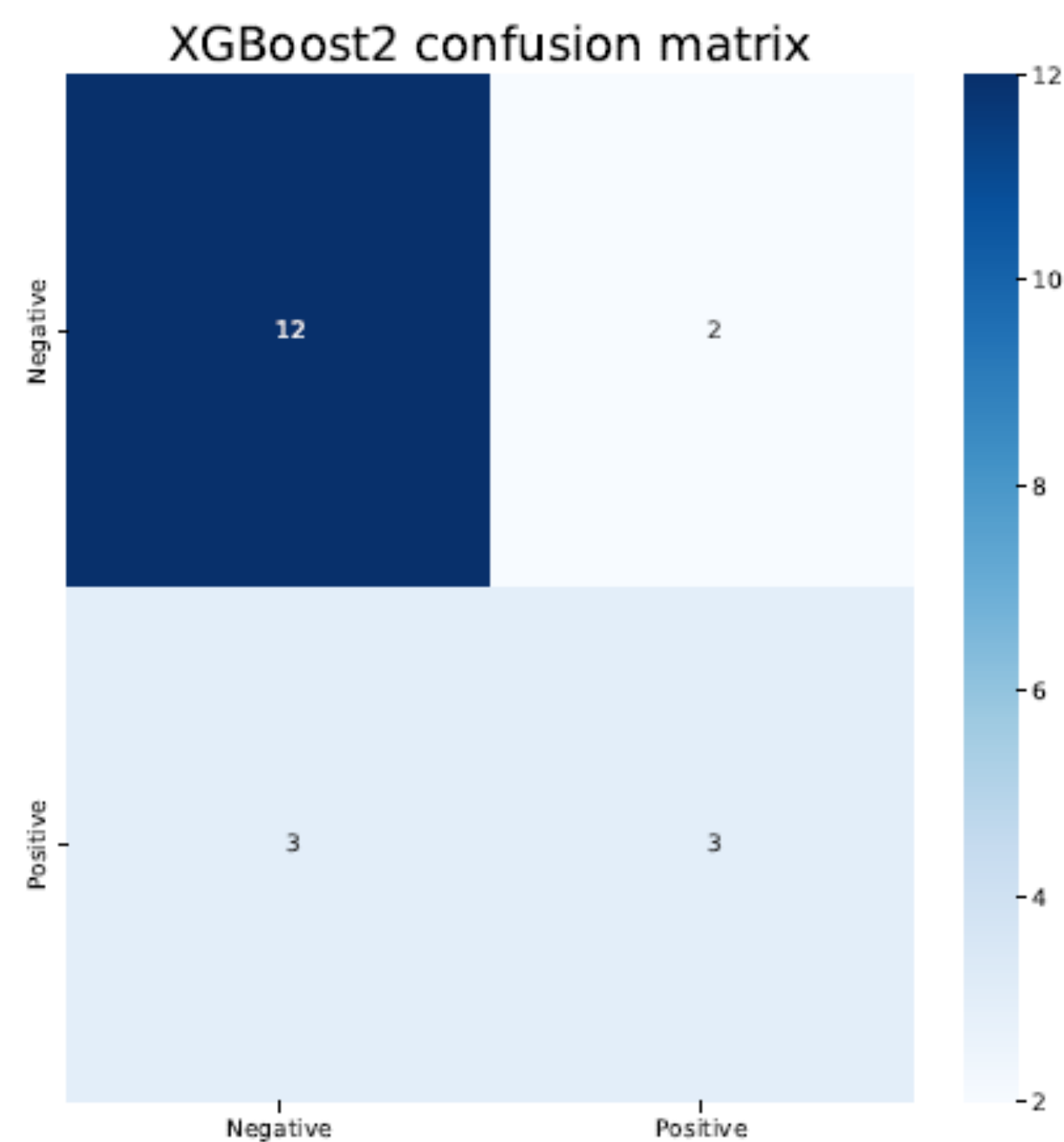
## 2 XG Boost Classifier

- 예측 결과 기존 데이터



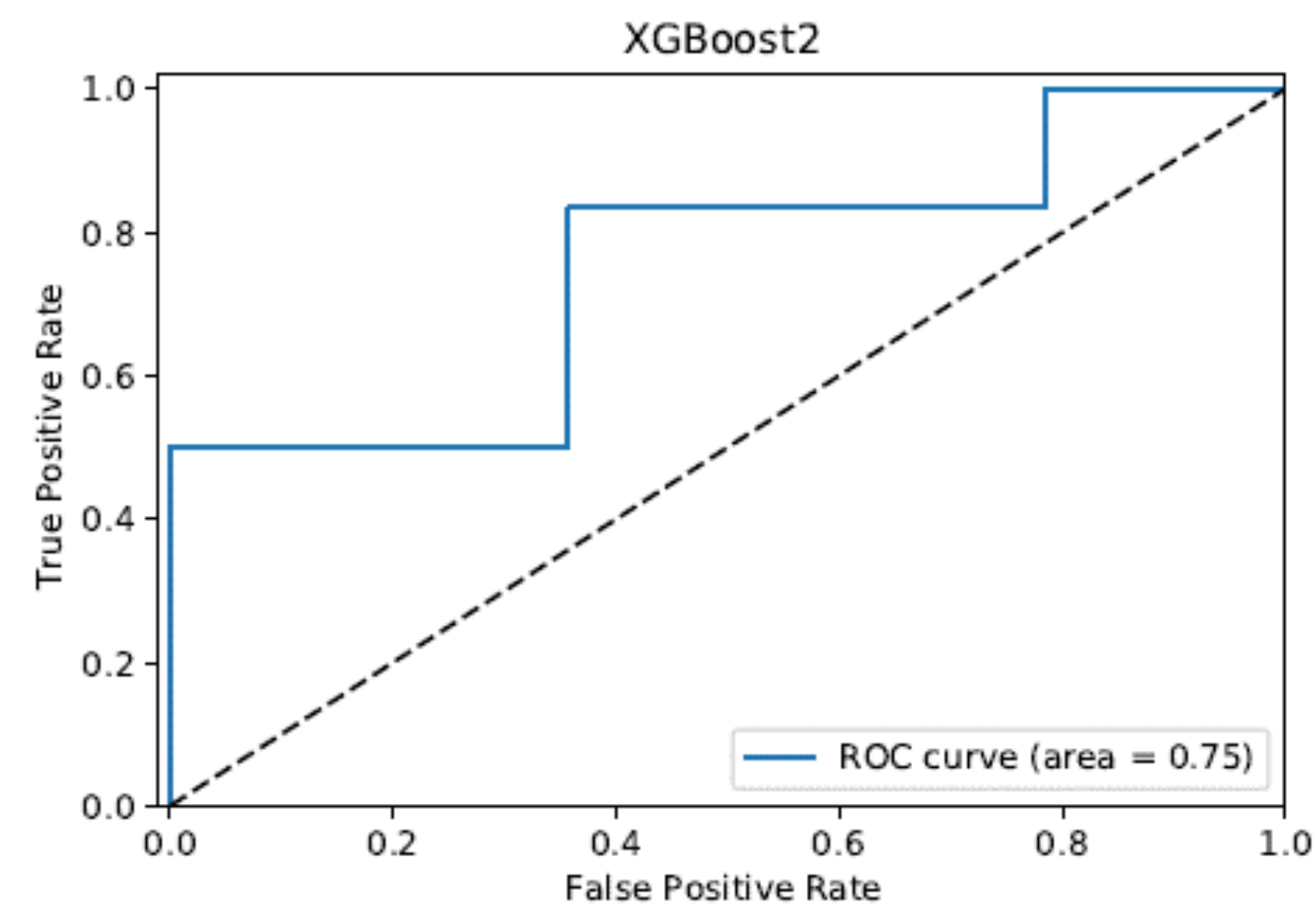
