

중고거래 플랫폼 내

가격 예측, 데이터 라벨링, 판매 확률 예측

KREAM, 중고나라, 당근마켓 플랫폼을 중심으로

2020131302 하진우

2021320322 윤민서

A PRESENTATION ABOUT

- 1 가격 예측
- 2 데이터 라벨링
- 3 판매 여부 예측

OUR OBJECTIVES

1. 분석 목표
2. 가설
3. 데이터 수집
4. 데이터 분석
5. 에러 사항
6. 결론

분석 목표

중고 거래 플랫폼 내 업로드된 아이템의 구매 가격을 예측

KREAM

HOME STYLE⁺ SHOP Q

A light gray hoodie with a small Supreme logo patch on the chest.

즉시 구매가
450,000원

Supreme Box Logo Hooded Sweatshirt Ash Grey - 23FW
슈프림 박스로고 후드 스웨트셔츠 애쉬 그레이 - 23FW

모든 사이즈

최근 거래가 499,000원 ▼90,000 (-15.3%)	판매가 \$168 (약 221,600원)	모델번호 -	출시일 23/12/07	대표 색상 Ash Grey
--	------------------------------	-----------	-----------------	-------------------

구매 450,000원 즉시 구매가 판매 620,000원 즉시 판매가

관심상품 1,930

추가 혜택 더보기

포인트 계좌 간편결제 시 1% 적립

결제 네이버페이 포인트 최대 2만원 지급! 외 3건

가설

1. 중고 아이템의 상세정보 [브랜드, 관심 수, 리뷰 수, 과거 거래 가격..]
바탕으로 구매 가격을 예측할 수 있을 것이다.
2. 1의 가격 예측 모델을 활용하여 거래 횟수가 적은 아이템의 구매 가격을
예측할 수 있을 것이다.

가설

소규모 데이터를 수집하여 구매 가격 예측이 가능한지 사전 검증 진행

- 중고거래 플랫폼 KREAM에서 8가지 상품군의 인기 top100 아이템 800개 크롤링
- 상품 상세 정보 [거래 횟수, 브랜드, 관심 수, 과거 거래 가격(4건)..] 바탕으로 현재 구매 가격 예측 Task 진행
- Task 결과 바탕으로 구매 가격 예측 가능 판단

Random Forest

MEAN SQUARED ERROR: 4612943907.14

ROOT MEAN SQUARED ERROR: 67918.66

MEAN ABSOLUTE ERROR: 21128.0

R-SQUARED: 0.97

LINEAR REGRESSION

MEAN SQUARED ERROR: 3624270249.57

ROOT MEAN SQUARED ERROR: 60201.91

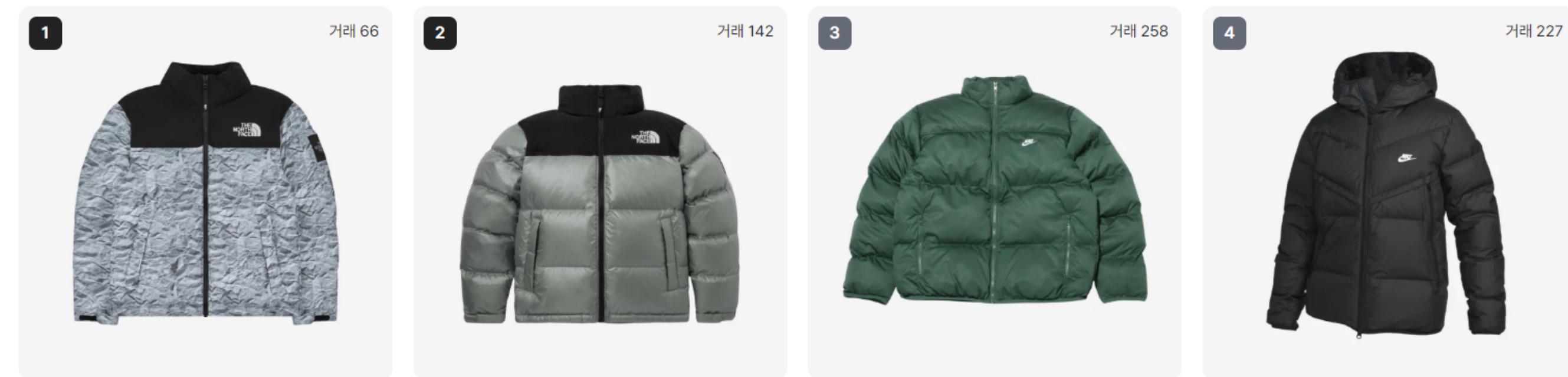
MEAN ABSOLUTE ERROR: 23757.12831365042

R-SQUARED: 0.97

데이터 수집

- 중고 거래 플랫폼 KREAM에서 10660개의 아이템을 크롤링

남성 패딩 인기 순위



The North Face

The North Face White Label Novelty Nuptse Down Jacket P White
노스페이스 화이트 라벨 노벨티 눕시 다운 자켓 프린트 화이트

181,000원

즉시 구매가

493 2

The North Face

The North Face White Label Novelty Nuptse Down Jacket Avocado
노스페이스 화이트 라벨 노벨티 눕시 다운 자켓 아보카도

빠른배송

343,000원

즉시 구매가

1,234 7

Nike

Nike NSW Puffer Jacket Fir - Asia
나이키 NSW 패딩 자켓 퍼 - 아시아

빠른배송

154,000원

즉시 구매가

2,911 7

Nike

Nike NSW Storm-Fit Windrunner Hoodie Jacket Black
나이키 NSW 스톰핏 윈드러너 후드 자켓 블랙

빠른배송

160,000원

즉시 구매가

1,342 7

데이터 수집

- 상품 상세 정보 수집

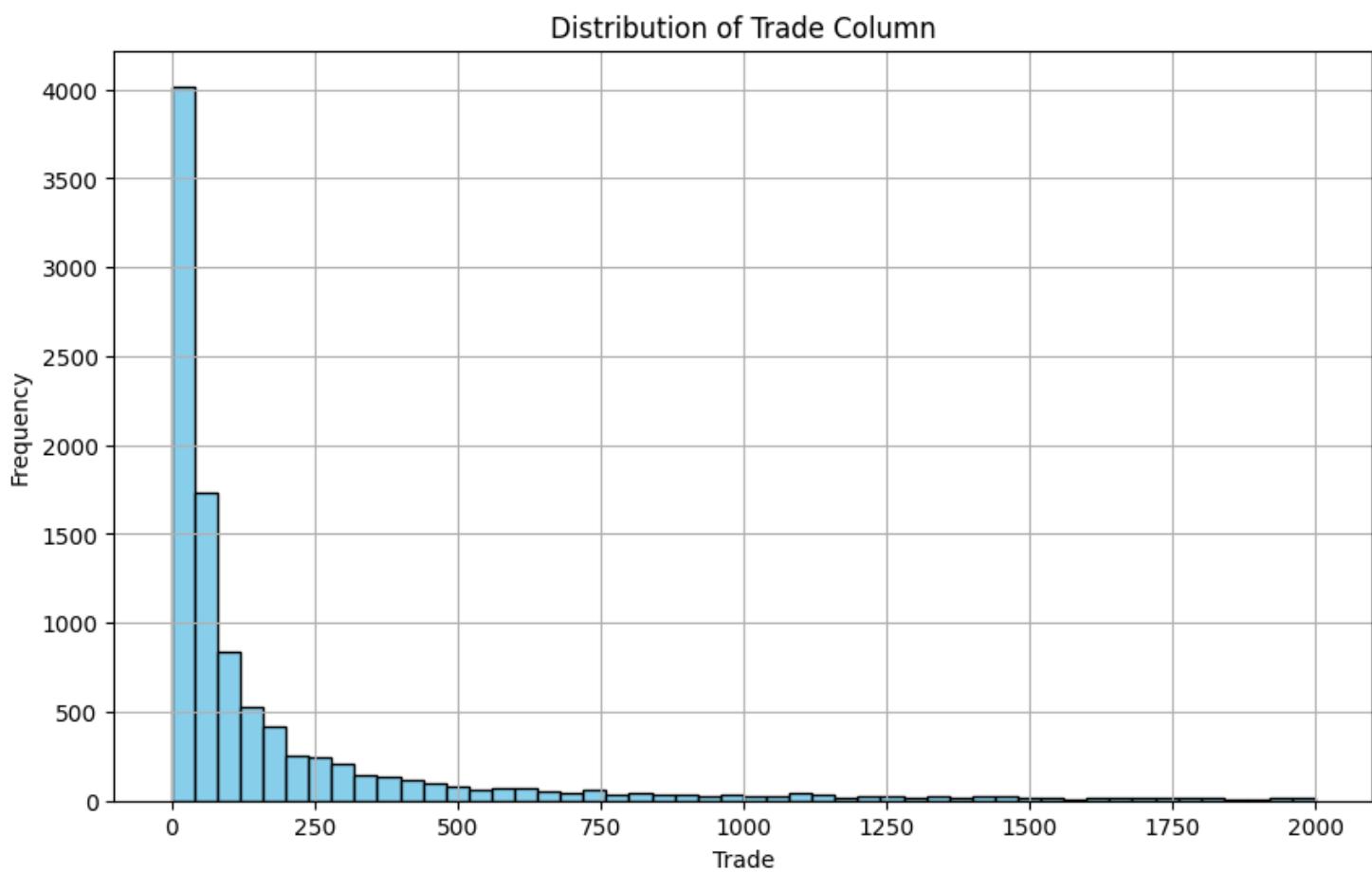
[링크, 거래 횟수, 브랜드, 상품명, 현재 가격, 관심 수, 리뷰 수, 과거 거래 가격(4건)]

Link	Trade	Brand	Name	price1	Interest	Review	price2	price3	price4	price5
https://kream-	거래 4.3만	Adidas	Adidas Samba OG Black	145,000	8.4만	845	133,000	141,000	152,000	157,000
https://kream-	거래 4.8만	Adidas	Adidas Samba OG	131,000	7.7만	984	149,000	130,000	141,000	141,000
https://kream-	거래 858	Adidas	Adidas x Wales Bonner	275,000	1만	49	283,000	299,000	248,000	291,000
https://kream-	거래 1,139	Adidas	Adidas x Wales Bonner	244,000	1만	68	317,000	289,000	350,000	300,000
https://kream-	거래 4,830	Adidas	Adidas Tobacco Gruen	179,000	2.8만	161	176,000	130,000	154,000	170,000
https://kream-	거래 1,193	Adidas	Adidas x Song for the	275,000	6,320	109	270,000	204,000	239,000	239,000
https://kream-	거래 3,029	Adidas	Adidas Tobacco Gruen	145,000	2.2만	111	189,000	189,000	180,000	163,000
https://kream-	거래 4.9만	Adidas	Adidas Superstar Core	113,000	8.7만	1,543	114,000	118,000	113,000	117,000
https://kream-	거래 8,912	Adidas	Adidas Yeezy Boost 350	325,000	1.4만	348	307,000	329,000	294,000	311,000
https://kream-	거래 1.3만	Adidas	(J) Adidas Superstar	99,000원	3만	245	94,000원	91,000원	99,000원	94,000원

