



oculus

# 웹 로그 데이터를 활용한 이상 접속 탐지 시스템 개발

# 목차 a table of contents

---

- 1** 프로젝트 개요 및 목표
- 2** 데이터 준비 및 분석
- 3** 모델 개발 및 결과
- 4** 모델 한계점 및 개선 방향
- 5** 결론 및 향후 계획





# Part 1

## 프로젝트 개요 및 목표

## Part 1 프로젝트 개요

---

목표

웹 로그 데이터를 분석하여 비정상적인 접근을 탐지하는 모델 구축

성과  
지표

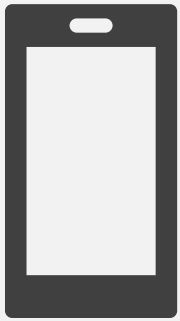
정확도, 정밀도, 재현율, F1 Score

활용  
가능성

비정상적인 접근을 실시간으로 탐지해 차단할 수 있는 시스템 제공

## Part 1 문제 정의

---



해결하려는 문제



시나리오



기대 효과

## Part 1 프로젝트 진행 순서

---







# Part 2

## 데이터 준비 및 분석



## Part 2 초기 데이터 파악

아이피주소	접속국가	접속상대	정상/비정상
213.152.161.239	NL	Global Layer B.V.	1
45.83.66.34	DE	Alpha Strike Labs GmbH	0
45.83.66.3	DE	Alpha Strike Labs GmbH	0
122.155.196.149	TH	CAT Telecom Public Company Limited	1
137.184.138.81	US	DIGITALOCEAN-ASN	1
45.83.67.202	DE	Alpha Strike Labs GmbH	0
45.83.67.16	DE	Alpha Strike Labs GmbH	0
45.83.65.67	DE	Alpha Strike Labs GmbH	0
45.83.64.115	DE	Alpha Strike Labs GmbH	0
152.70.112.97	US	ORACLE-BMC-31898	1

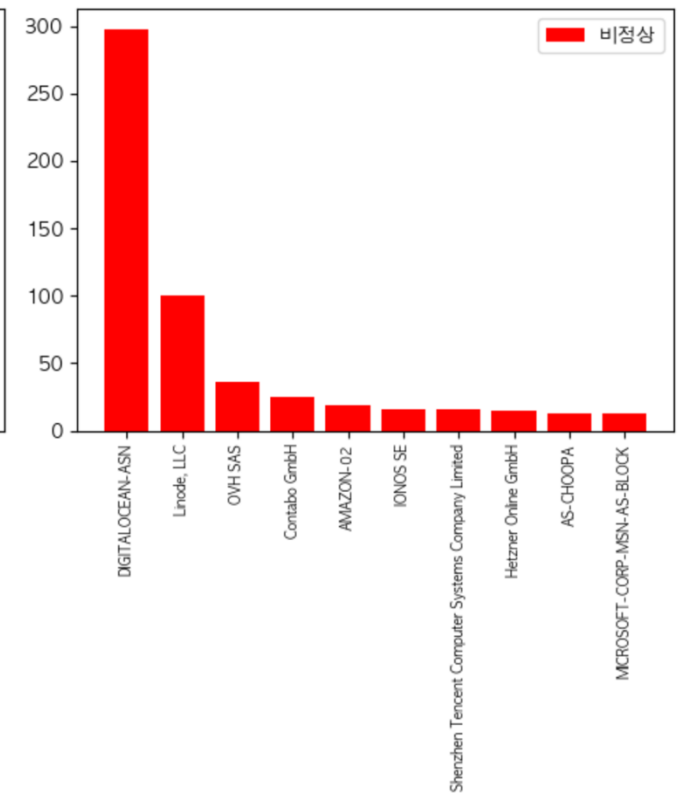
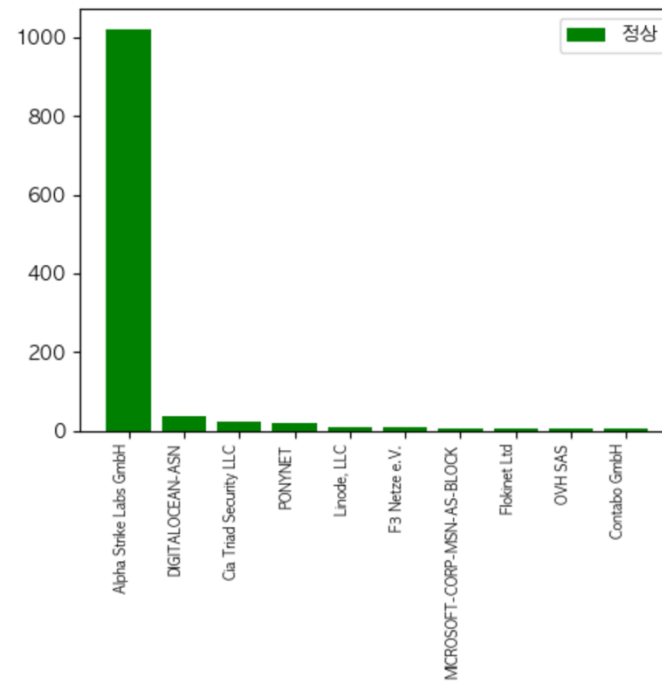
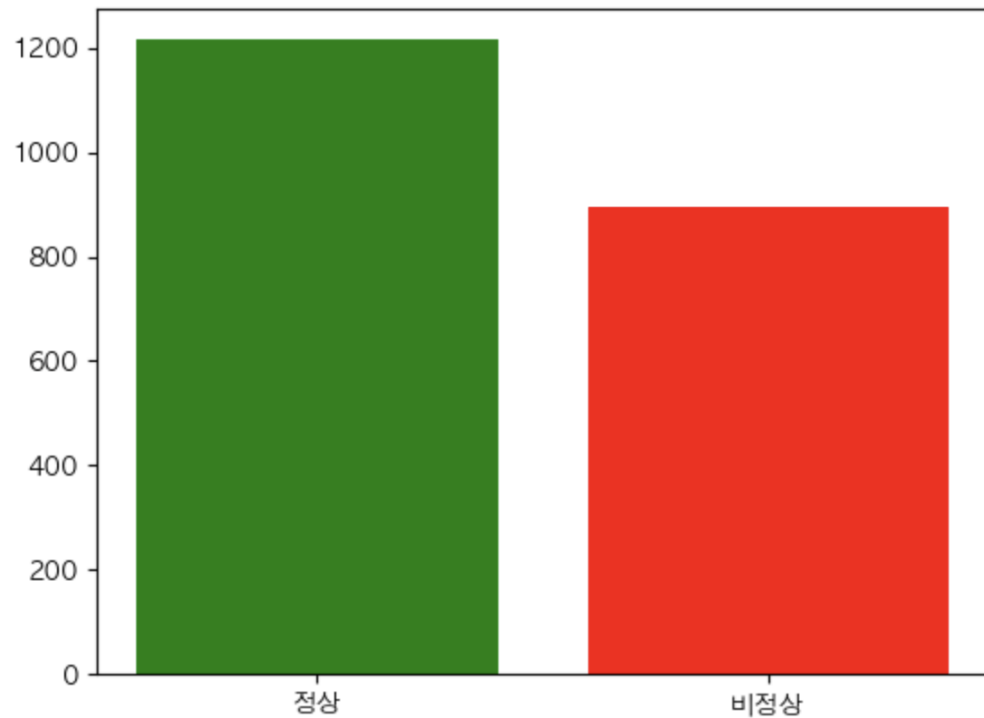
### LOG4SHELL DATASET

