

대구로 어플



이상거래탐지 모델을 통한 소비자 불만 사전예측

제 6회 대구 빅데이터 분석 경진대회

SPADA

서동옥, 이현정, 최준혁

CONTENTS

01

문제 인식

02

소비자 불만 정의 및 가정

03

사용 데이터

04

모델링

05

성능 (학습 데이터)

06

성능 (테스트 데이터)

07

결과

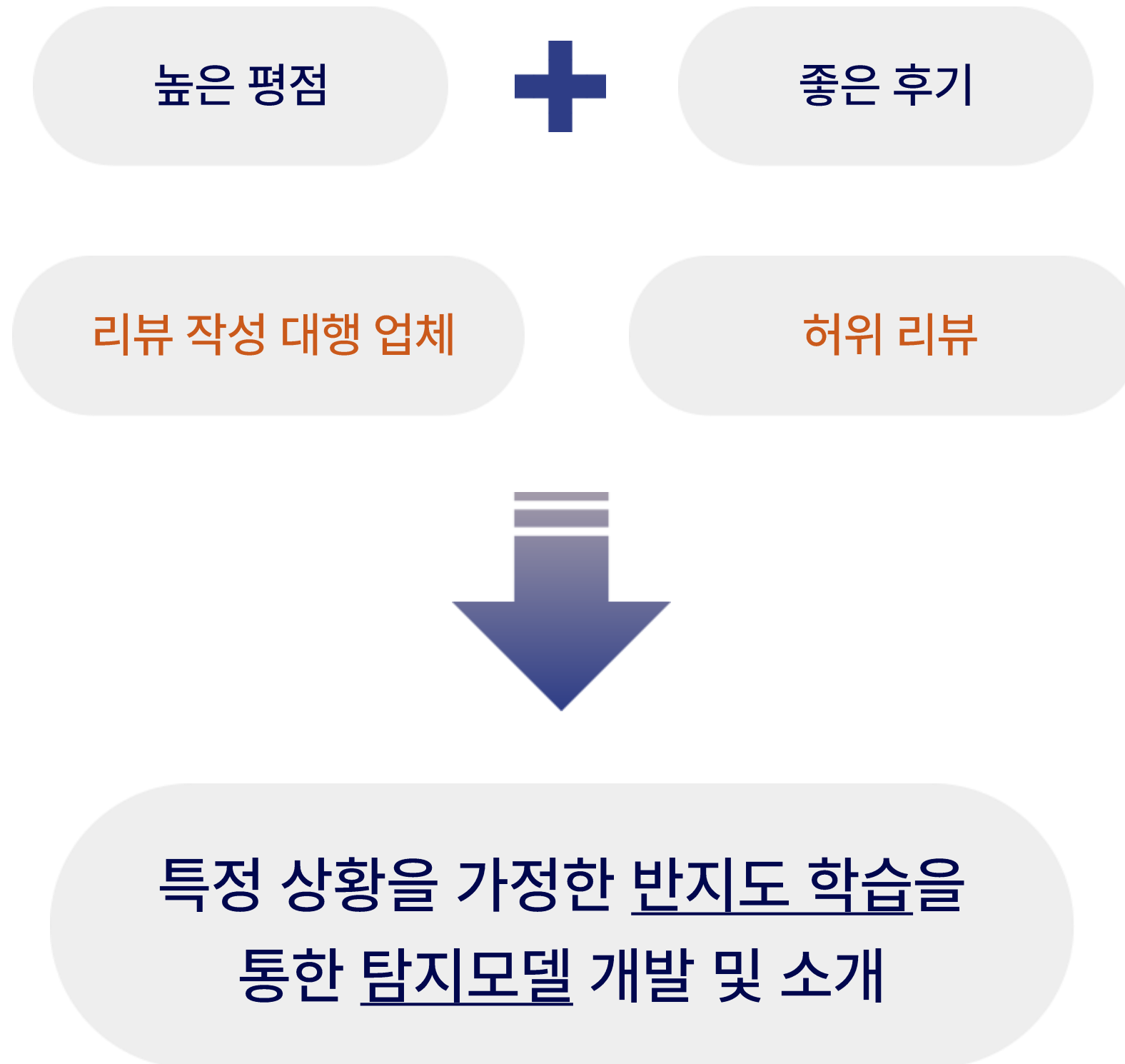
08

결론

09

Appendix

01 문제 인식



맛집 플랫폼 리뷰 1건+트래픽

맛집 플랫폼 리뷰 10건+트래픽

맛집 플랫폼 리뷰 30건 +트래픽

2.맛집 플랫폼 추천(좋아요)상품

리뷰 추천 25건 - 2만원

리뷰 추천 50건 -4만원

리뷰 추천 100건- 9만원

3.맛집 플랫폼 저장 및 즐겨찾기 서비스

맛집 플랫폼 저장 및 즐겨찾기 서비스 25건-2만원

맛집 플랫폼 저장 및 즐겨찾기 서비스 50건-4만원

* 자료 출처 : 크몽

02 소비자 불만 정의 및 가정



case1

신고하기

언제나 맛나요 ㅋㅋㅋ



2024.06.18



신고하기

피자랑 치킨 항상 너무 맛있습니다
감사합니다 ㅎㅎ