### 2 XG Boost Classifier

- 예측 결과 전처리 데이터



성능향상

Grid Search CV, Stratified K-fold,  
Early Stopping, Evaluation Set 지정

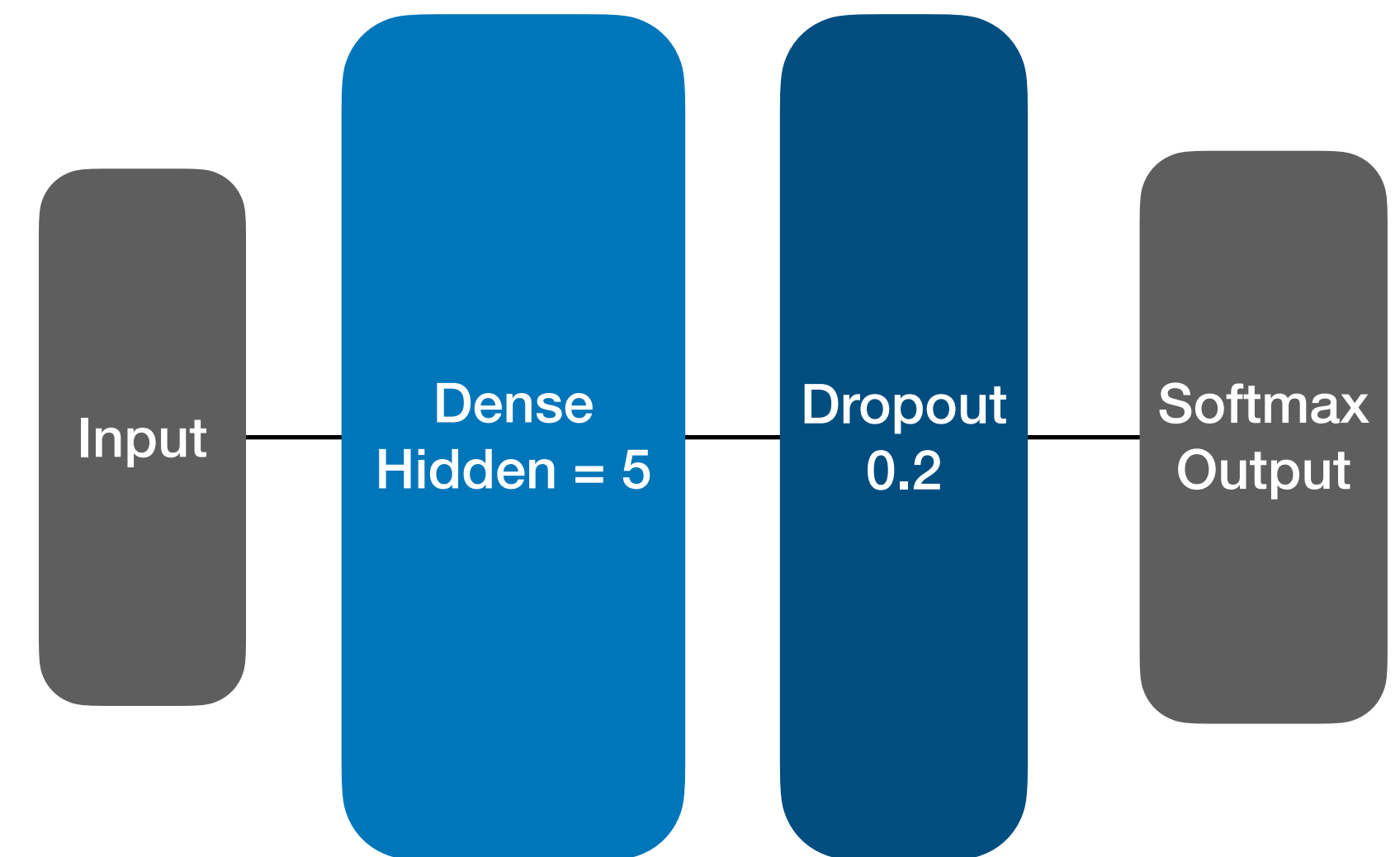


# 1 Simple One Layer Neural Network

딥러닝의 가장 기본적인 신경망