2111.rows x 4.columns

- 아이피 주소 : 네트워크 고유 식별자
- 접속 국가 : 현재 패킷을 보내는 국가
- 접속상대 : 현재 패킷을 받는 곳
- 정상/비정상 : 레이블 값 (4개 레이블 구성 → 정상:1, 비정상:0)



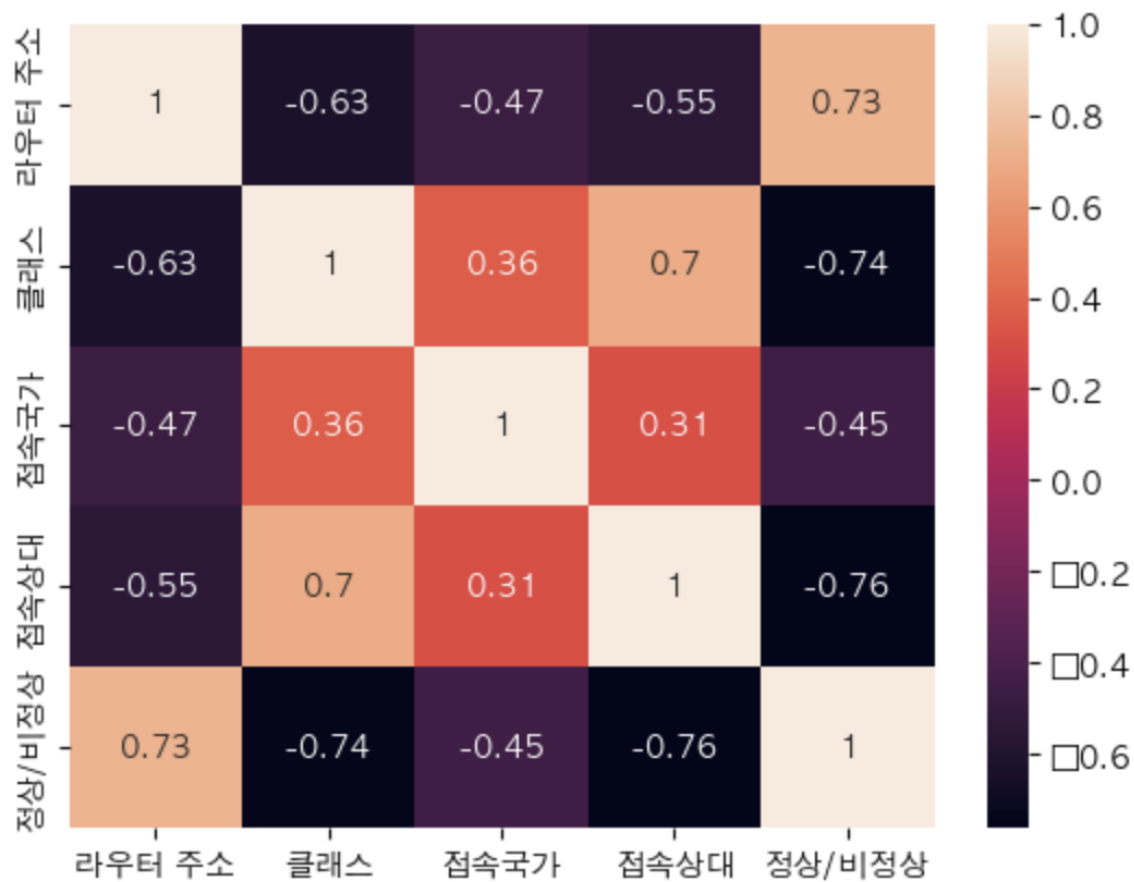
## Part 2 정상/비정상 데이터 및 접속상대 별 분포




## Part 2 최종 데이터 및 상관관계 분석

	라우터 주소	클래스	접속국가	접속상대	정상/비정상
0	157.245.108.1	B	IN	DIGITALOCEAN-ASN	1
1	194.195.246.1	C	DE	Linode, LLC	1
2	194.195.244.1	C	DE	Linode, LLC	1
3	194.195.244.1	C	DE	Linode, LLC	1
4	194.233.164.1	C	DE	Linode, LLC	1
...	...	...	...	...	...
1283	45.83.67.1	A	DE	Alpha Strike Labs GmbH	1
1284	45.83.67.1	A	DE	Alpha Strike Labs GmbH	1
1285	45.83.67.1	A	DE	Alpha Strike Labs GmbH	1
1286	45.83.67.1	A	DE	Alpha Strike Labs GmbH	1
1287	45.83.67.1	A	DE	Alpha Strike Labs GmbH	1