특정 사용자가 하나의 가게에만 리뷰작성,
모두 **높은** 별점으로 작성

=> 불편사항 : 예상과 다른 맛으로 인한 소비자 불만,
정상적인 가게의 매출 타격



case2

신고하기

피자맛이 어색해서 까봤더니 토마토소스가 없어요 빵위에 붉은
색은 페퍼로니 기름입니다 여기는 원래 피자에 토마토소스 안
바르나요?



특정 사용자가 하나의 가게에만 리뷰작성,
모두 **낮은** 별점으로 작성

=> 불편사항 : 정상적인 가게의 매출 타격

03 사용 데이터

- 데이터 형식

리뷰 데이터

'SHOP_ID', 'CUST_ID', 'REVIEW_TIME', 'REVIEW_RANK'



주문 데이터

'SHOP_ID', 'ORDER_DT', 'CUST_ID', 'ORDER_ST', 'COMP_TIME', 'SHOP_CONF_TIME', 'CANCEL_DT', 'PACK_YN'



결제 데이터

'SHOP_ID', 'CUST_ID', 'PAY_GBN', 'TOT_AMT', 'DISC_AMT', 'DEST_PAY_GBN', 'CARD_PAY_GBN', 'CARD_APP_TIME',
'CARD_APP_AMT', 'PACK_DISC_AMT'



