

# 데이터에서 가치를 찾는 데이터 분석가, 허주혁입니다.

---

## Contact

---

Github

<https://github.com/hjuhyeok>

Phone

010-2476-5021

Email

wngur1205@naver.com

## ABOUT ME



이름 : 허주혁

번호 : 010-2476-5021

연구분야

- 불균형 데이터 분석
- 불량 예측 모델 개발

좌우명 : 나는 할 수 있다.

저는 제조업의 불량 예측과 품질 개선에 관심이 있습니다.  
특히 데이터를 통해 새로운 가치를 창출하는 것을 즐깁니다.

### 학력

2017.02 운천고등학교 졸업

2023.08 충북대학교 정보통계학과 졸업  
GPA : 3.98 / 4.5

2025.08 연세대학교 통계데이터사이언스학과 졸업  
GPA : 4.19 / 4.5

### 실무 경력

2022 이노포스트

- 보행안전지수 개발 및 시각화

2023,2025 연세대학교 데이터사이언스연구소

- 석/박사 대상 통계 상담 및 의사결정 지원

2024 연세대학교 바른ICT연구소

- 악성댓글 비용 시뮬레이터 개발

### 프로젝트

2022 천안시 교통사고 취약지역 도출

2023 맞춤형 추천을 제공하는 개인화 추천 시스템 모델 구축

2024 Tabular Data Generation Using Generative Models

### 논문

2025 ADBoost: Boosting with Tree-Based Hybrid-sampling for Imbalance Data

- Statistics and Computing 저널 투고

### 수상 경력

2017 충북대학교 성적우수 장학생

2022 천안시 데이터 기반 시각화 아이디어 공모전 우수상

### 프로그램 활용 능력

PYTHON	★	★	★	★	★
R	★	★	★	★	★
TABLEAU	★	★	★	★	☆
QGIS	★	★	★	★	☆
SQL	★	★	★	☆	☆
MINITAB	★	★	★	☆	☆

### 자격증 및 공인 성적

2020 워드프로세서, 사회조사분석사 2급

2021 데이터 분석 준전문가(ADsP)