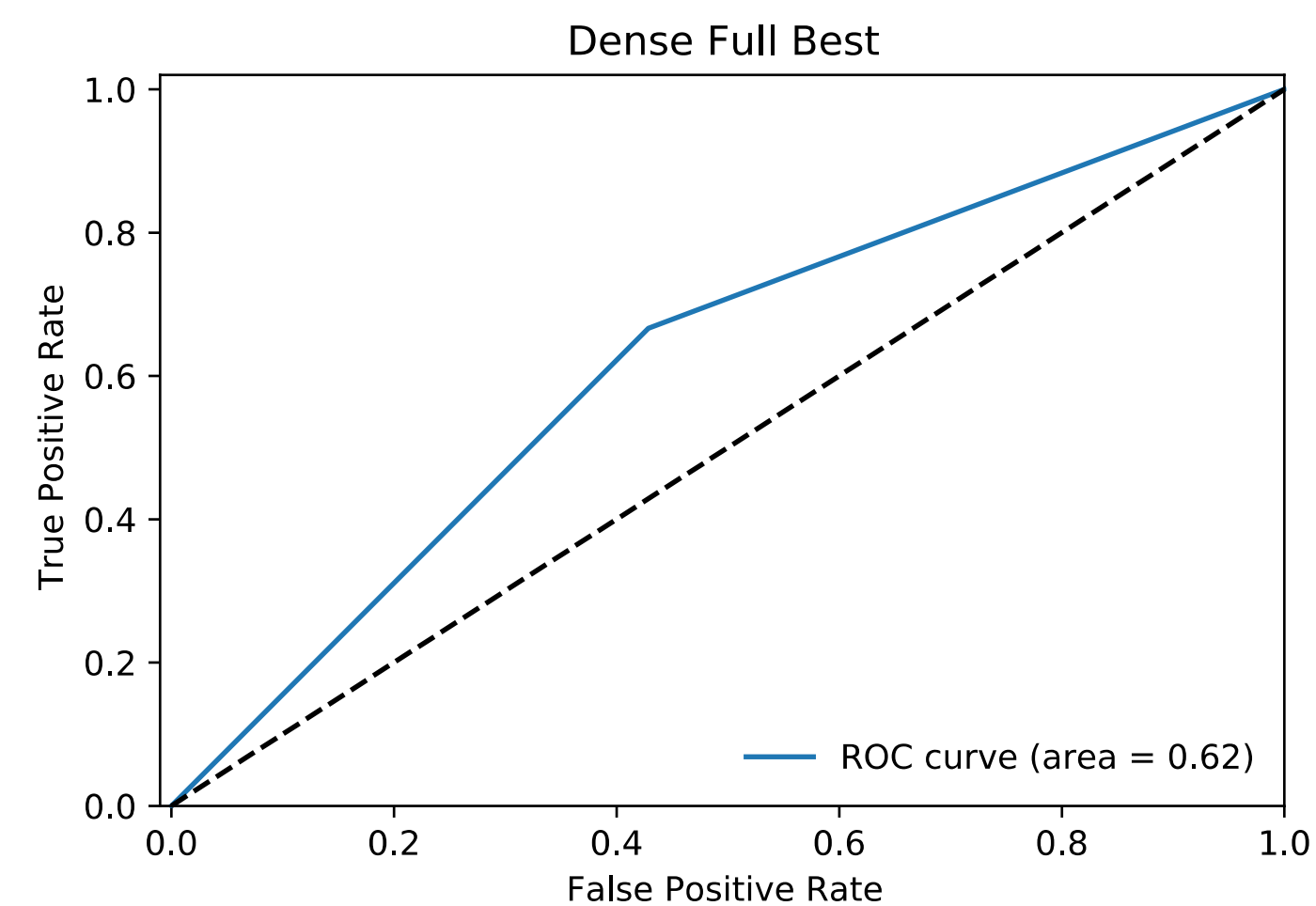
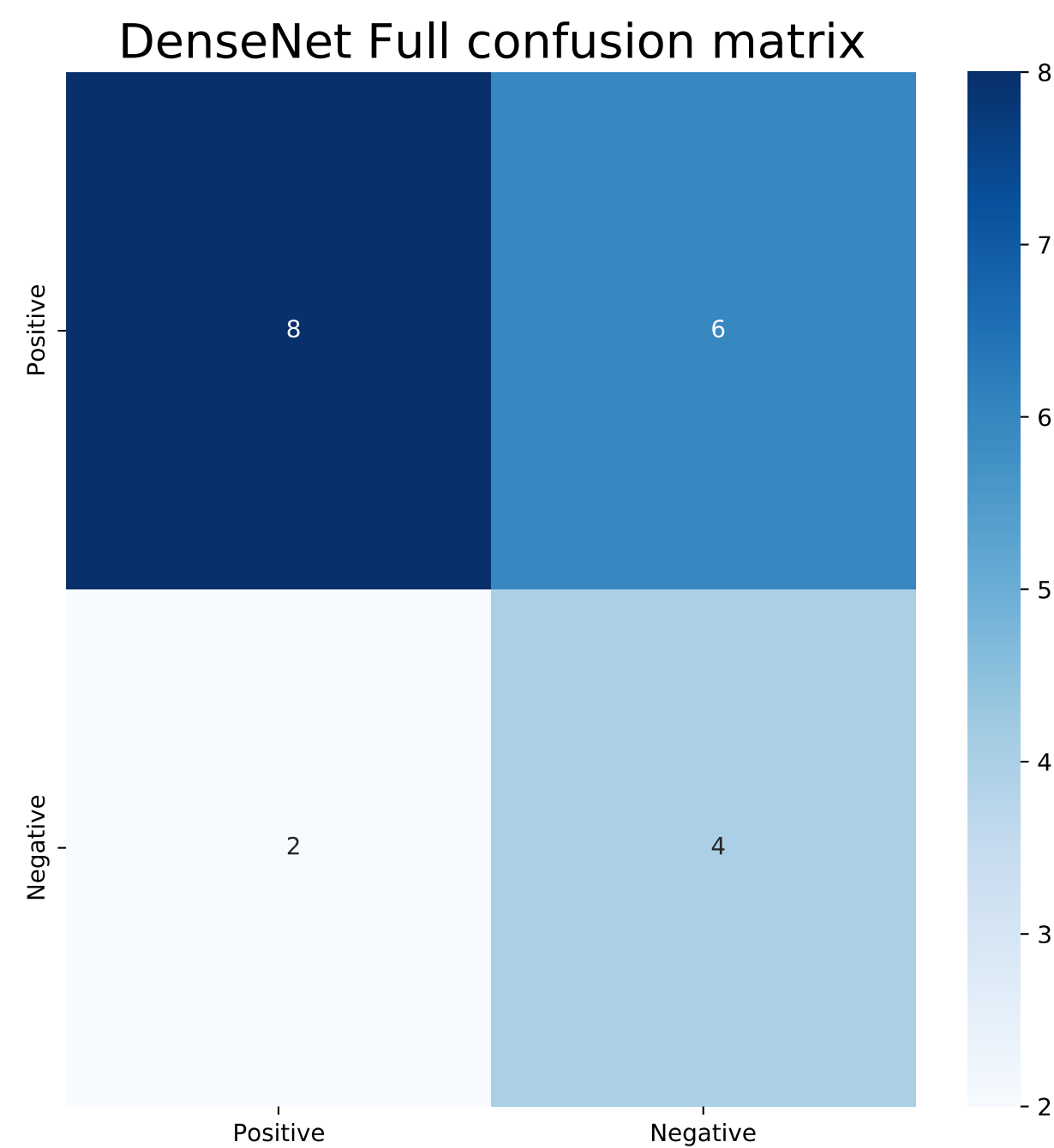
- 모델 선정 이유