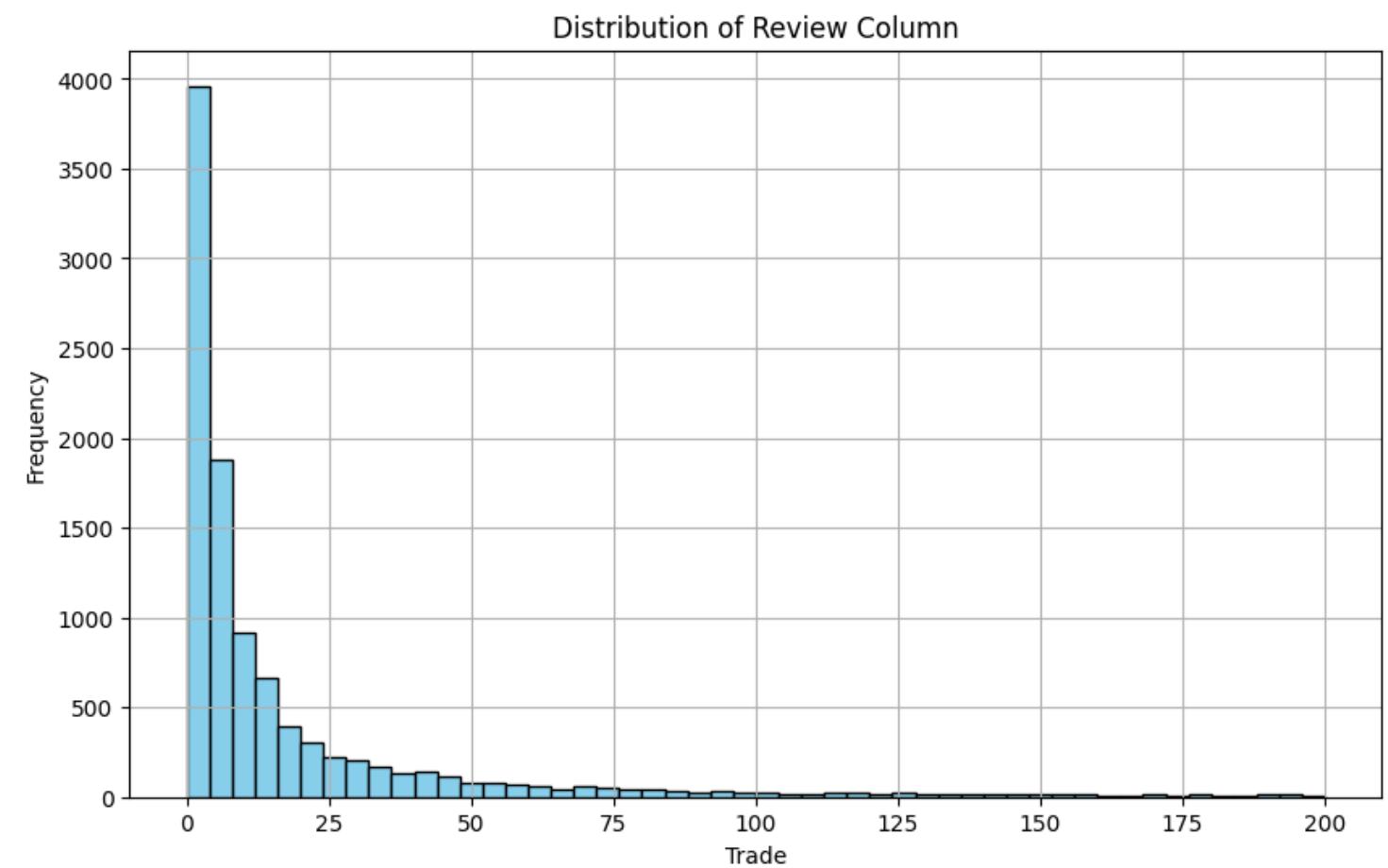
데이터 분석

1) EDA

Trade



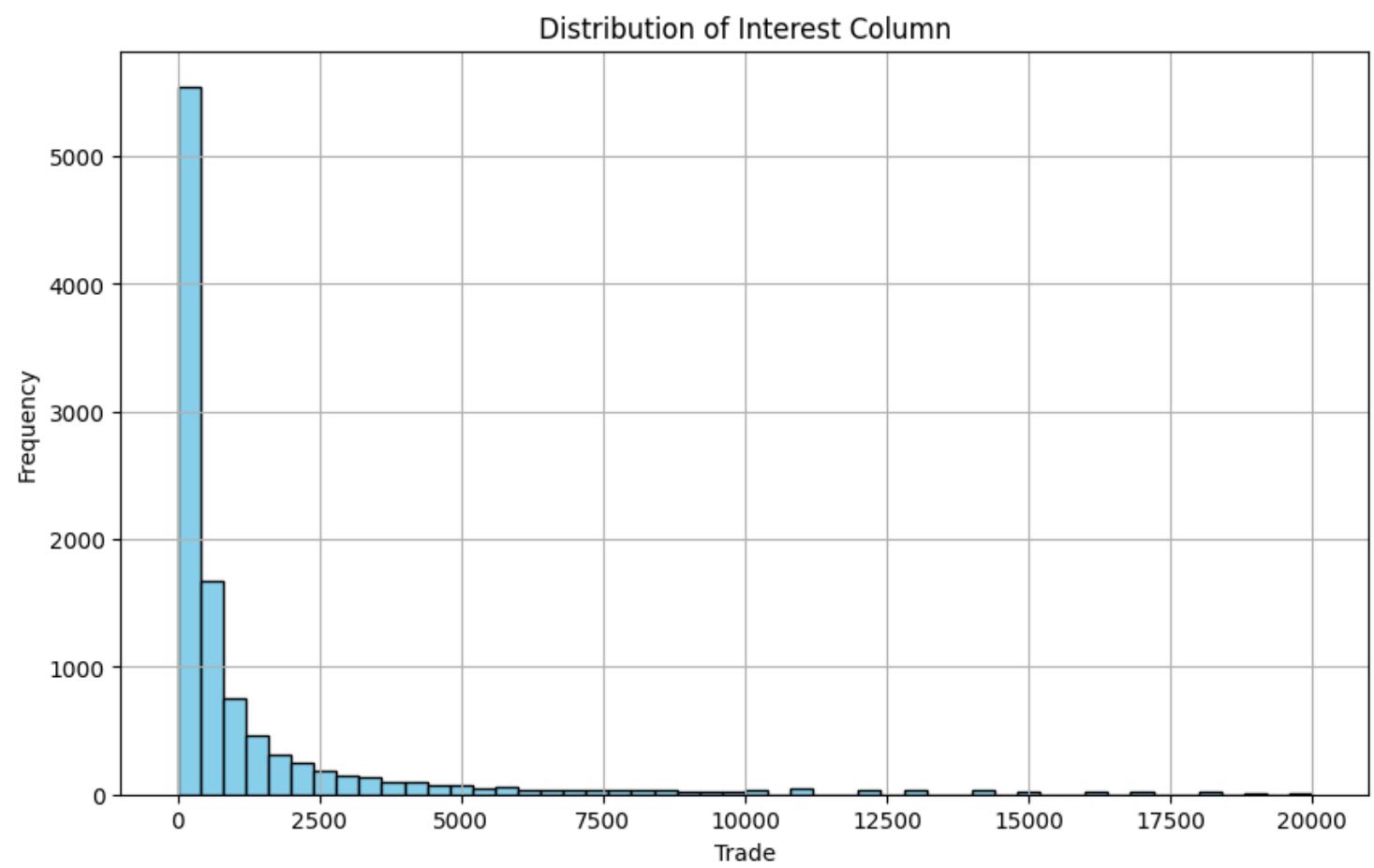
Review



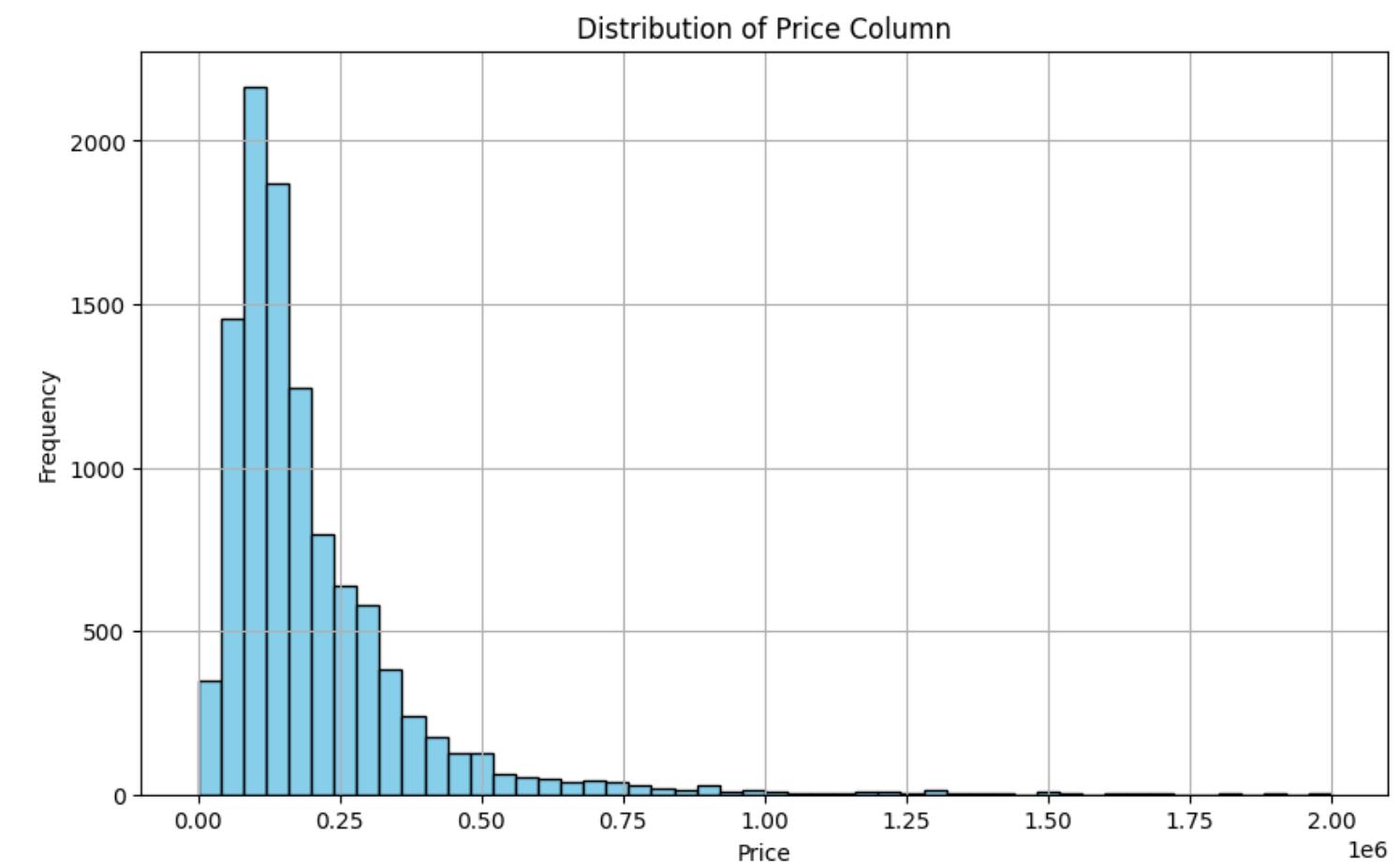
데이터 분석

1) EDA

Interest



Price



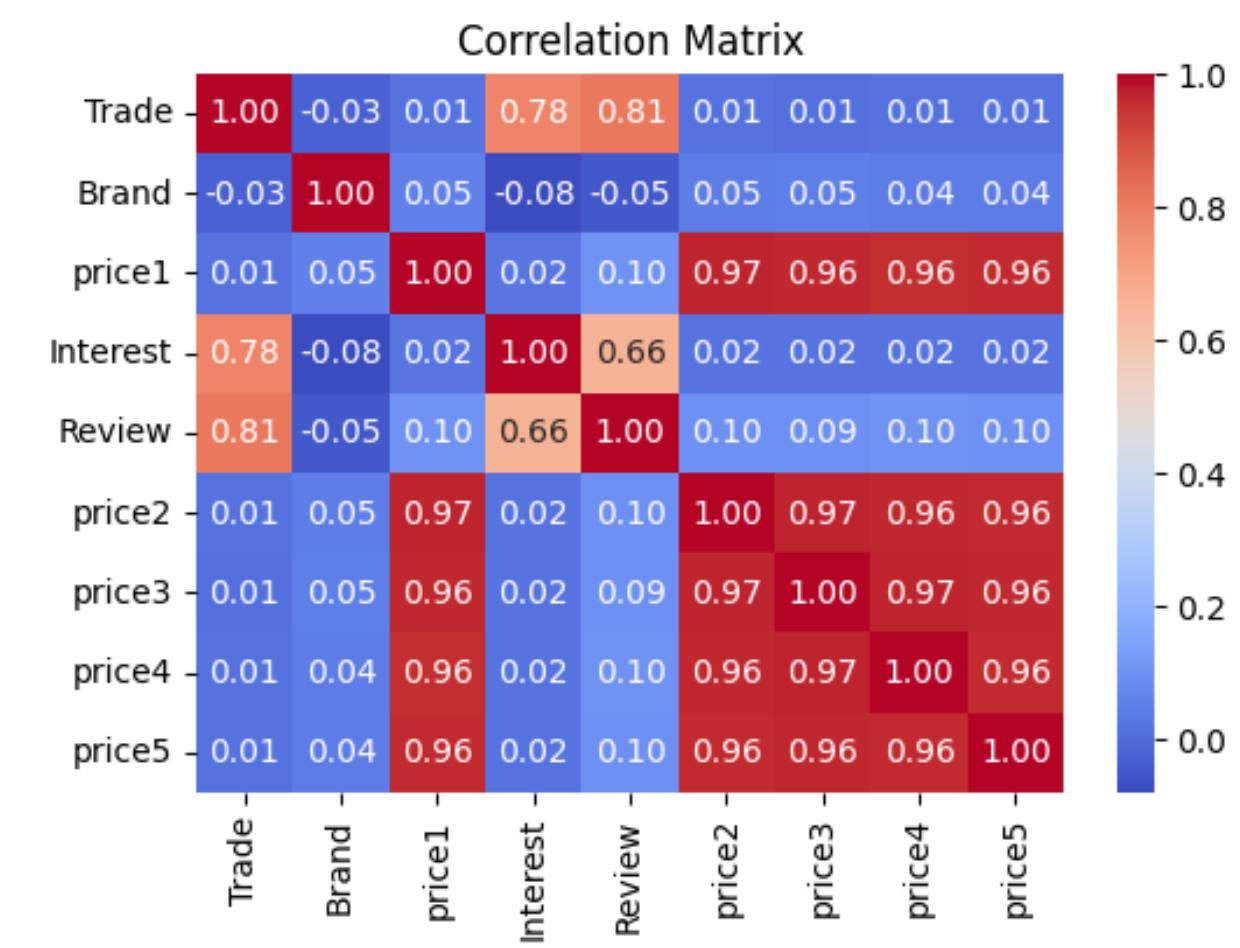
데이터 분석

2) 전처리

- Name(Korean), Link column drop
- Trade, Interest, Review column 내 문자 제거
- Brand column 라벨 인코딩
- 문자 변수 data type 정수/실수로 변환