- 데이터 특징

라벨 데이터

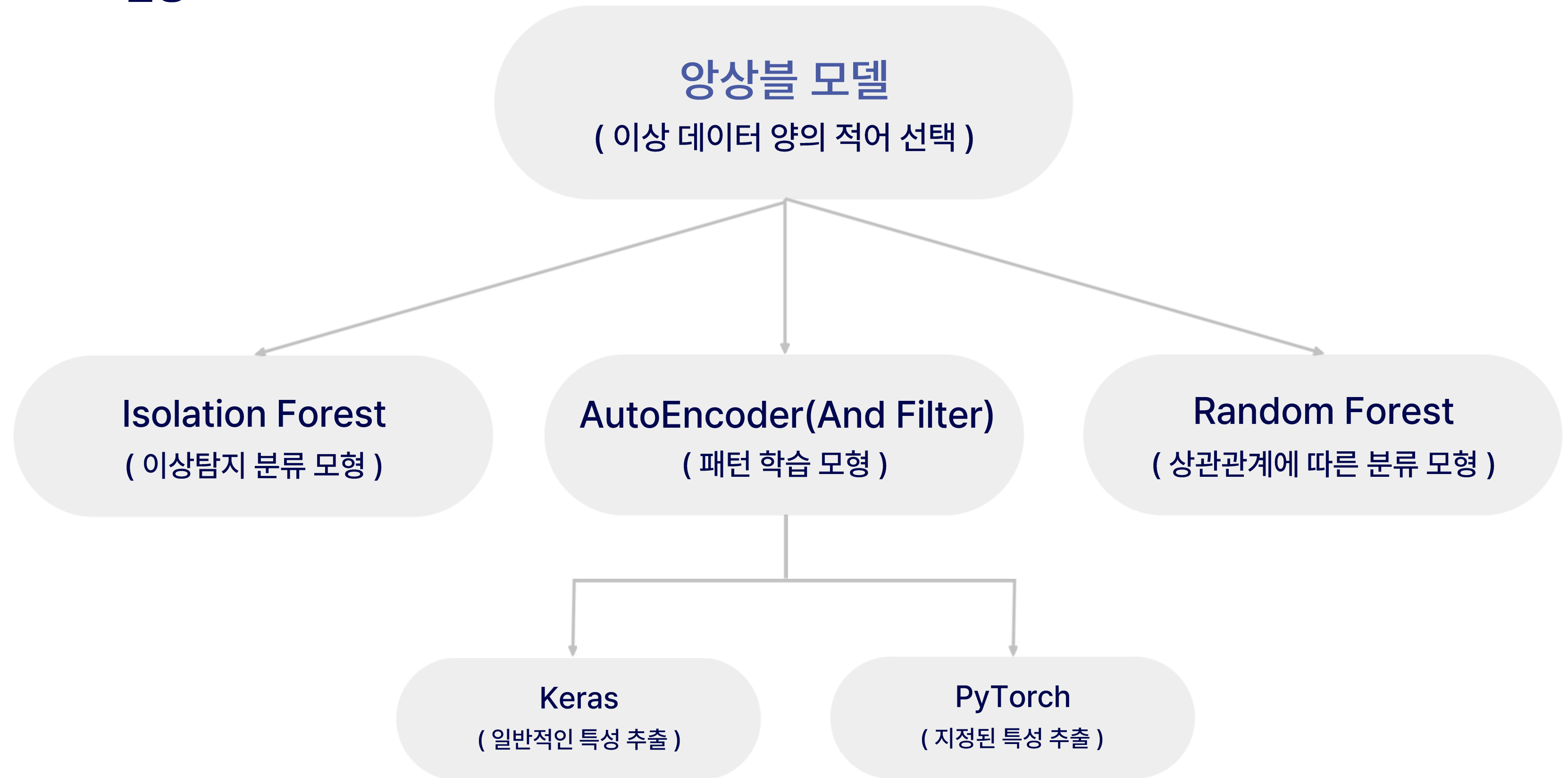
=> 중구의 대구로 데이터 (학습용)

비라벨 데이터

=> 대구 전체의 대구로 데이터 (테스트용)