다소 단순한 구조로, 간편한 사용



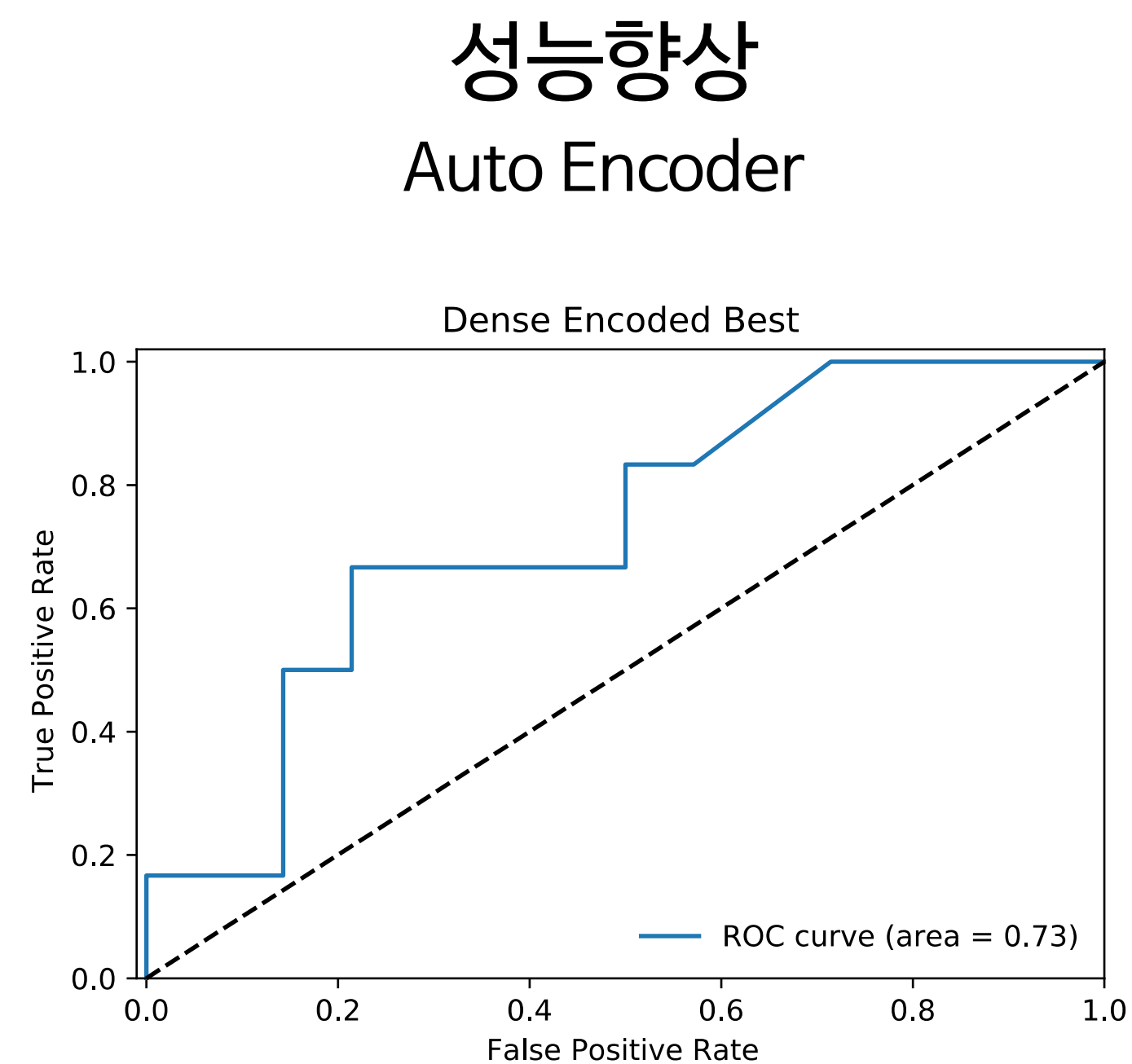