- Price1 ~ Price 5 column의 결측치 최근 거래 가격으로 대체

3) 상관 분석

- 현재 가격(Price1)과 높은 상관관계를 가진 변수 분석
- 현재 가격(Price1)와 최근 4개 가격(Price 2~Price 5)
높은 상관성 확인



데이터 분석

4) 모델링

Random Forest

MEAN SQUARED ERROR: 3082989184.63

ROOT MEAN SQUARED ERROR: 55524.67

MEAN ABSOLUTE ERROR: 29673.02717900656

R-SQUARED: 0.95

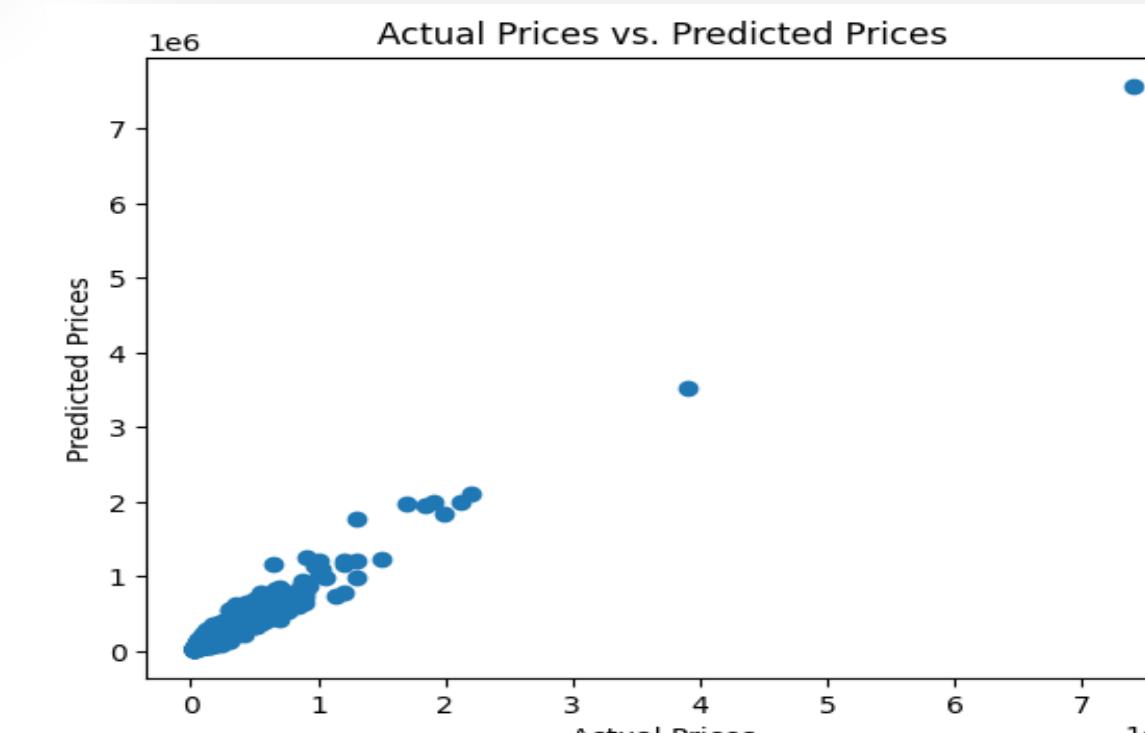
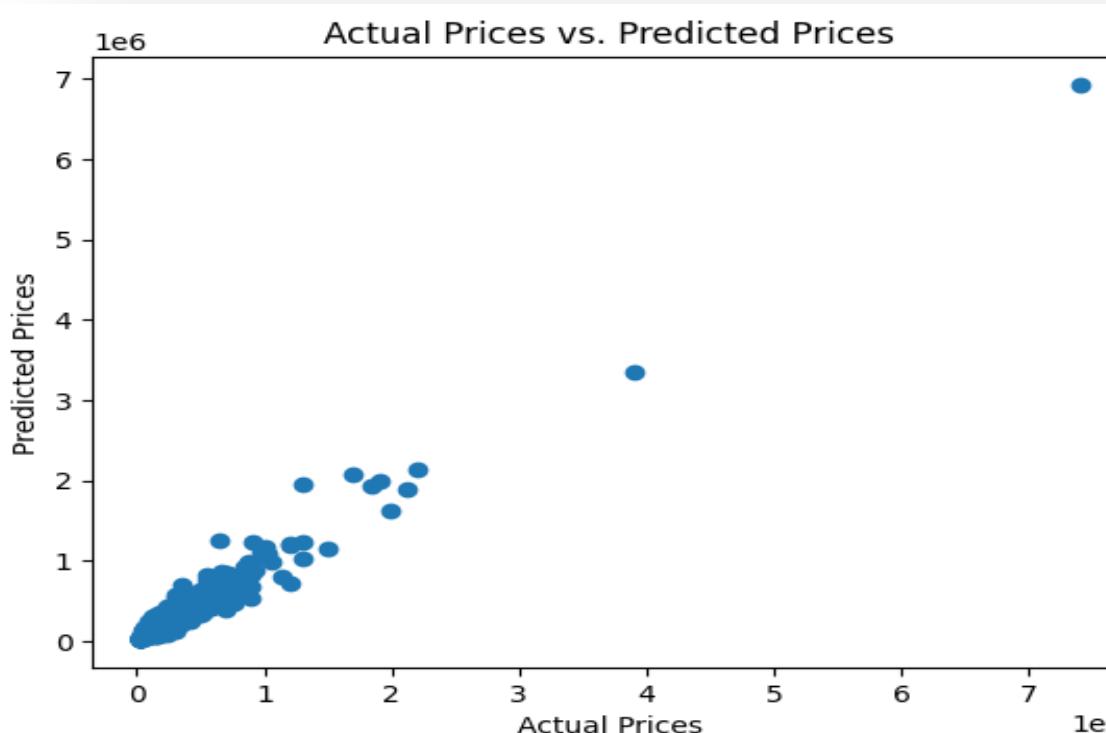
LINEAR REGRESSION