2022 컴퓨터 활용능력 1급, 품질경영기사, SQL 개발자 (SQLD), 빅데이터 분석기사

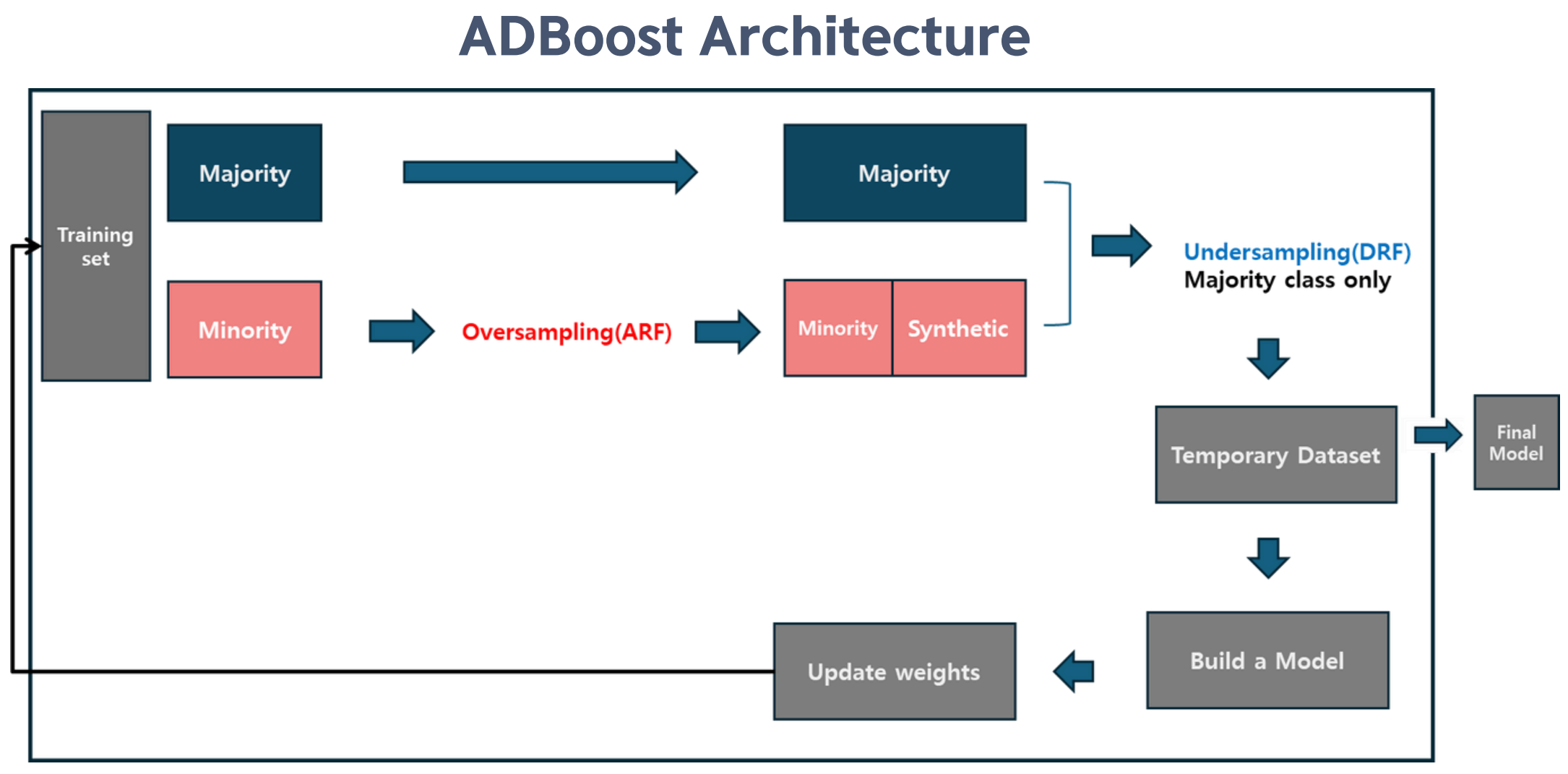