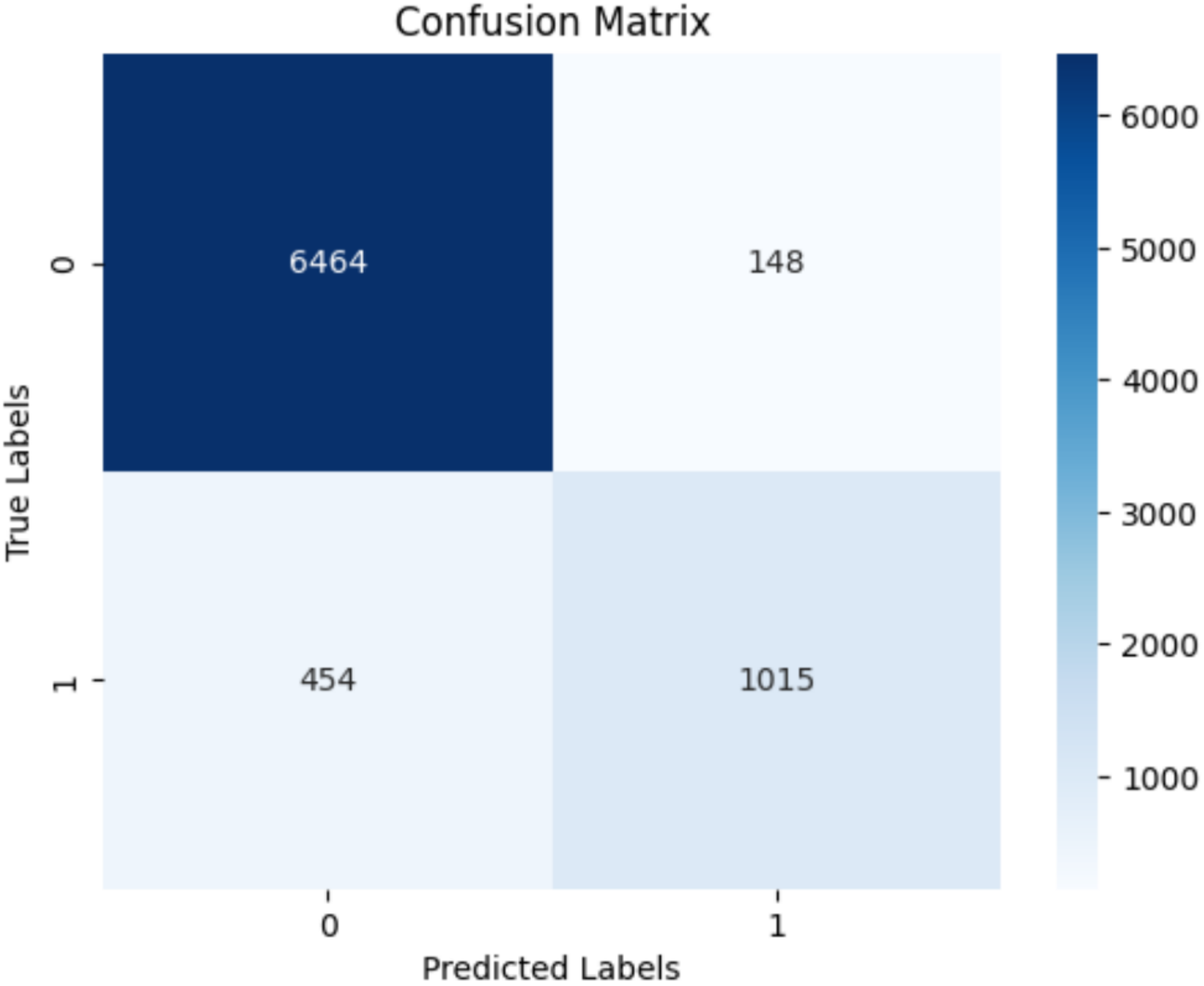
* 제공: 대구 빅데이터 활용센터, 기간 (2022 ~ 2023)