MEAN SQUARED ERROR: 2505942101.61

ROOT MEAN SQUARED ERROR: 50059.39

MEAN ABSOLUTE ERROR: 28737.003969071025

R-SQUARED: 0.96

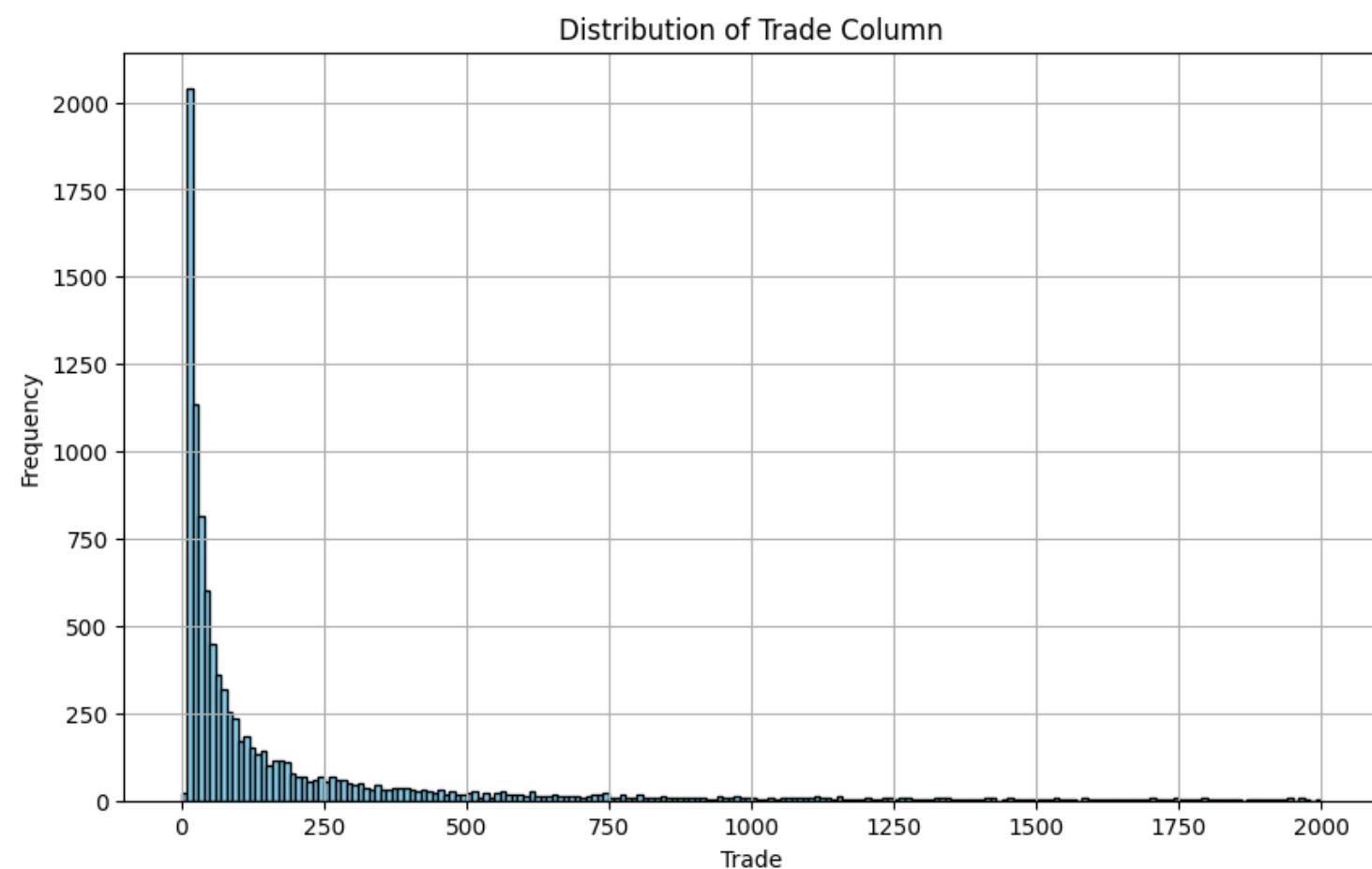


데이터 분석

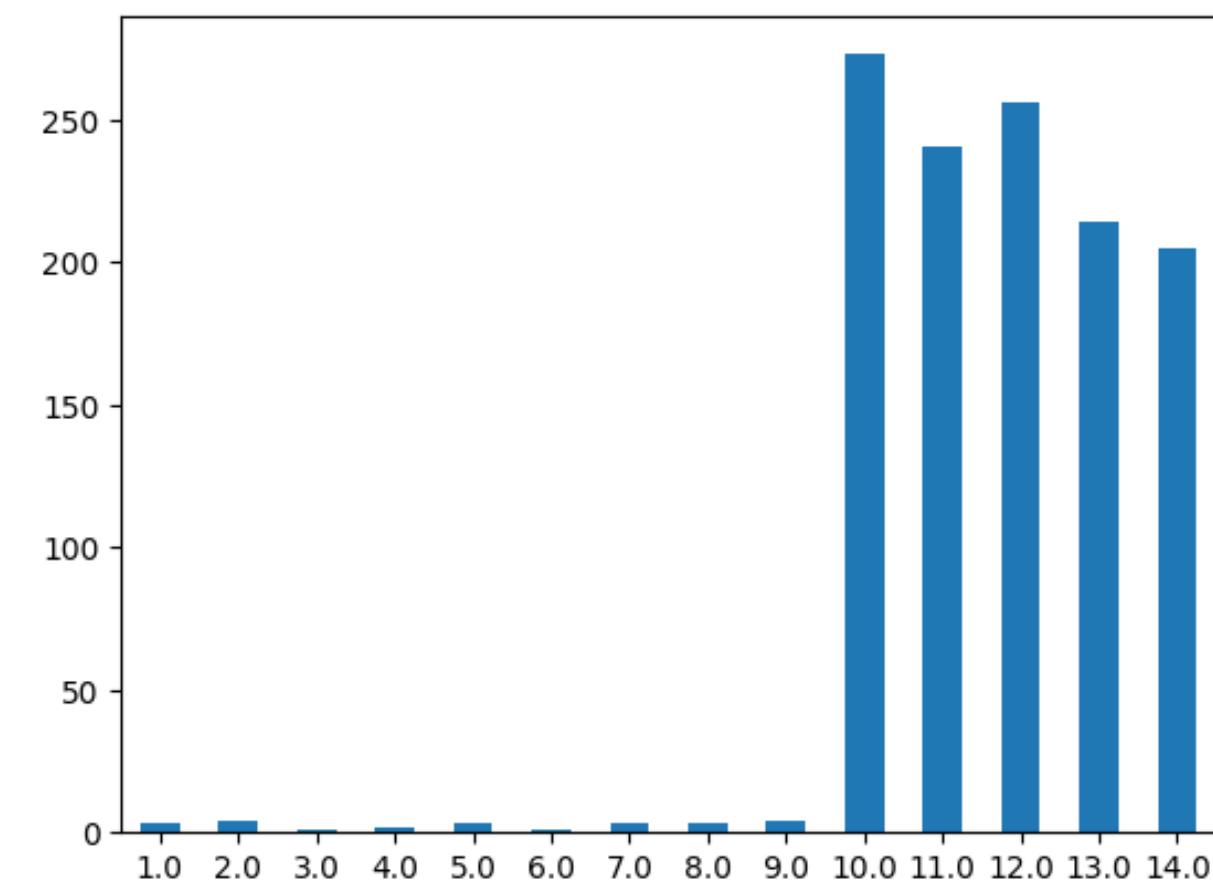
5) 거래 횟수(Trade)가 적은 아이템 구매 가격 예측

- 거래 횟수 하위 10%(1-14회)에 해당하는 데이터 추출

Trade 전체 분포



Trade 하위 10% 분포



데이터 분석

6) 모델링

Random Forest

MEAN SQUARED ERROR: 2022664542.80

ROOT MEAN SQUARED ERROR: 44974.04

MEAN ABSOLUTE ERROR: 27962.83950617284

R-SQUARED: 0.93

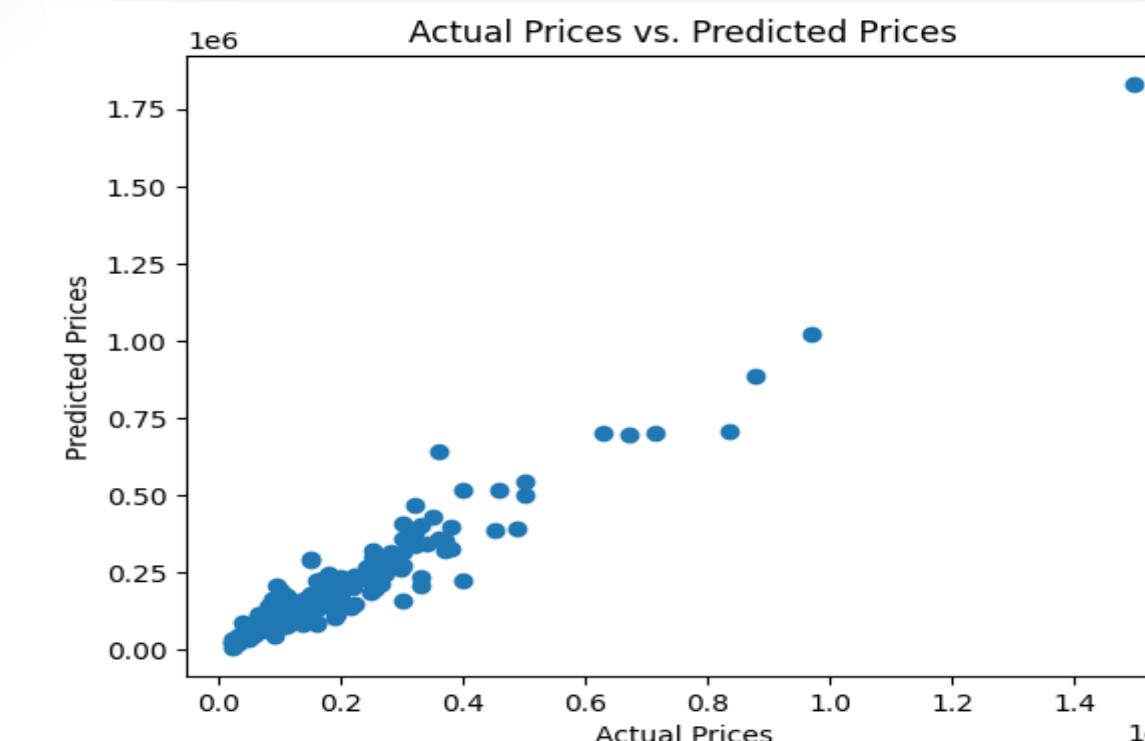
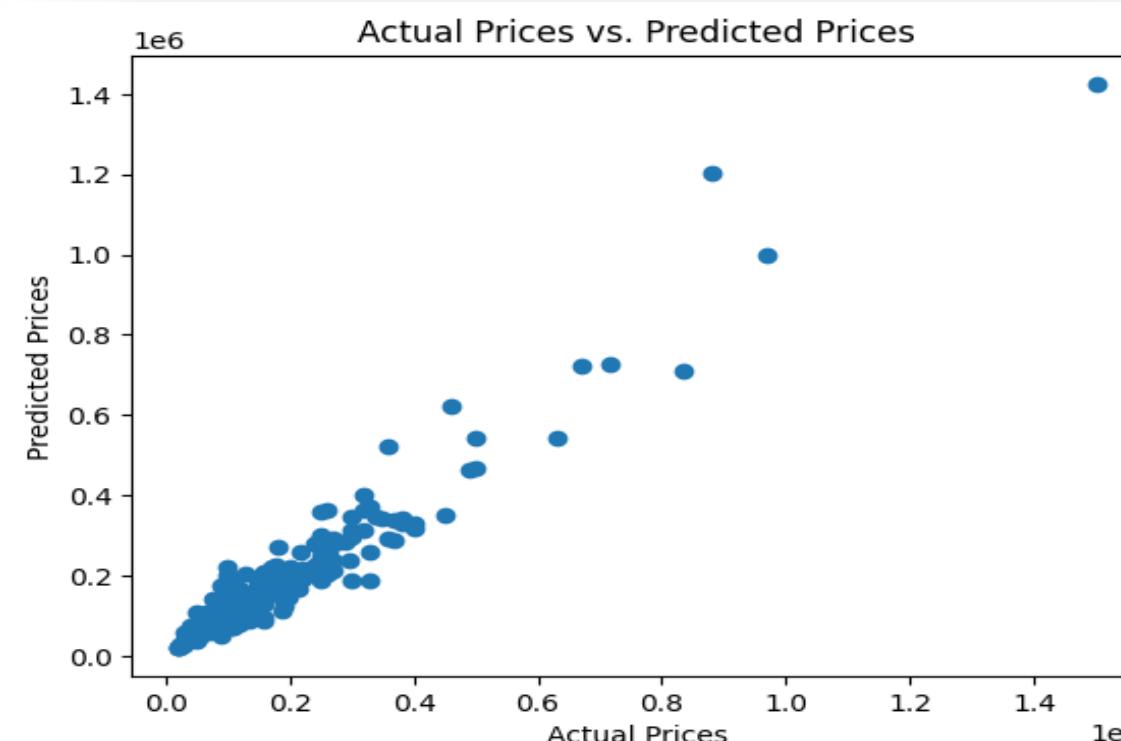
LINEAR REGRESSION