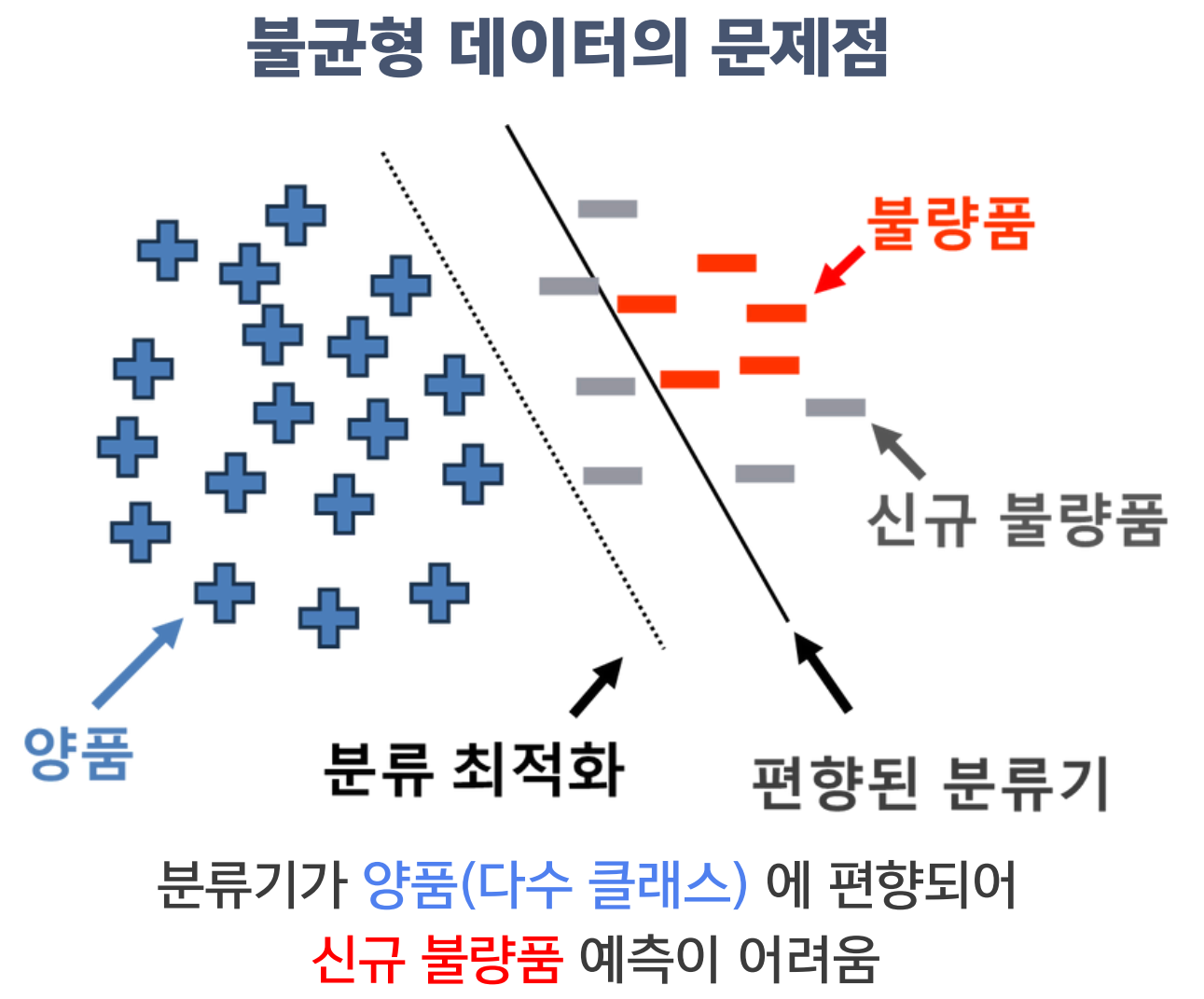
2024 TOEIC 785/990