04 모델링



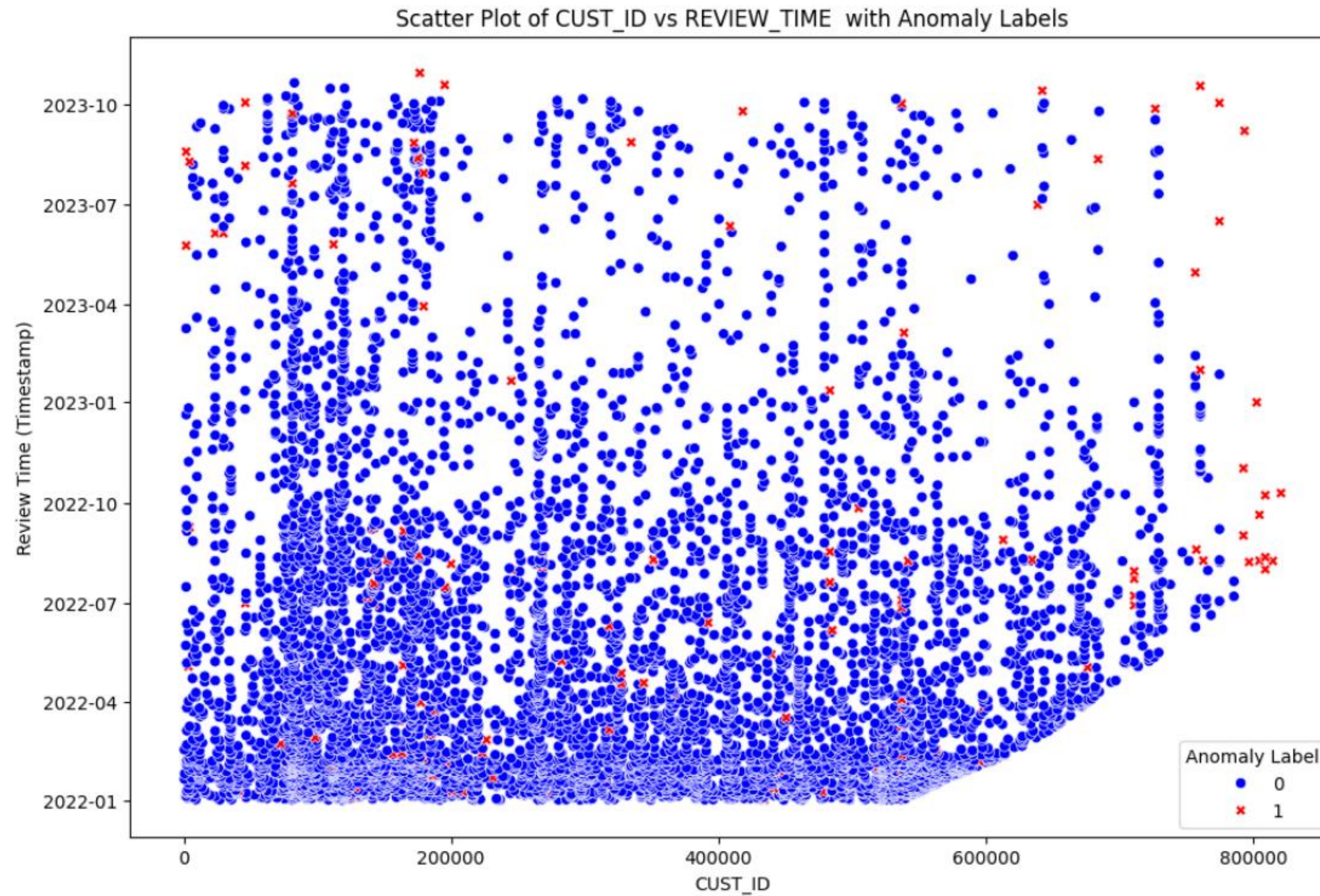
05 성능 (학습 데이터)