MEAN SQUARED ERROR: 2497447263.91

ROOT MEAN SQUARED ERROR: 49974.47

MEAN ABSOLUTE ERROR: 30022.199306504204

R-SQUARED: 0.91



결론

1. 중고 아이템의 상세정보 [브랜드, 관심 수, 리뷰 수, 과거 거래 가격..]를 바탕으로 Random Forest, Linear Regression을 활용해 구매 가격 예측이 가능하다.
2. 위 모델을 활용하여 거래 횟수 하위 10%(1-14회)에 해당하는 아이템의 구매 가격 예측이 가능하다.
3. 과거 가격 데이터 추가 및 사이즈 변수까지 고려하여 데이터셋을 구성한다면 보다 높은 정확도와 실용성을 갖출 수 있을 것이라 기대된다.

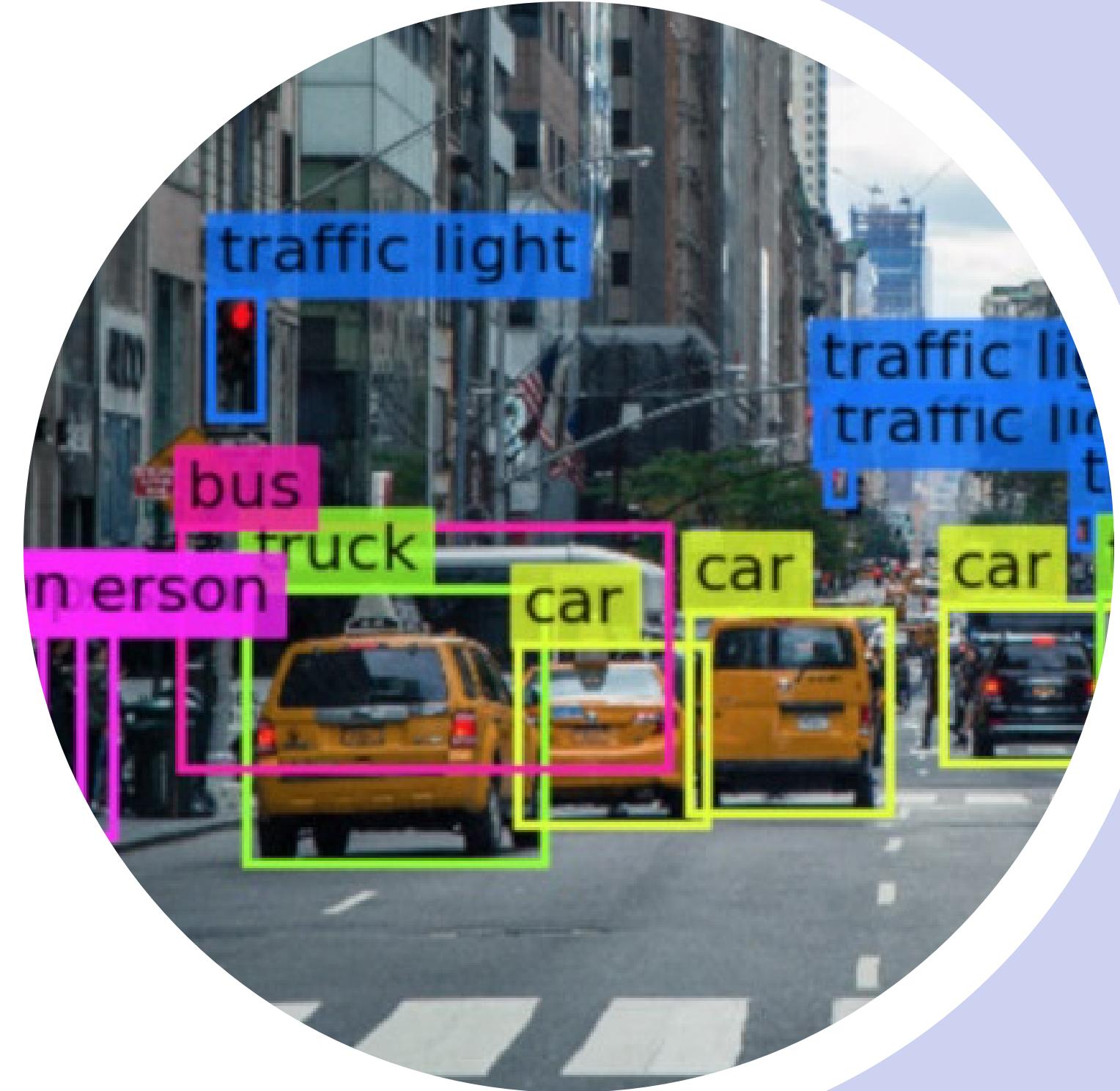
분석 목표

판매 가격 예측, 판매 여부 예측

- 중고거래 플랫폼의 운영자의 입장에서 분석

데이터 라벨링 (새로운 Task)

- 중고거래 플랫폼의 이용자의 입장에서 분석



분석 목표

중고거래 플랫폼의 이용자의 입장에서 분석

- 올해 6월에 삼성에서 자가수리 프로그램을 시작
- 그러나 정품 부품은 가격이 상당히 비쌈

일반화된 목표

- 라벨이 필요한 어느 물품이든 적용할 수 있는 방법론을 만들기



자가수리 프로그램 안내

을 보유하고 계신 고객은 삼성전자서비스 홈페이지를 통해 정품 부품과 수리도구를 구매하여 저 대상 모델별 구매 가능한 부품 종류와 수리 가이드를 확인해 보세요.

※ 먼저 영상을 시청하신 후 자가 수리 여부를 결정하시기 바랍니다.

※ 자가 수리를 시도하신 이후에는 유상 수리만 가능합니다.

※ 방수·방진을 보증 받기 위해서는 서비스 센터에서 수리 받으셔야 합니다.

자가수리 매뉴얼 확인, 부품 구매



데이터 수집

이것이 데이터 분석이다 with 파이썬

- Ch.05 데이터 종합 분석 예제
 - 5.1 중고나라 휴대폰 거래가격 예측하기

중고나라 휴대폰 거래 데이터셋: 4176개

create_date: 판매(or 구매) 게시글이 올라온 시점

price: 게시글 작성자가 제안한 휴대폰의 거래가격

text: 게시글의 제목과 본문을 합친 텍스트 데이터

phone_model: 휴대폰의 기종

factory_price: 휴대폰의 공시가격

maker: 휴대폰 제조사

price_index: 판매 게시글이 올라온 시점에서의 휴대폰 물가 지수 데이터

수 데이터



최종 방법론

GPT-4 Labeling

- 2차 발표에서 GPT-4가 정한 카테고리가 나쁘지 않다고 판단
- 아래와 같은 규칙으로 label을 지정, 그러나 결측치 발생
- 중고거래 사이트에 상품을 판매하려는 사람의 특성 상 자신의 물건의 좋은 점만 어필하려고 함 → 레이블 불균형