1288 rows × 5 columns





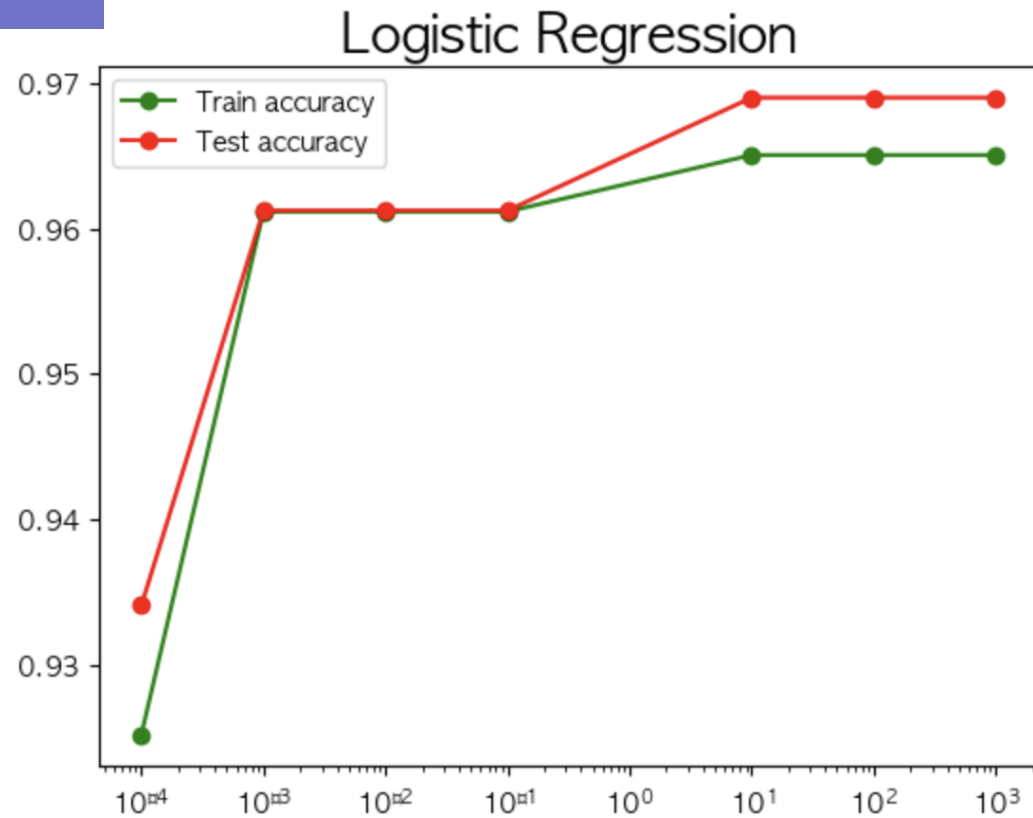


# Part 3

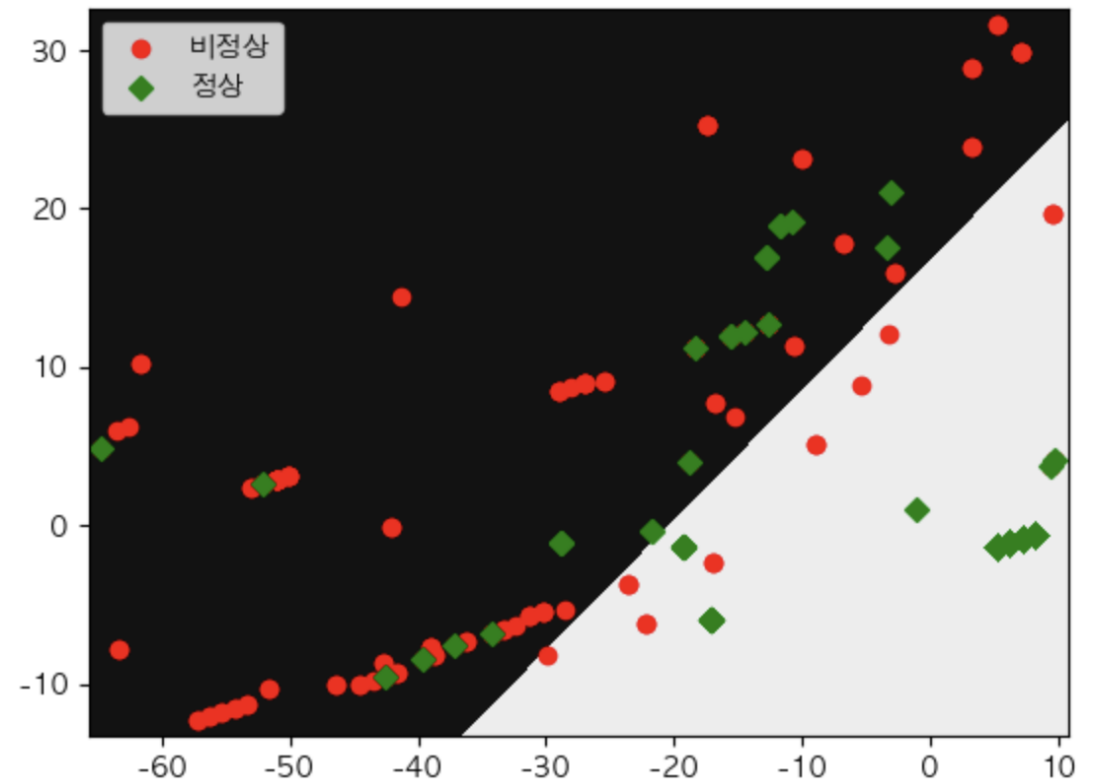
## 모델 개발 및 결과

## Part 3 Logistic Regression 그래프 파악

A



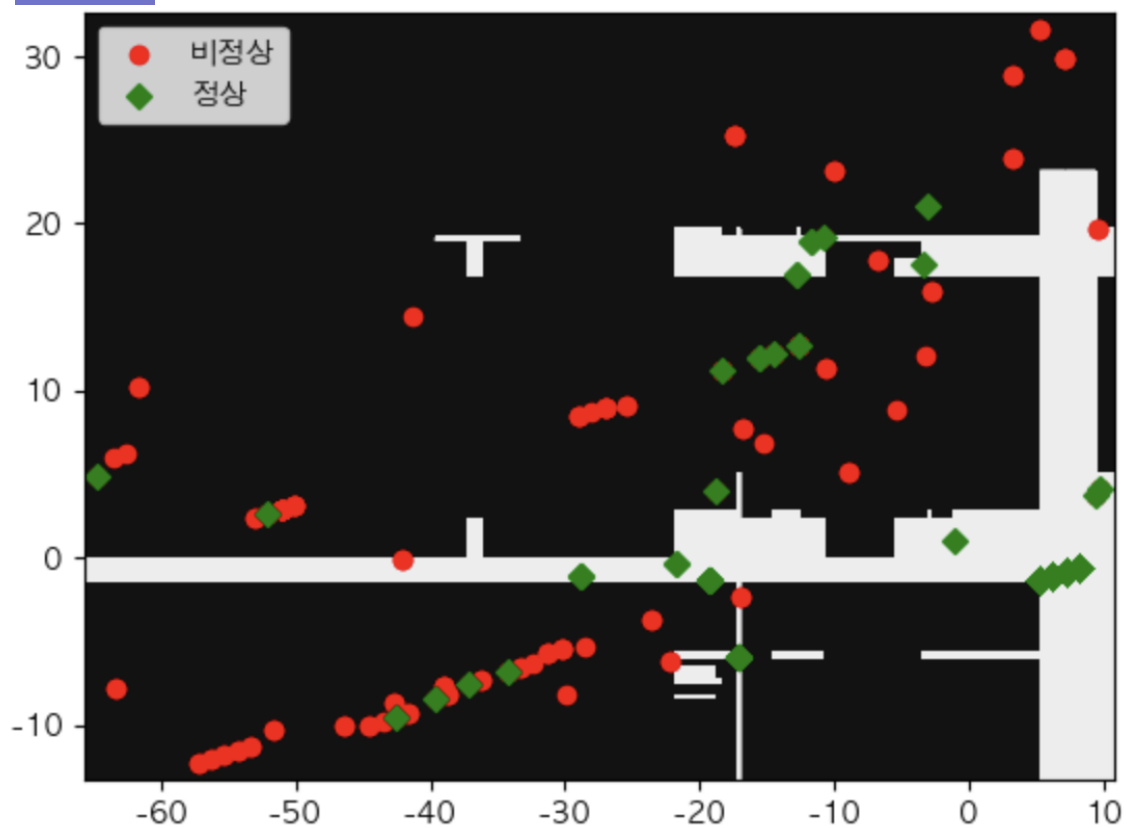
B



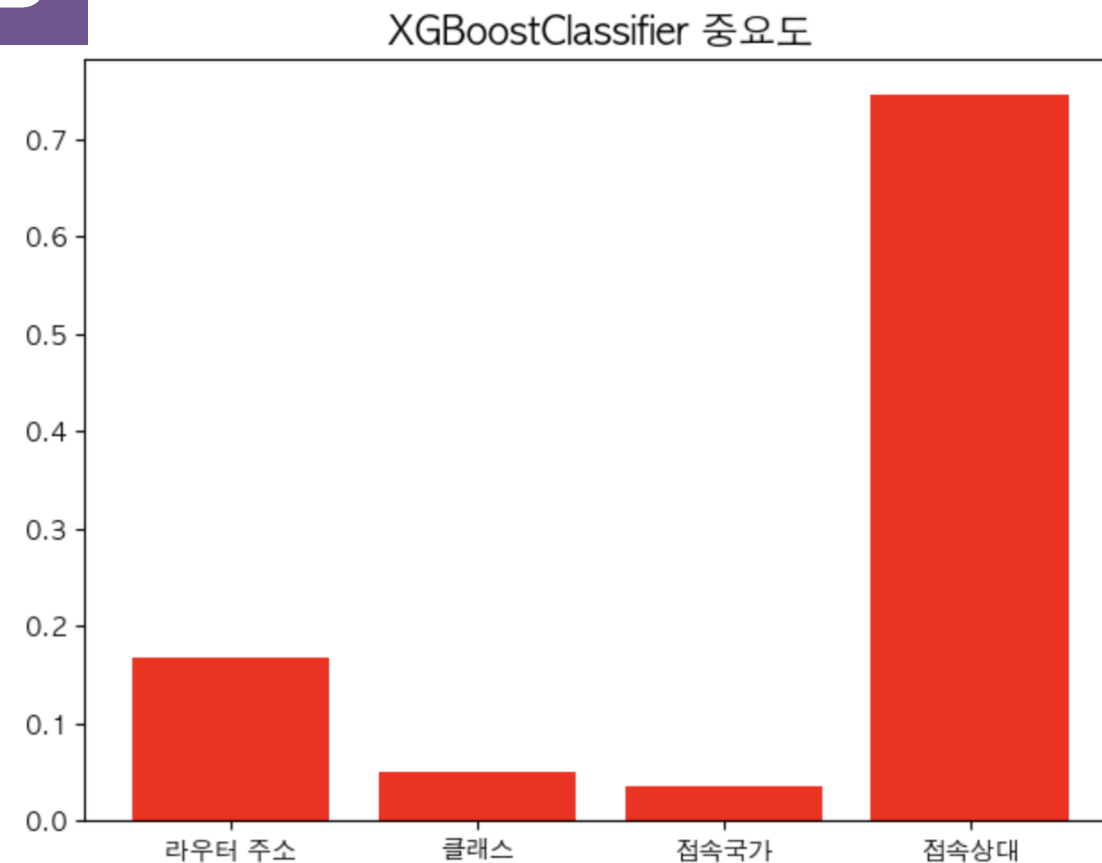