2025 OPIC IH LEVEL

논문

# ADBoost: Boosting with Tree-based Hybrid-sampling for Imbalanced Data

국문 : 불균형 데이터 처리를 위한 결정트리 기반 하이브리드 샘플링 부스팅 기법



**현재 상태** Statistics and Computing 저널 투고

**기여도**



**논문 개요**

본 연구는 불균형 데이터 환경에서 소수 클래스 예측 성능을 향상시키는 것을 목표로 합니다. 이를 위해 Adversarial Random Forest 기반 오버샘플링과 Double Random Forest 기반 언더샘플링을 결합한 하이브리드 샘플링 기법을 제안하였습니다. 또한 이 기법을 Boosting과 통합하여 불량 예측 성능을 개선하는 ADBoost 알고리즘을 개발하였습니다.

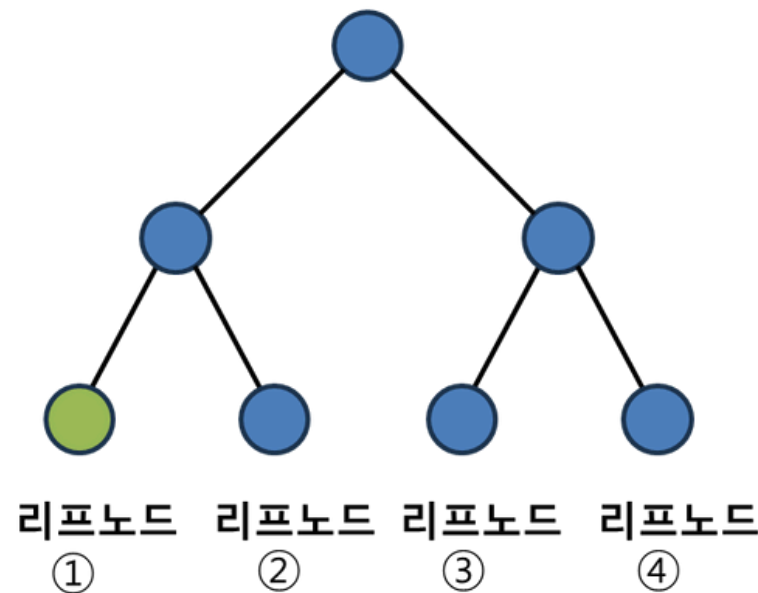
논문

# ADBoost: Boosting with Tree-based Hybrid-sampling for Imbalanced Data

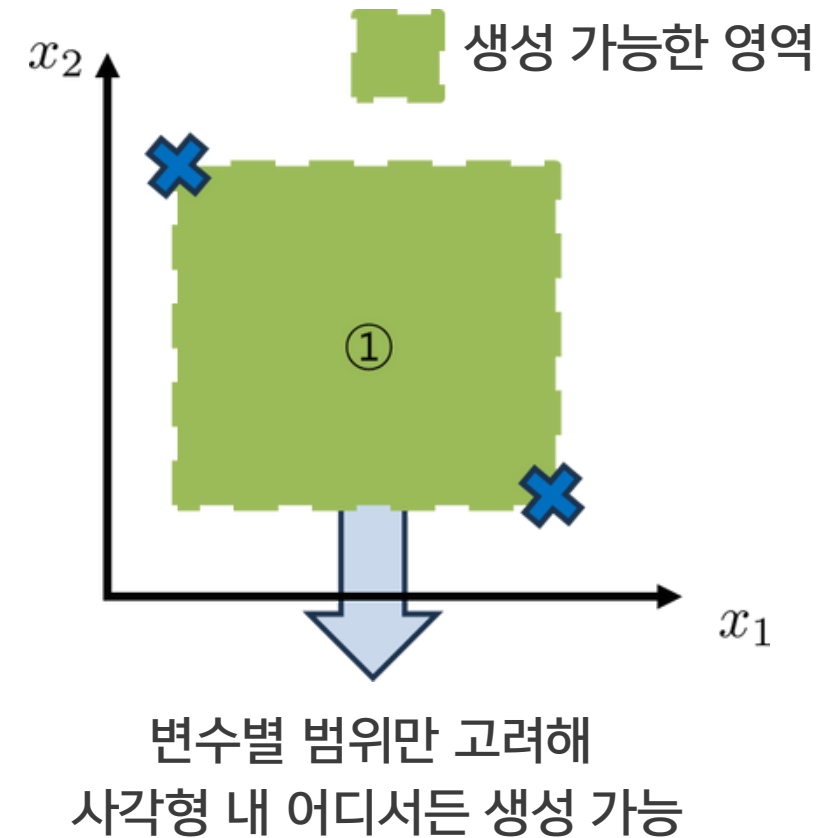
국문 : 불균형 데이터 처리를 위한 결정트리 기반 하이브리드 샘플링 부스팅 기법