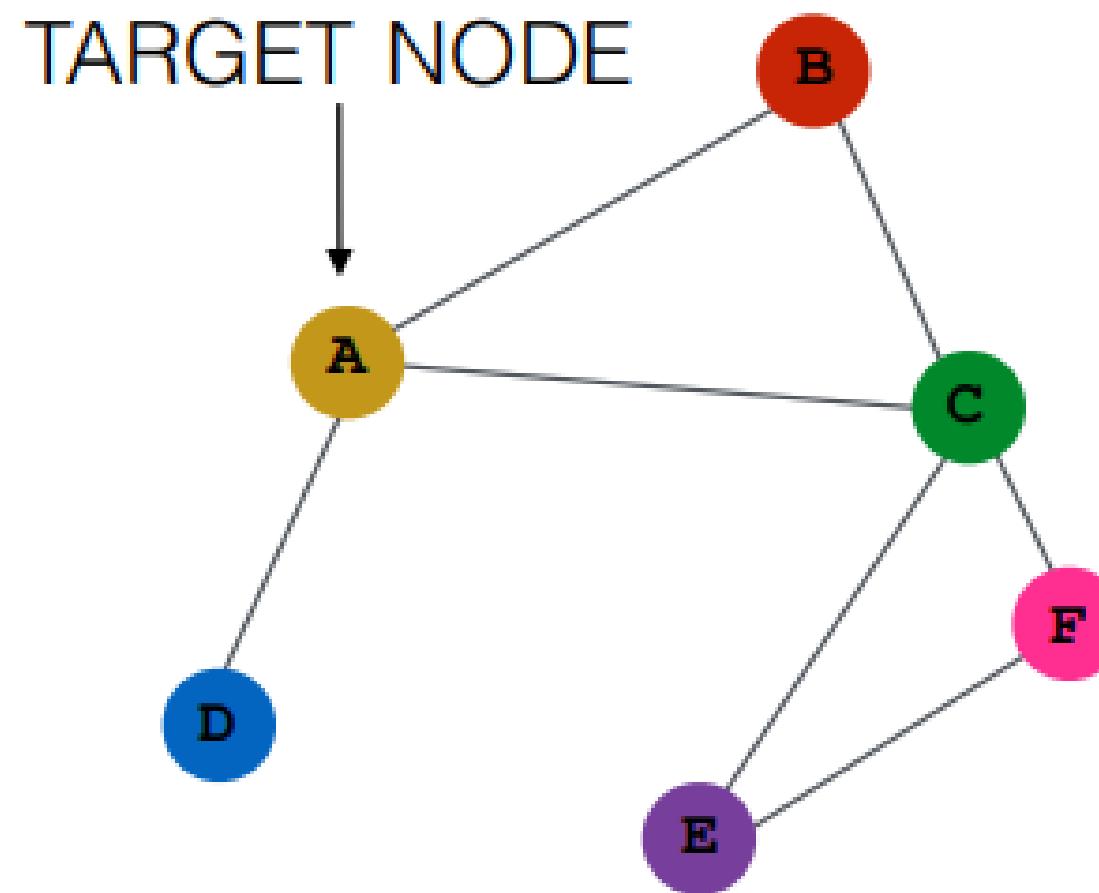
```
updated_patterns = {
    1: r"흠집 없음|배터리 상태 좋음|새상품|미사용|완벽한 상태|아주 좋은 상태|최상|A급|완전 새것|새 것 같은거
의 새것",
    2: r"사용감 조금 흠집 몇 개|좋은 상태|사용감 적음|양호한 상태|B급|관리 잘 됨|사용감 약간|약간의 기스",
    3: r"보통 상태|사용감 있음|흠집 있음|정상 작동|소모품 교체|C급|평균적인 상태|일반 사용감|통상적인 사용
감",
    4: r"액정 파손|사용에 불편함|부분적 고장|수리 필요|작동에 문제|D급|약간의 고장|일부 기능 이상|사용에 지
장",
    5: r"매우 나쁨|고장|작동하지 않음|부품 누락|심각한 손상|E급|심각한 결함|많은 손상|고장남|수리 불가"
}
```

```
print(phone_data['label'].value_counts())
1.0    605
5.0    100
3.0     49
2.0     30
4.0     21
Name: label, dtype: int64
```

최종 방법론

결측치 처리 - 그래프 기반 방법

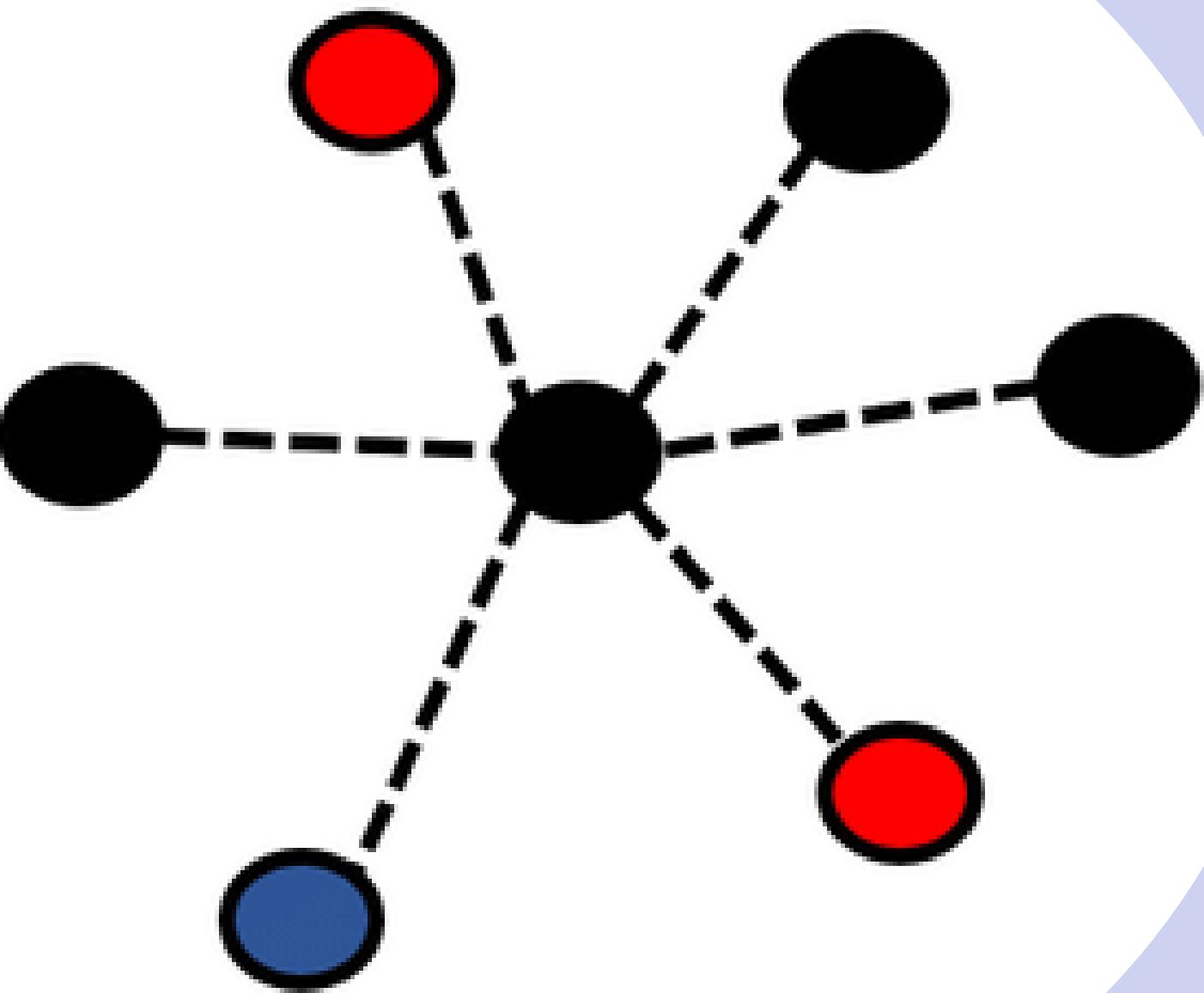
- 데이터 간의 관계를 그래프로 모델링하고, 이 그래프 구조를 사용하여 레이블 정보를 전파
- 일반적인 semi-supervised learning 방법 중 하나인 LabelPropagation 알고리즘과 유사하지만, 데이터 간의 관계를 더 복잡하게 모델링할 수 있음
- text(휴대폰 설명) feature를 tf-idf vector를 통해 수치화
 - 게시물의 text에 대한 코사인 유사도가 0.5가 넘으면 edge가 있는 것으로 간주
 - Data(x=[4951, 3], edge_index=[2, 5970], y=[4951])
- label 불균형 해결을 위해 에포크마다 각 레이블에서 21개씩 뽑아서 학습 → 가장 적은 레이블인 4번이 21개라서



가설

데이터 간의 복잡한 관계도 잘 잡아낼 수 있는 그래프의 특성 상 괜찮은 결과가 나올 것

텍스트에 등장하는 단어의 유사도 기반으로 간선을 구성하였기 때문에 합리적인 방향으로 레이블이 전파될 것



데이터 분석 결과

사용 모델

- GCN, GAT, JKNet
 - 각각 Convolution, Attention, Skip connection 기반
- Train: 텍스트를 통해 지정된 레이블이 존재하는 데이터
- Test: 나머지 결측 데이터
- JKNet의 경우 Train acc는 92%까지 상승

Method	Accuracy	F1 Score
GCN	47.62	45.83
GAT	20.00	6.67
JKNet (6 layers)	69.52	67.68

```
phone_data['label'].value_counts()
1.0    1878
5.0    948
2.0    942
3.0    613
4.0    570
Name: label, dtype: int64
```

데이터 분석 결과

갤럭시 s3 lte 16기가 판매합니다 액정 조금 깨진것 말고는 정상작동 다 됩니다 ... 사진과 같이 액정에 약간 금이 가 있는 상태이며 테두리쪽은 생활기스가 정도 있는 편입니다 ... → 3

skt 베가아이언2 팝니다 망가짐3만원 → 5

sh-860 와인폰3 직접거래 모델명 sh-860 와인폰3상 태 생활기스 조금 있습니다 ... 유심만 너어서 사용하심 됩니다 ... → 2

애플라이트 64기가 팝니다 아이폰6플러스 블루

2기가 팝니다 직거래 갤럭시s6엣지 32기가 품명
로 팝니다~~~ 새상품급 실기스조차 없어요 직접거
란 폰 단품판매합니다 직접거래 sk g5 티탄 폰 단품