## Part 3 XGBoost 결정 그래프와 특성 중요도

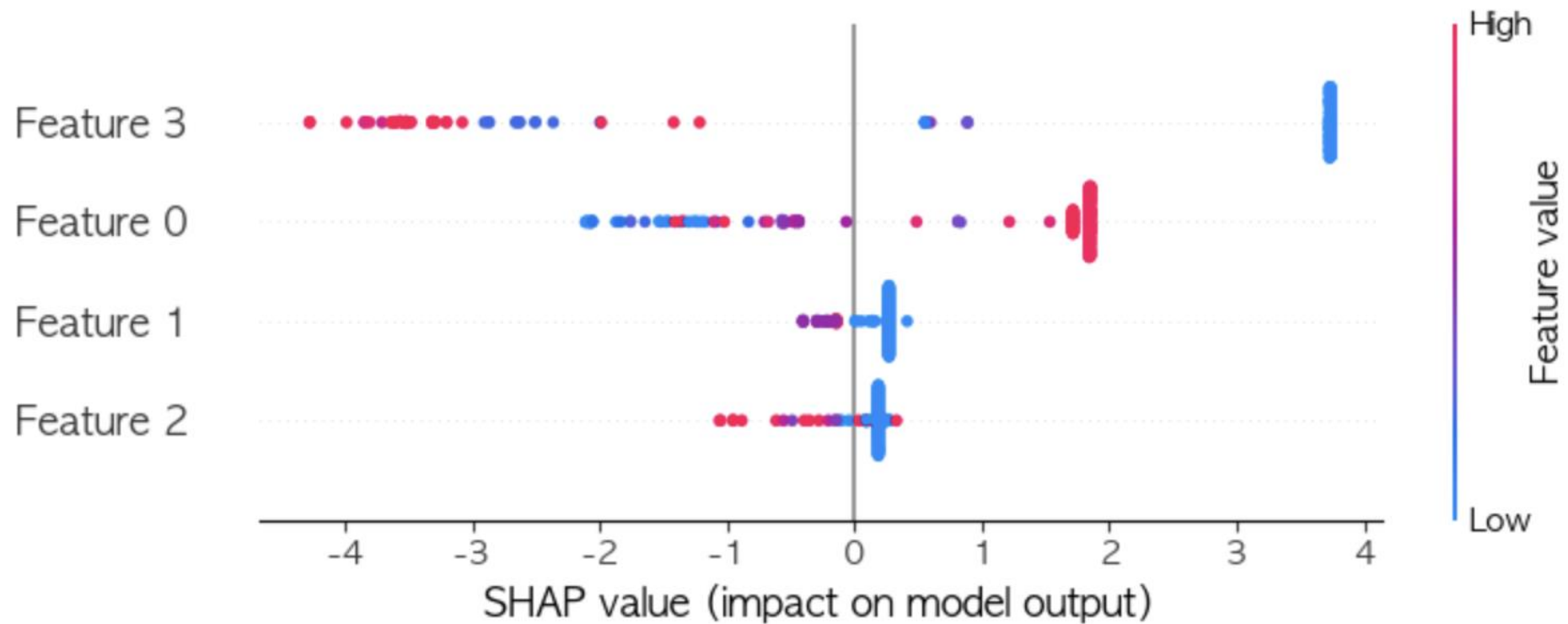
A



B

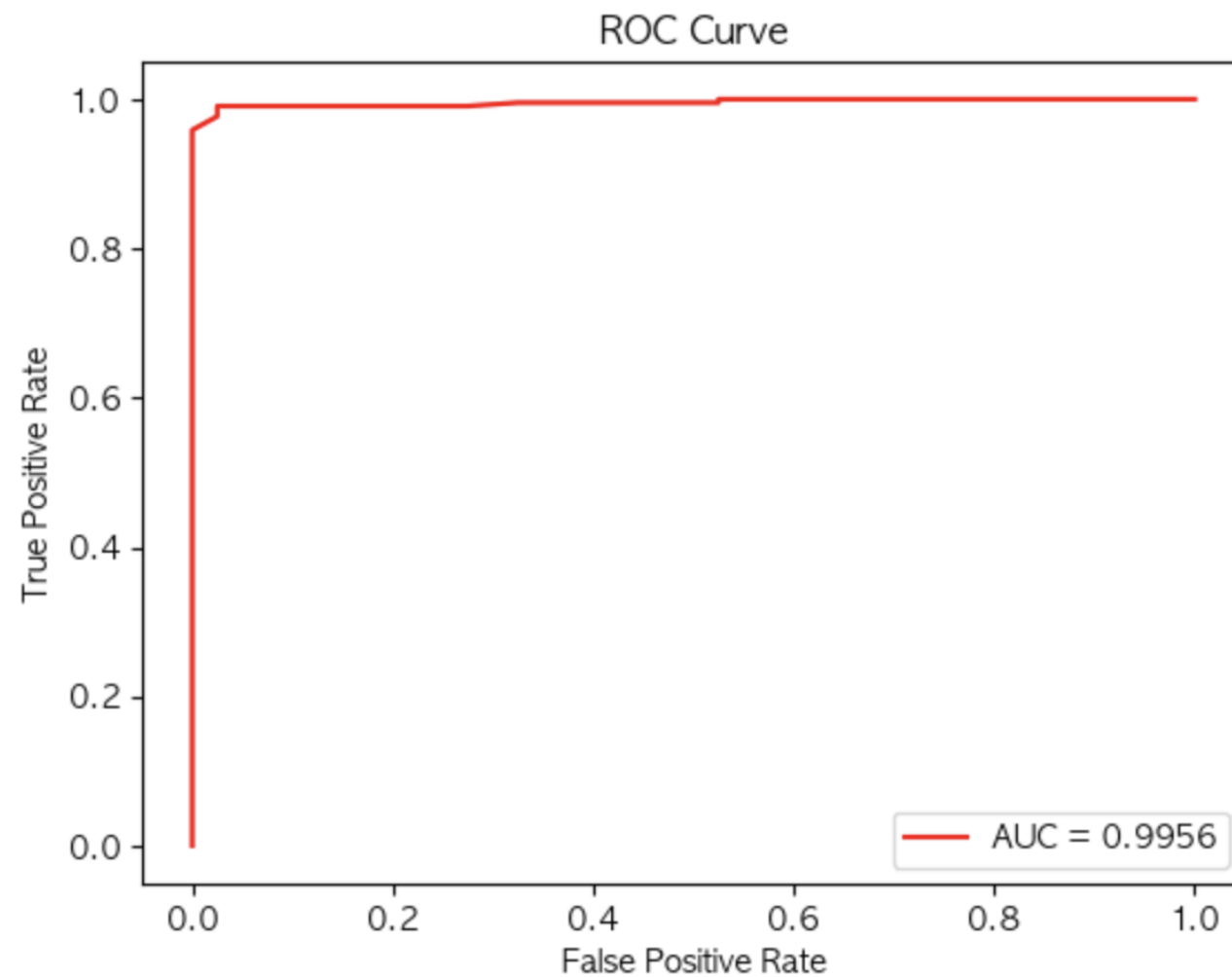


## Part 3 SHAP 그래프





## Part 3 ROC CURVE

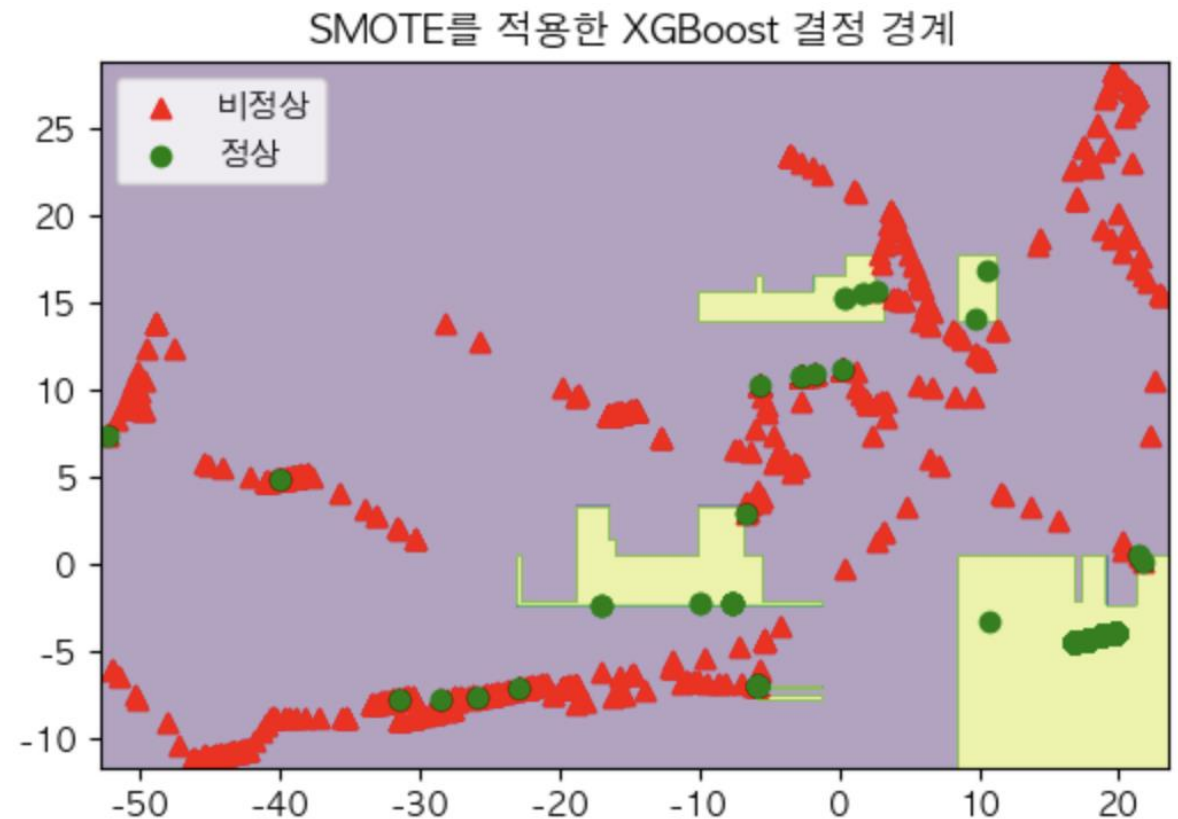
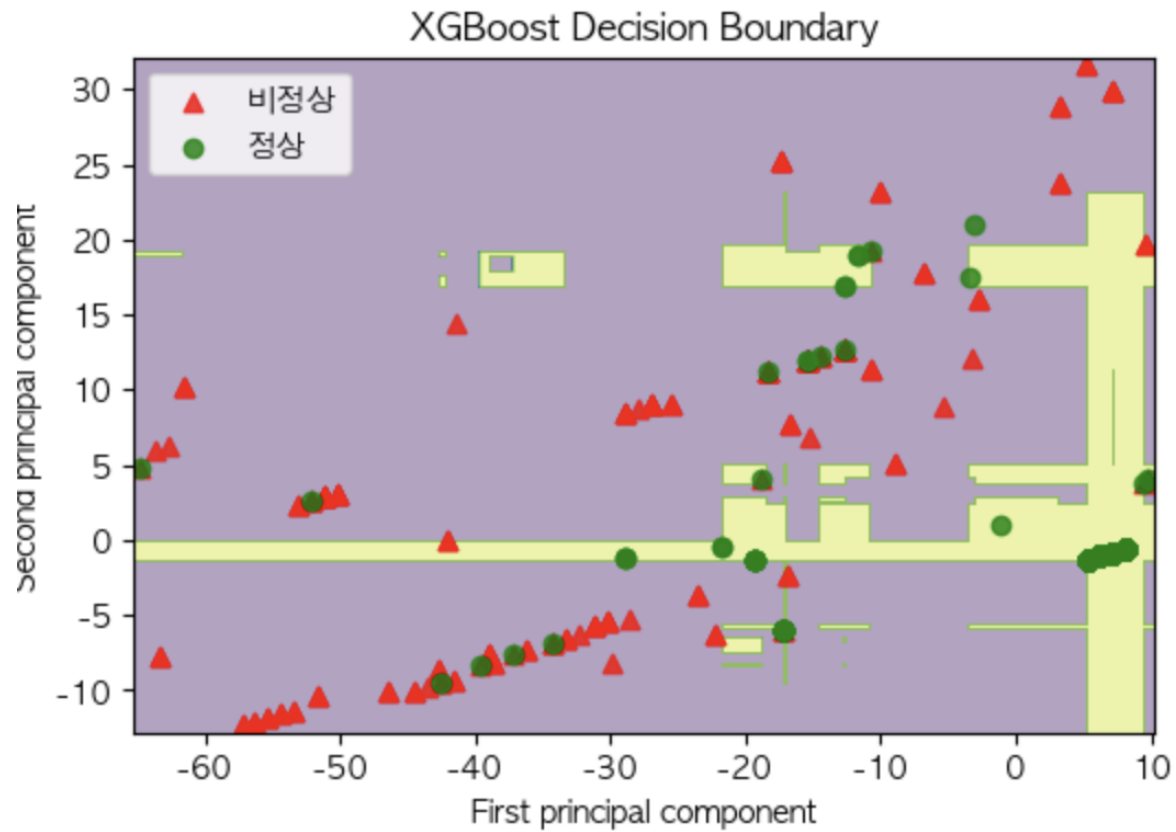


# Part 4

## 모델 한계점 및 극복 방향



## Part 4 기존 모델과 