## Adversarial Random Forest(ARF)

개념



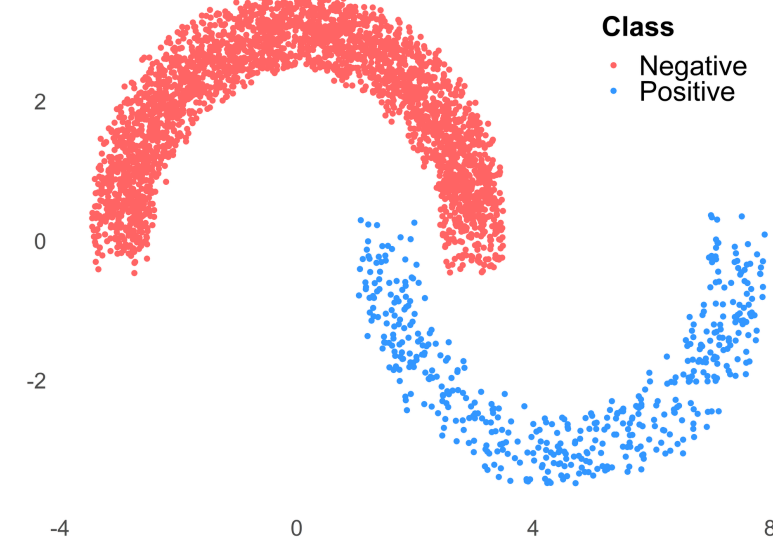
생성 방식



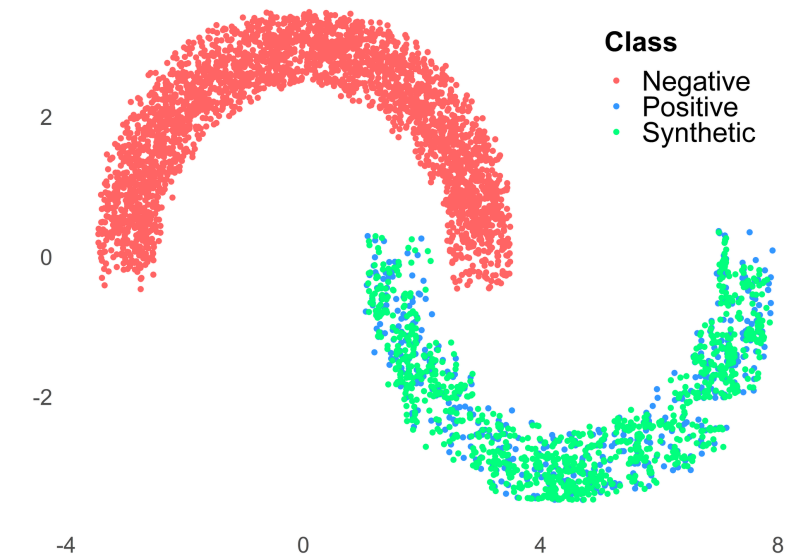
- ▶ 리프노드 내 변수별로 샘플링해 비선형적으로 데이터를 생성
- ▶ 고차원에서도 조건부 독립성을 활용해 복잡한 경계를 잘 반영

## Data Oversampling

Before



After



- ▶ 소수 클래스와 유사한 특성을 지닌 Synthetic Data(불량 데이터) 생성
- ▶ 불균형을 완화하여 모델이 불량 데이터 패턴을 학습할 수 있도록 지원