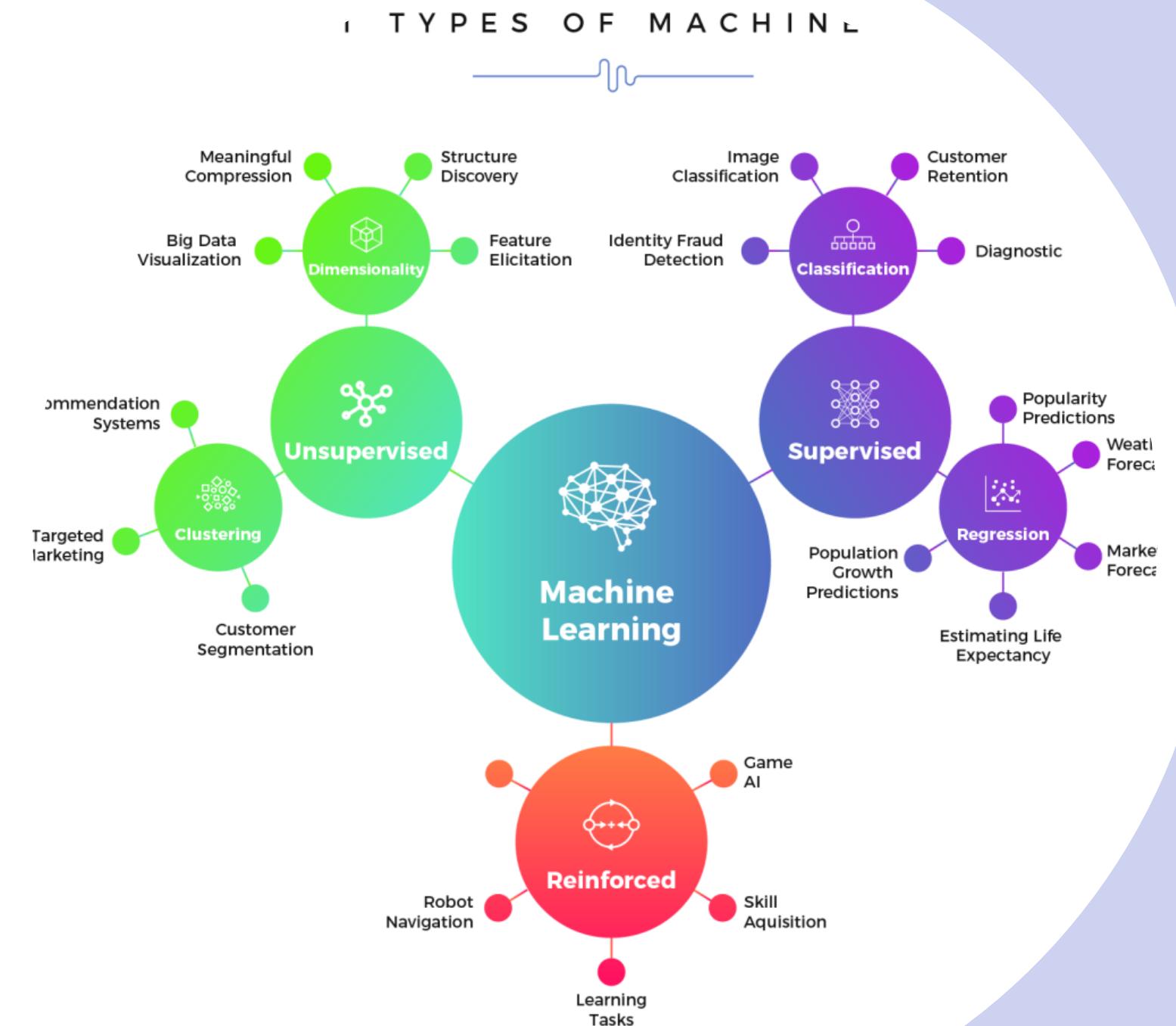
sony 엑스페리아 c5 ultra e5506 16gb 미사용

추가 실험

우리가 직접 붙인 레이블을 다른 모델에게 분류하라고 시킨다면 어떤 결과가 나올까?

1. 텍스트 포함, 텍스트 미포함, 텍스트 차원 축소
2. Logistic Regression, SVM, Random Forest, MLP

(1.)과 (2.)를 조합한 12가지 경우에 대하여 실험



추가 실험

Method	Text Feature	Accuracy	F1 Score
Logistic Regression	100D	0.39	0.35
	SVM	0.49	0.47
	Random Forest	0.64	0.64
	MLP	0.40	0.32
Logistic Regression	0D	0.39	0.34
	SVM	0.46	0.43
	Random Forest	0.67	0.67
	MLP	0.21	0.10
Logistic Regression	2D (t-SNE)	0.41	0.34
	SVM	0.46	0.43
	Random Forest	0.67	0.67
	MLP	0.42	0.32

시도해본 방법론

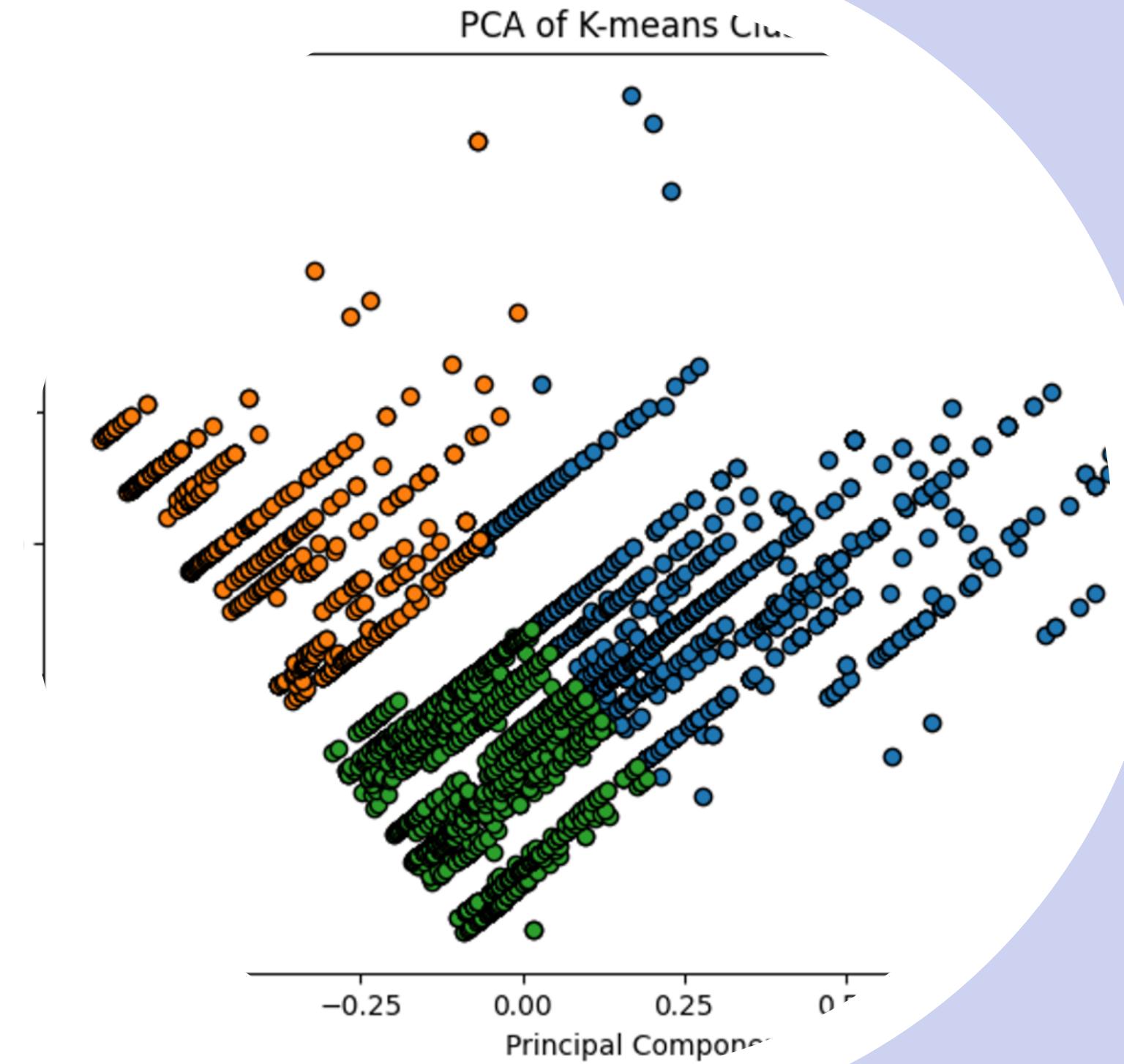
1. LabelPropagation, LabelSpreading
2. K-Means Clustering

결과 예시

… 액정에 기스가 나 있는 것처럼 보이지만 액정 보호 필름을 붙인 상태라 떼고 새걸로 다시 붙여 사용하시면 아마 깨끗할거예요 ㅎㅎ … → 사용 불가

… 부모님이 항상 액정보호 필름 붙이고 다이어리 케이스 사용하면서 가볍게 사용하였던 폰이라 상태 정말 좋습니다 배터리도 2개 전부 센터에서 교환 가능 할 때 교환 하여 상태 좋습니다 … → 사용 불가

… 구성품 본체 및 충전기, 케이블, 액정보호필름 사용하지 않은 신품입니다 …
→ 정상은 아니지만 사용에 지장 없음



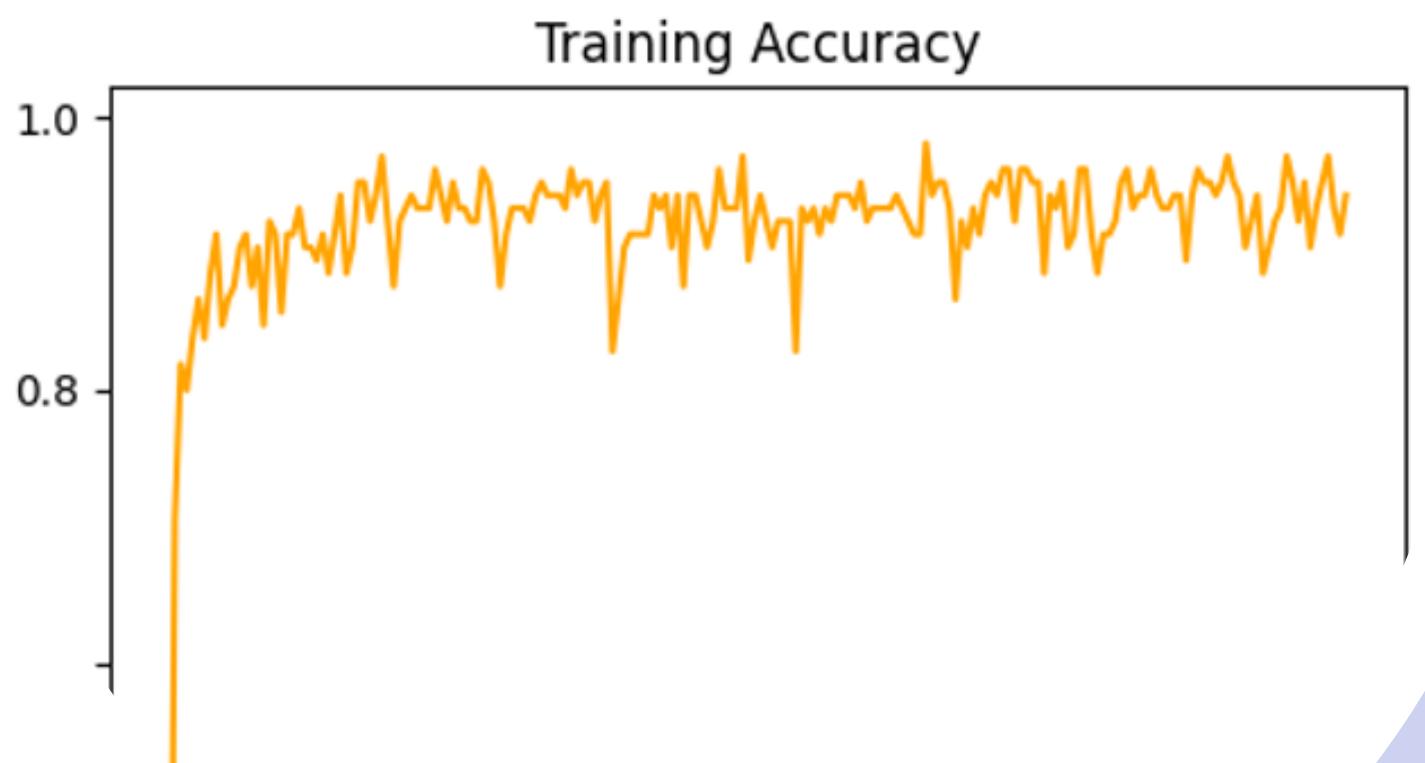
에러 사항

1. JK-Net

- 샘플링에 크게 의존한 학습 방식을 사용했기 때문에 훈련 집합과 평가 집합 사이에 큰 정확도 차이 관찰 → 과적합 (92% vs. 70%)