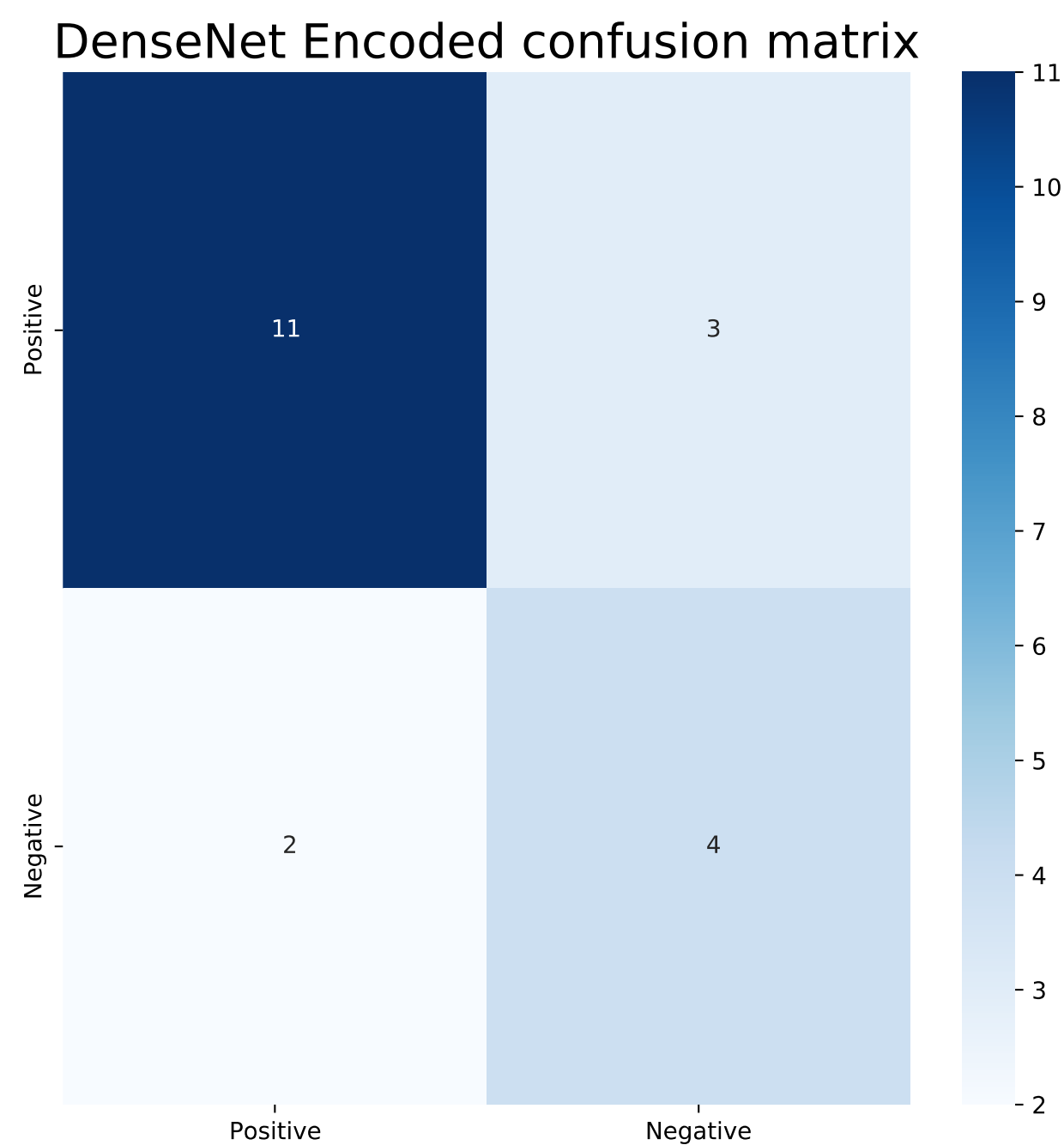
# 1 Simple One Layer Neural Network