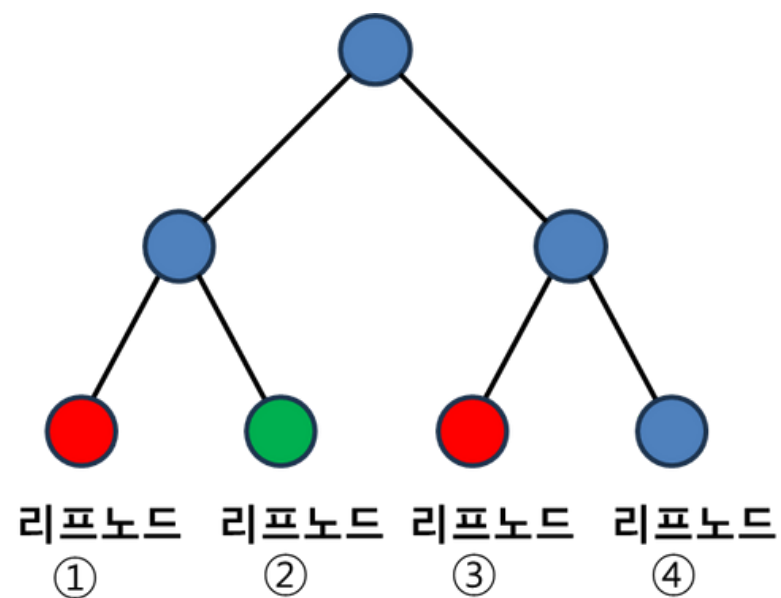
논문

# ADBoost: Boosting with Tree-based Hybrid-sampling for Imbalanced Data

국문 : 불균형 데이터 처리를 위한 결정트리 기반 하이브리드 샘플링 부스팅 기법

## Double Random Forest(DRF)

개념



제거 방식

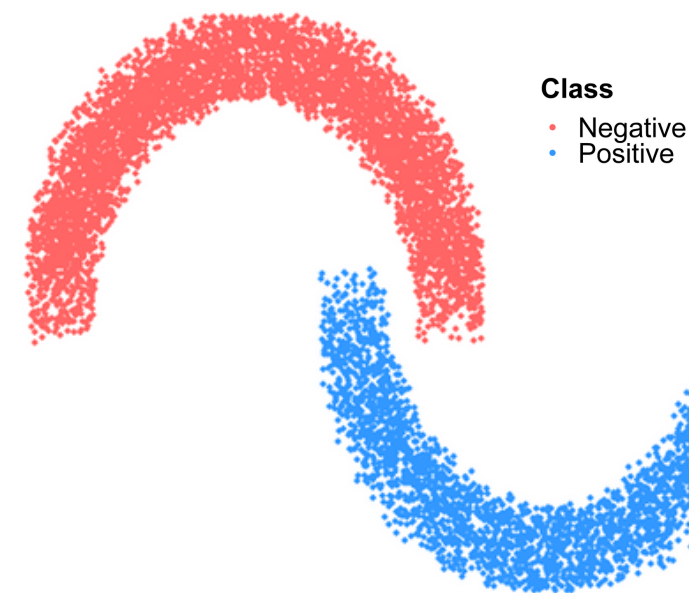
리프노드 번호	클래스 구성	유형
①	양품	제거 대상
②	양품, 불량품	보존
③	양품	제거 대상
④	불량품	보존

● ● : 데이터 보존  
● : 제거 대상

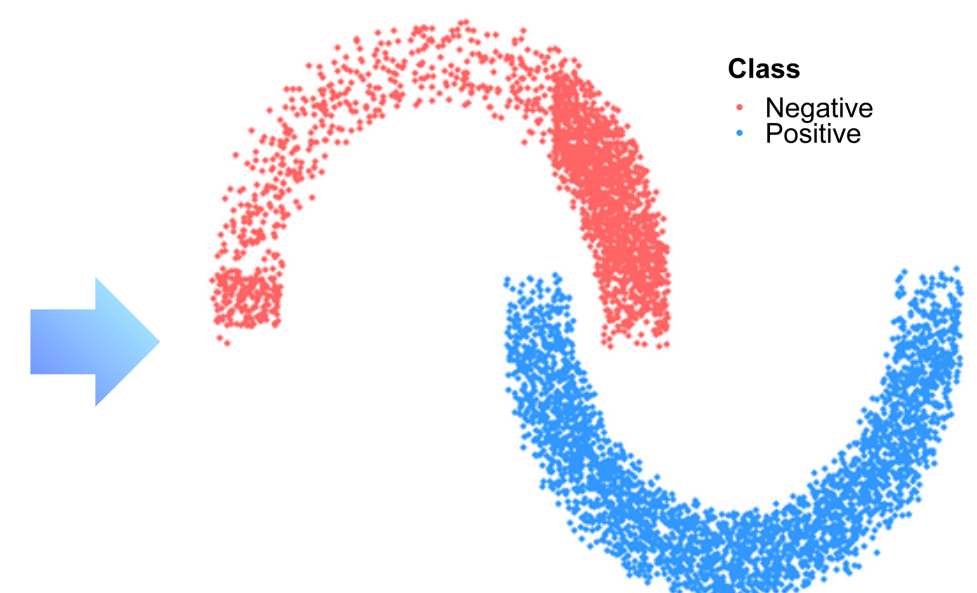
- ▶ 양품 데이터만 존재하는 리프노드를 제거
- ▶ 과적합 방지 및 학습 효율 향상