	Is
Accuracy	
Recall	
F1 Score	



Ensemble
92.55%
69.09%
77.13%

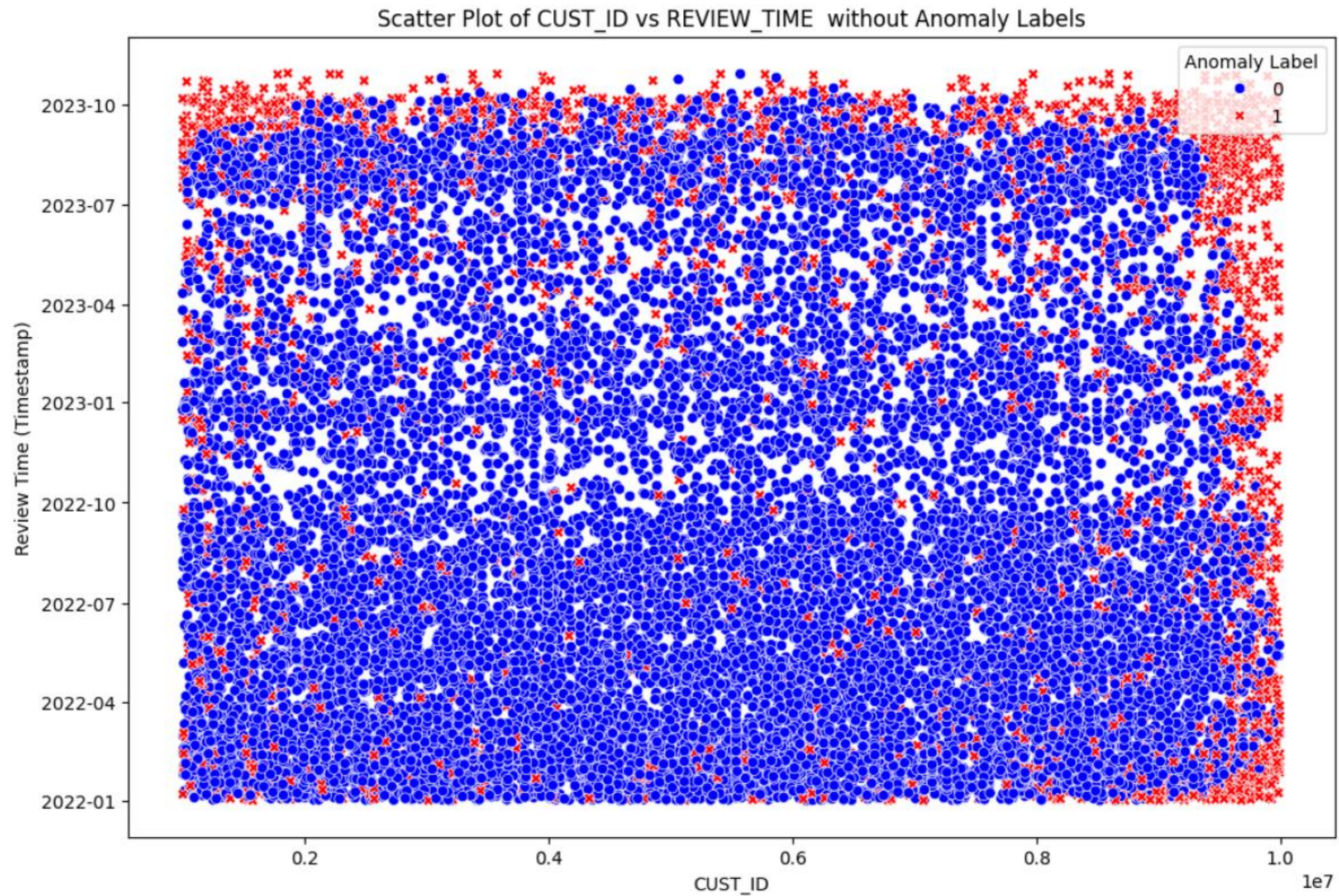
05 성능 (학습 데이터)



정상 스캐터링 (3)



06 성능 (테스트 데이터)



07 결과

허위 리뷰 대상자의 리뷰 데이터	해당 사용자 주문 데이터	해당 사용자 결제 데이터
-------------------	---------------	---------------

한 가맹점에서만 주문
=> 모델링에서 이상데이터로 판단

이상거래로 판단된 고객의 주문데이터를
확인해보니 취소건과 한 가맹점에서만
주문했음을 확인

거래금액 확인
=> 낮은 금액대의 주문 후 리뷰 작성하였음을
정성적으로 평가 가능

▶1 payment_df.loc[payment_df["CUST_ID"]== 9485582]