- 예측 결과 기존 데이터



# 1 Simple One Layer Neural Network

- 예측 결과 전처리 데이터





## 2 LSTM

단방향 노드를 가진 네트워크

- 모델 선정 이유

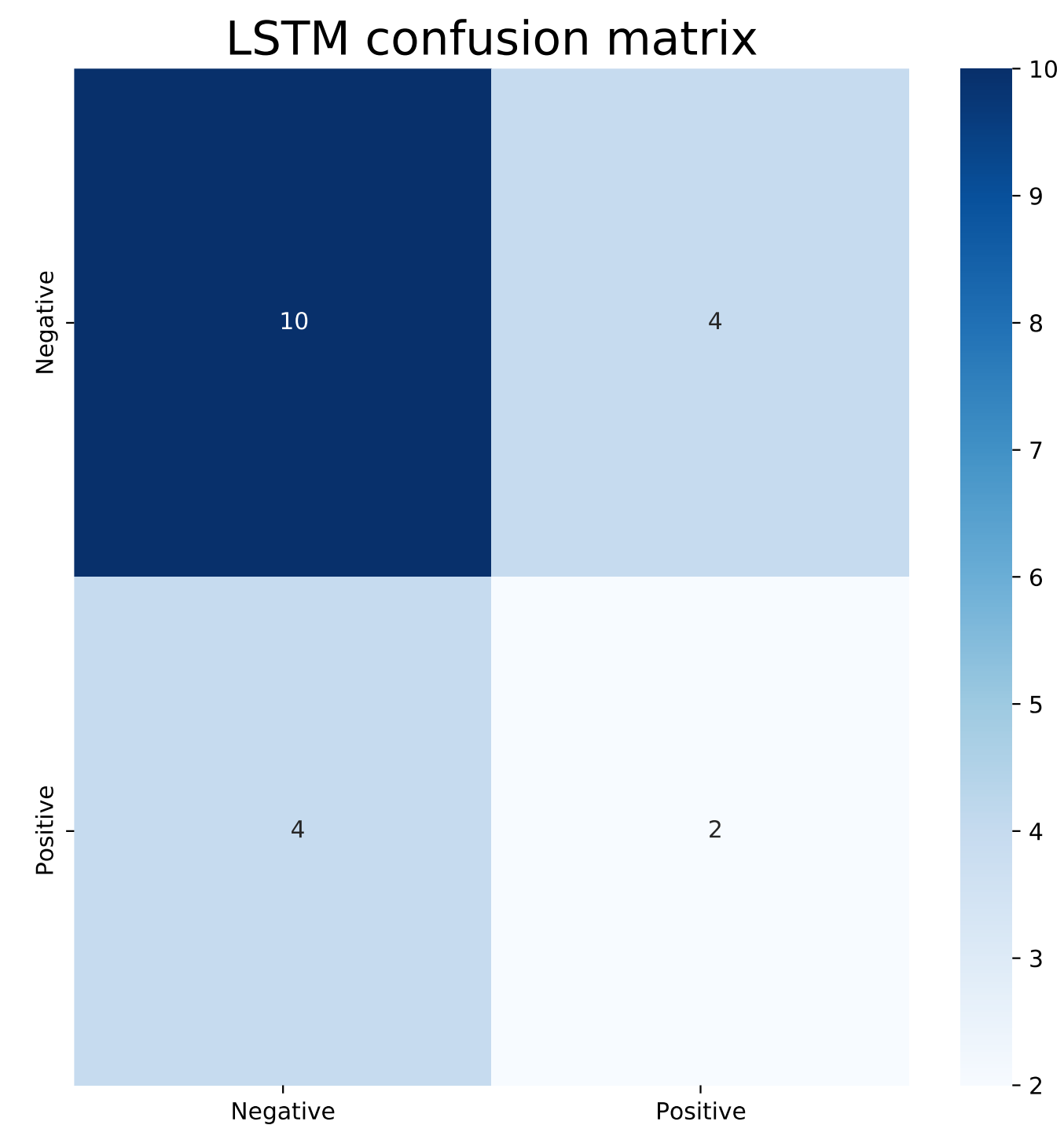
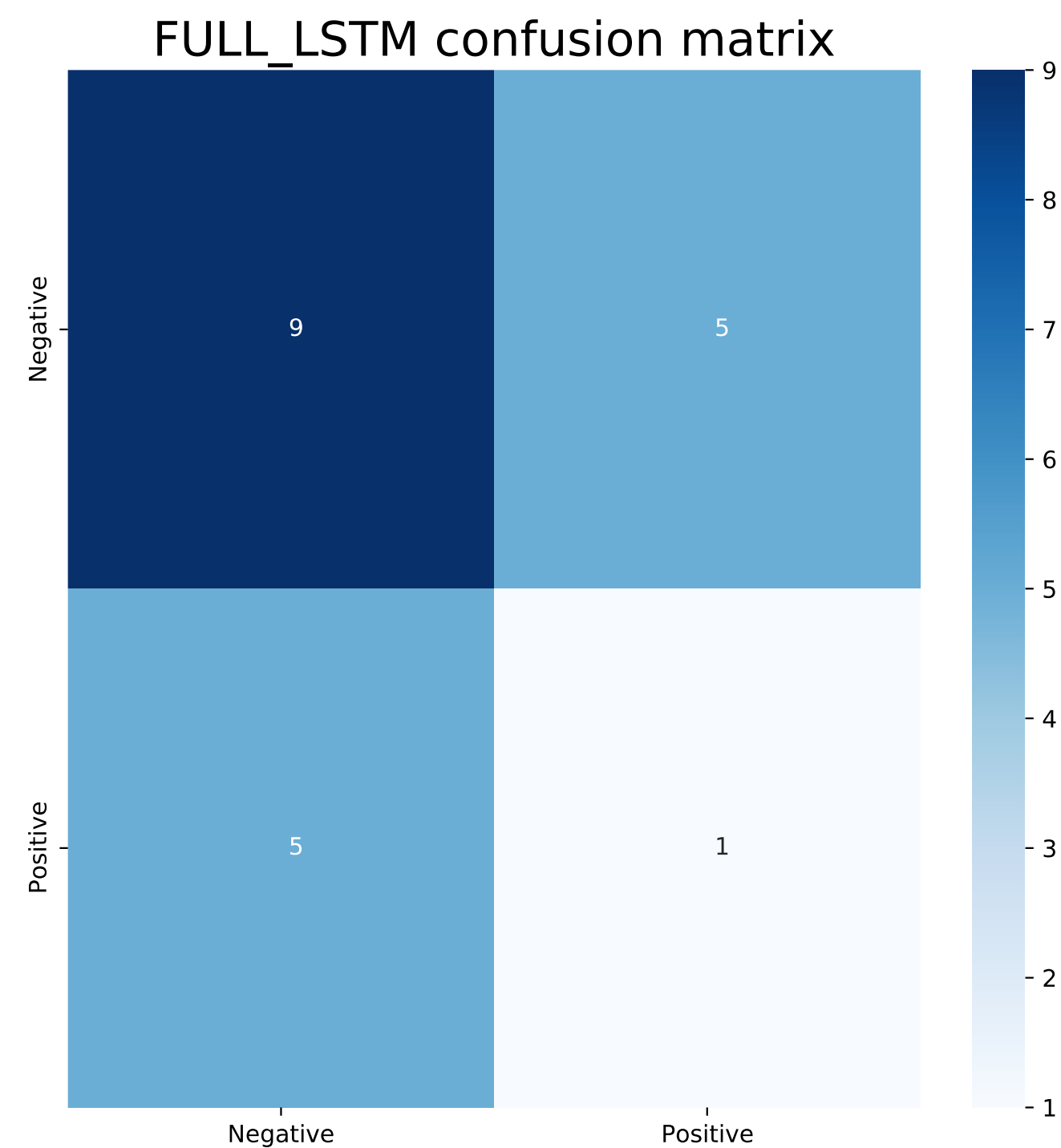
Vanishing Gradient 문제 해소 LSTM은 방향노드를 가지기에 다음 레이어로 gradient 전달

연속성 있는 데이터에 많이 활용

## 2 LSTM

- 예측 결과 기존 데이터

전처리 데이터



### 2 LSTM

#### - 예측 결과

성능향상

Auto Encoder

Dense-layer를 활용한 단층신경망보다 다소 낮은 성능 향상

오토인코더로 인한 연속성 상실이 오토인코더에서 성능 향상을 보이지 못하게 만든 원인  
(Logistic Regression이 인코딩 데이터에 대해서 낮은 성능을 보이는 것과 비슷한 이유)

LSTM의 파라미터가 단층신경망보다 많지만, 그에 비해 데이터셋이 부족해서 학습 저하  
(전체적으로 학습 정확도가 하락)

의사의 보조적 역할을 수행

4조 고준서 권혜현 김하림 전규리

# 감사합니다

자유롭게 질문해주세요!