## Data Undersampling

Before



After



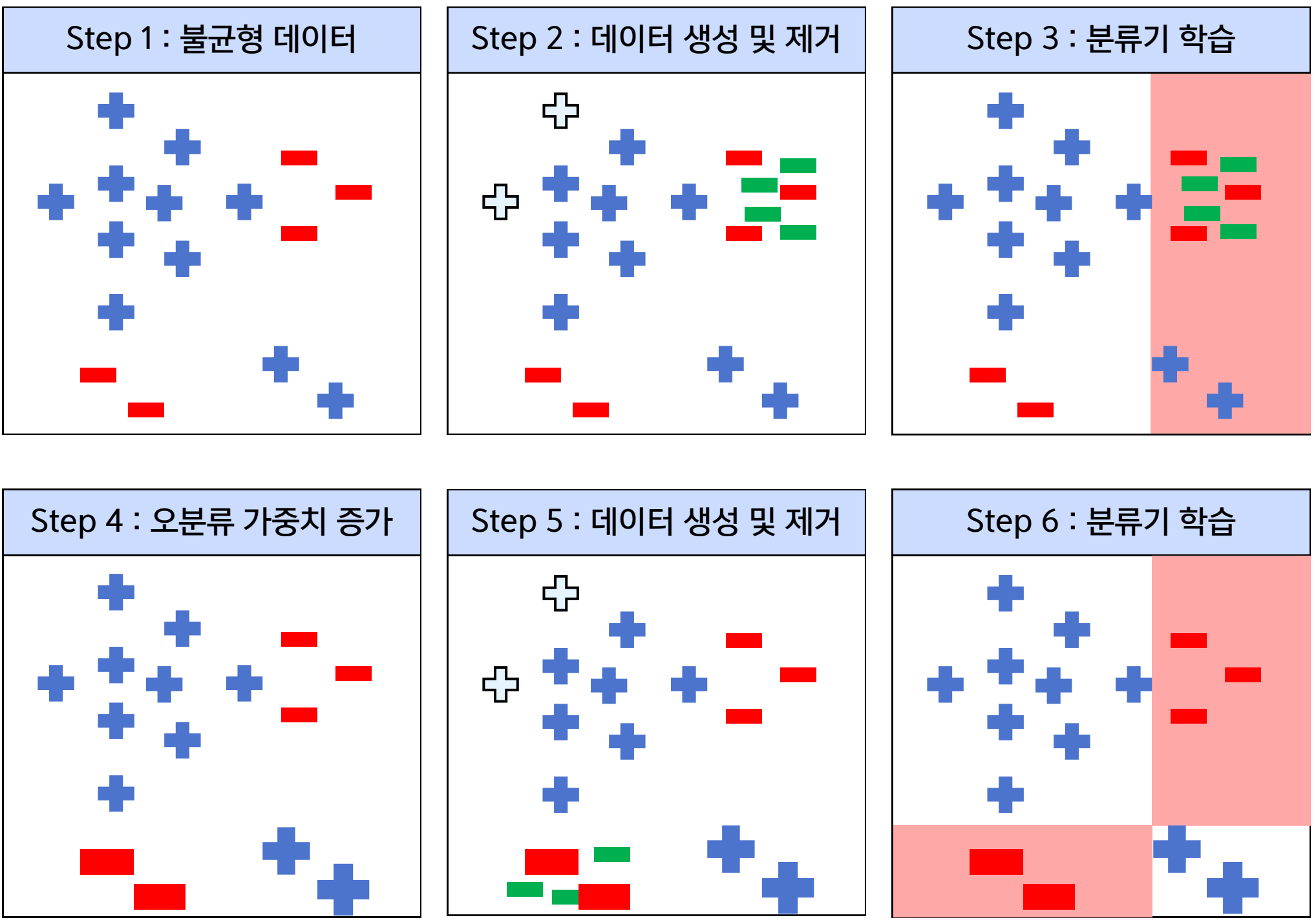
- ▶ 불필요한 양품 데이터 제거를 통한 경계 정보 강화
- ▶ 복잡한 경계를 반영해 불량 예측 성능 향상에 기여