↕

	SHOP_ID	CUST_ID	PAY_GBN	TOT_AMT	DISC_AMT	DEST_PAY_GBN	CARD_PAY_GBN	CARD_APP_TIME	CARD_APP_AMT	PACK_DISC_AMT
16422	9380485	9485582	4.0	0.035070	0.000000	2.0	0.0	2022-02-12 08:34:06	0.035070	0.0
16423	9380485	9485582	4.0	0.035070	0.069926	2.0	0.0	2022-01-03 09:33:41	0.029409	0.0
16424	9380485	9485582	4.0	0.035070	0.069926	2.0	1.0	2022-01-16 10:39:43	0.029409	0.0
16425	9380485	9485582	4.0	0.035070	0.069926	2.0	1.0	2022-01-16 10:42:30	0.029409	0.0
16426	9380485	9485582	4.0	0.035070	0.022772	2.0	0.0	2022-06-01 08:38:06	0.033226	0.0
22361	9602334	9485582	4.0	0.038076	0.049505	2.0	0.0	2022-02-01 18:35:01	0.034068	0.0

08 결론

시사점

허위 리뷰에 대한 1차적 필터링 가능

패턴을 통한 리뷰의 건전성 확보

안좋은 결과를 가진 모델이 있으나
앙상블 모델에는 사용가능

1

2

3

보완점

리뷰데이터와
결제/주문 데이터의 싱크 불일치

데이터 라벨링의 명확한 기준

양질의 데이터를 통한 복잡한 모델링 필요

Insight

허위 리뷰로 의심되는 리뷰어들을 1차 필터링하는데 성공,
이상탐지 모델을 통해 리뷰 및 거래 데이터에서 비정상적인 패턴을 보이는 사용자를 효과적으로 탐지 가능

END of Presentation

대구로 어플
이상거래탐지 모델을 통한 소비자 불만 사전예측

SPADA | 서동욱, 이현정, 최준혁

문의 : hcan1445@gmail.com

09 Appendix – 자연어 처리(리뷰)

- 리뷰 내용 반출 불가
- '데이터베이스(DB)권 ‘ 침해 우려
- 대구로 앱 크롤링 허가 요청

이 사람과 주고받은 메일

메일 검색



상세 ▾

☐ ▾ 읽음 삭제 스팸차단 | 답장 전체답장 전달 | 이동 ▾ 더보기 ▾



☐ ☆ ✉ 서동욱 문의) 빅데이터 분석 경진대회 SPADA팀 대구로 크롤링 및 추가데이터 문의 🔍 ↗ 08.07 09:00

사용하려는 방법은 다음과 같습니다.
GET 방식 예제: <https://review.daeguro.co.kr:45008/v5/review?projectType=DELIVERY&jobGbn=2&shopCode=4862&pageNum=1&pageCnt=20&photoYn=N>

사용할 응답 결과 예시

```
{
  "orderNo": "2668168", --> 라벨 인코딩을 통한 마스킹 (주문 식별번호)
  "shopCode": "4862", --> 라벨 인코딩을 통한 마스킹 (가게 식별코드)
  "custCode": "347348", --> 라벨 인코딩을 통한 마스킹 (고객 식별코드)
  "custNickName": "slo***", --> 라벨 인코딩을 통한 마스킹 (고객 닉네임)
  "myReviewYn": "N",
  "content": "직장사람들이 맛있대요 ㅎㅎ",
  "starScore": 5,
  "orderMenu": "(NEW)동마카롱 6종선물세트, 스페셜티급 아메리카노",
  "insertDate": "2022-07-26 19:19:20",
  "visibleGbn": "A",
  "answerContent": "",
  "answerVisibleGbn": "",
  "answerInsertDate": "",
  "images": [
    "https://review.daeguro.co.kr:45008/review-images/4862/2668168_01_20220726191919.jpg"
  ],
  "isUse": "Y",
  "isUpdateable": "N",
  "blindStandDt": "",
  "blindEndDt": "",
  "blindType": ""
}
```


09 Appendix – 결제, 주문데이터

- 리뷰, 결제, 주문 데이터 길이 불균형

⇒ 리뷰에서 발견된 주문을 주문 또는 결제 데이터에서
검색하면 개수가 다름

⇒ 동시에적으로 사용불가 -> 지표로서 활용

⇒ 정성 평가용 데이터