리샘플링 모델 차이 비교





# Part 5

## 결론 및 향후 계획

## Part 5 불균형 데이터 한계 극복

	Logistic Regression	XGBoost Classifier	SMOTE XGBoost Classifier
훈련 점수	0.965	0.988	0.989
테스트 점수	0.969	0.986	0.985
정밀도	0.9724	0.9851	0.9839
재현율	0.986	0.9942	0.993
F1 Score	0.9792	0.9896	0.9885



## Part 5 모델의 한계

---

데이터의 다양성과  
범용성 부족

모델의 결정경계  
불완전성

실시간 탐지  
성능 제한

설명 가능성  
부족

## Part 5 실무 적용 방안

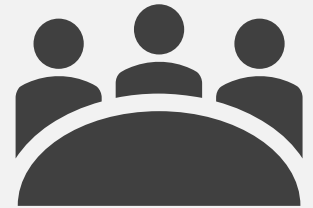
---



탐지 시스템 연계



실시간 운영 및 테스트  
배포



확장 가능성

The background features a vibrant, abstract design with flowing, wavy shapes in shades of blue, purple, and teal. A large, semi-transparent sphere with a blue-to-yellow gradient is positioned in the center-left. To its upper left, two smaller spheres of the same gradient are stacked vertically. The Korean text '감사합니다' is centered within the large sphere in a clean, white, sans-serif font.

감사합니다