논문

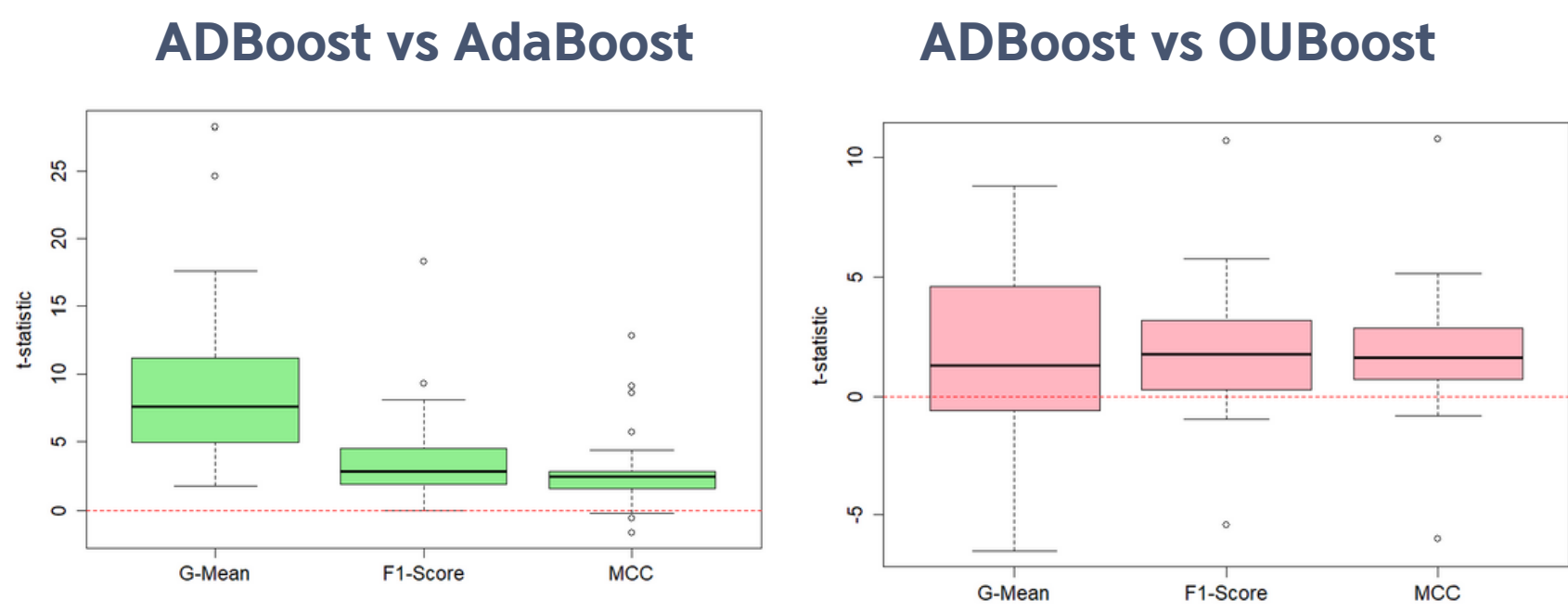
ADBoost: Boosting with Tree-based Hybrid-sampling for Imbalanced Data

국문 : 불균형 데이터 처리를 위한 결정트리 기반 하이브리드 샘플링 부스팅 기법

Boosting 학습 과정



ADBoost vs 기존 부스팅 모델 성능 비교



	AdaBoost	SMOTEBoost	RUSBoost	OUBoost	ADBoost
G-Mean	0.6949	0.7316	0.7104	0.8049	<b>0.8740</b>
F-Score	0.5772	0.6064	0.5927	0.6082	<b>0.6491</b>
MCC	0.5624	0.5847	0.5737	0.5794	<b>0.6276</b>

- ▶ ARF : 의미 있는 불량 데이터 생성 → 데이터 다양성 확보
- ▶ DRF : 양품 데이터 제거 → 경계 학습 강화
- ▶ ADBoost : 클래스 간 균형 있는 분류 성능 확보
- ▶ 실험 결과 : 클래스 균형 분류 달성, 주요 지표(G-Mean, F-Score, MCC)에서 **최고 성능 기록**

# 읽어주셔서 감사합니다!

---

## Contact

---

Github

<https://github.com/hjuhyeok>

Phone

010-2476-5021

Email

[wngur1205@naver.com](mailto:wngur1205@naver.com)