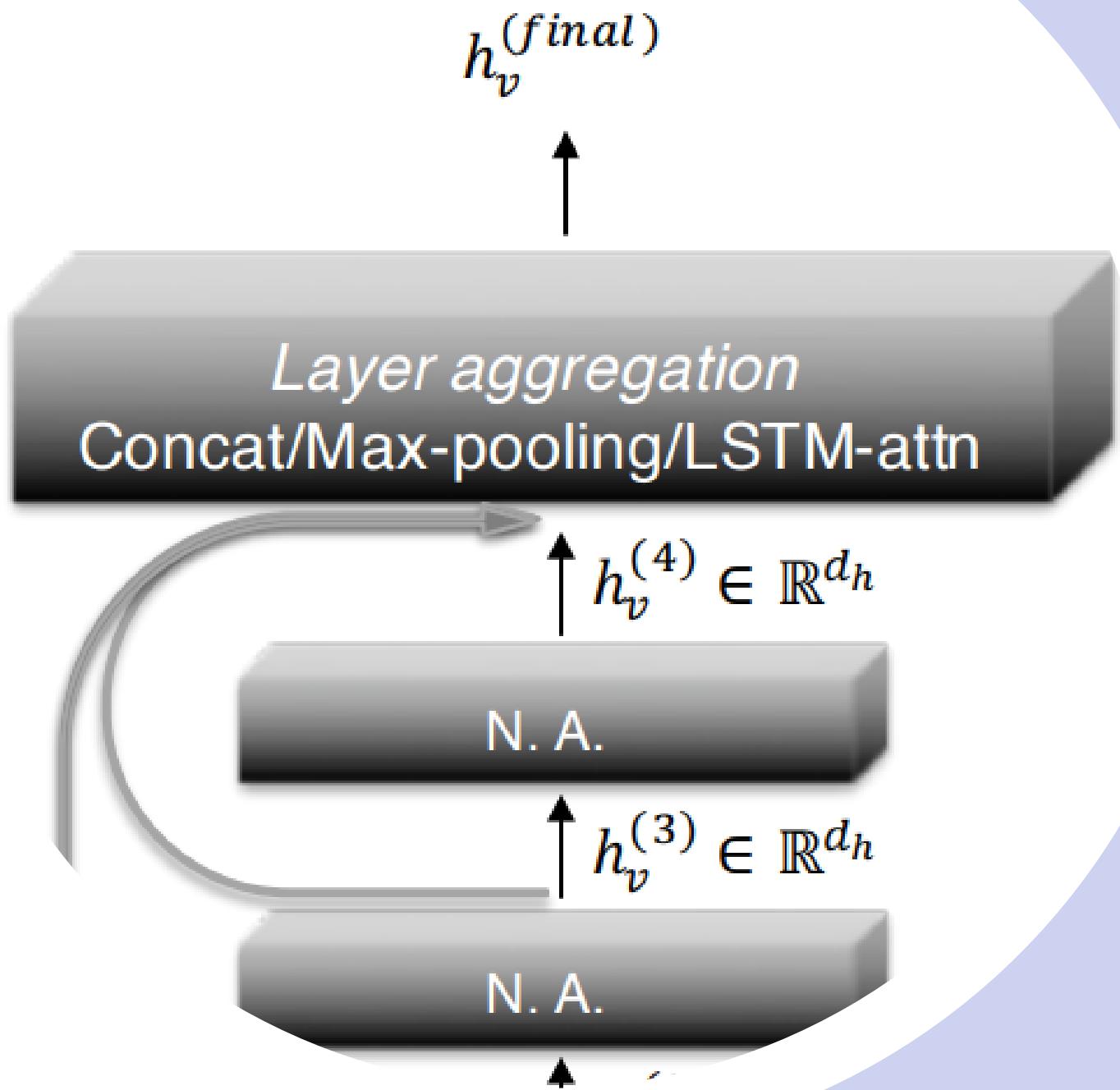
2. 간선 구성

- 단순히 tf-idf vector의 코사인 유사도가 높은 것을 기준으로 잡았기 때문에 다른 특성을 잘 반영하지 못함



결론

- 그래프 신경망에 기반하여 목적에 맞는 레이블을 만들어내는 것은 효과적인 방법론
- 그래프를 구축할 때 텍스트 변수와 자연어처리 기법을 결합하는 것도 효과적
- 계층을 깊게 쌓으면 oversmoothing 현상이 일어나는 convolution, attention 기반 기법보다는 계층을 깊게 쌓으면서도 지역적 특성을 잘 반영할 수 있는 skip connection 기반 방법이 더욱 효과적



분석 목표

중고 거래 플랫폼 내 업로드된 아이템의 7일 이내 판매 여부 예측
Traditional ML vs. LLM

전여친이 준 향수

뷰티/미용 · 3일 전

3,000원

딱 한번썼고요 향은 좋아요 전여친이 줬는데 쓰기 좀 그래서 팔아요

관심 18 · 채팅 29 · 조회 737

분석 목표

중고 거래 플랫폼 내 업로드된 아이템의 7일 이내 판매 여부 예측

Task의 의의

- 이용자가 상품을 등록하는 과정에서 7일이 지나도록 팔릴지에 대한 여부를 "텍스트만을 통한 추론한 결과"와 함께 제공한다면 이용자 입장에서 제목이나 상품 설명을 더욱 매력적이고 구체적으로 작성할 수 있게 하는 계기가 됨
- 이미 7일이 지난 상품의 경우에도 "끌올" 기능을 통해서 시간을 업데이트 할 수 있고 이때에도 7일 이내 판매 가능성을 제공한다면 더 나은 서비스가 될 것으로 예상

Task의 도전적인 점

- 제목과 상품 설명만, 즉 텍스트만 이용하여 이진 분류 문제를 해결하는 것이 유효한가?
- 텍스트를 제외한 다른 변수들만을 이용해서 분류 문제를 해결하는 것은 너무나 간단한 문제이고, 2차 발표에서 그 성능이 입증되었음을 확인함

가설

중고 거래 플랫폼 내 업로드된 아이템의 7일 이내 판매 여부 예측

Traditional ML vs. LLM

- LLM이 우세할 것으로 예상
- 모델 매개변수의 개수가 많아질 수록 성능이 크게 늘어날 것으로 예상
- Prompt tuning의 영향이 LoRA fine-tuning보다 클 것으로 예상

데이터 수집

중고 거래 플랫폼 내 업로드된 아이템의 7일 이내 판매 여부 예측

크롤링을 진행했을 당시에 7일 전 게시물 주소

<https://www.daangn.com/articles/658000000> (대략적인 값)

레이블 불균형을 방지하기 위하여 크롤링을 진행할 주소의 범위를 신중하게 결정

```
import random

address = []

for i in range(100000):
    address.append(651000000 + random.randint(1, 15000000))

address.sort()
```

데이터 수집

매너온도, 제목, 카테고리, 게시되고 지난 시간(끌올 포함), 상품 설명, 관심, 채팅, 조회



트위드러브

송파구 가락1동

69.6 °C

매너온도

나트 코트

여성의류 · 1일 전

20,000원

살때 비싸게 구매한 옷이지만

안입을 것같아 옷장비우는 차원에서

정말 저렴하게 보내요....

관심 8 · 채팅 13 · 조회 161

프롬프트 튜닝의 원칙

1. 지시

- '작성하기', '분류하기', '요약하기', '번역하기', '정렬하기' 등의 명령어를 사용하여 달성하고자 하는 결과를 모델에 지시

2. 특이성

- 모델에 실행시킬 지시와 작업을 구체적으로 설명
- 프롬프트가 상세하고 구체적일수록 결과는 더 좋아짐
- 더 나은 결과를 내는 특정한 토큰이나 키워드는 존재하지 않음

3. 해야 하는 것과 하지 말아야 할 것

- 프롬프트를 설계할 때 '하지 말아야 할 것'보다는 '해야 하는 것'에 초점을 두기

데이터 분석 - 2차 발표 이후의 결과들 (에러 사항 포함)

kullm-polyglot-12.8b-v2 + zero-shot prompt

[예측하기]

제목과 상품 설명을 참고해서 이 상품이 플랫폼에 업로드되고 나서 지난 시간이 7일 이상일지 또는 7일 미만일지 예측해 줘.
간단하게 생각하면 상품이 판매되는 데에 7일 미만이 걸릴지 또는 7일 이상이 걸릴지 예측하는 것과 비슷해.
네가 출력해야 하는 것은 상품이 플랫폼에 업로드되고 나서 지난 시간이 7일 이상일지 또는 7일 미만일지 예측한 결과야.
상품이 판매되는 데에 7일 미만이 걸릴 것 같으면 '0'이라고 답하면 되고 7일 이상이 걸릴 것 같으면 '1'이라고 답하면 돼.

{eval_input}

예측 결과:

- 전체적으로 좋지 못한 텍스트 생성

예측 결과: {}, {₩n}

데이터 분석 - 2차 발표 이후의 결과들 (에러 사항 포함)

kullm-polyglot-12.8b-v2 + 2-shot prompt

[예측하기]

...

제목: GNet H2블랙박스

상품 설명: 배선미포함 SD카드미포함 입니다. 생각있으신분 연락주세요

예측 결과: 1

제목: 청남방

상품 설명: 아주 깨끗한 천남방 내놔요. 지금 계절에 있기 좋은 청남방 이에요. 관심 있는 분 연락주세요.

예측 결과: 0

{eval_input}

예측 결과:

- Polyglot의 학습 데이터 중 대부분이 블로그 → LLM 본연의 특성을 볼 수 있었음

예측 결과: 000안녕하세요~~

데이터 분석 – 2차 발표 이후의 결과들 (에러 사항 포함)

kullm-polyglot-12.8b-v2 + 1 epoch fine-tuning + prompt tuning

- “단계별로 생각하라” 프롬프트
 - LLM 중에서도 큰 모델에 적용한다면 의도한 텍스트를 생성하기 좋은 방법
- 그러나 12.8B 규모의 LLM에서는 큰 개선이 관찰되지 않았음

제공된 제목과 상품 설명을 근거로, 이 상품이 플랫폼에 업로드된 후 7일 이상 지났는지 아니면 7일 미만인지 단계별로 생각하며 예측해주세요.

먼저, 제목과 상품 설명의 내용을 분석하여 상품의 유형, 목표 시장, 및 경쟁력을 파악하세요.

그 다음, 이러한 요소들이 상품의 판매 가능성에 어떤 영향을 미칠 수 있는지 고려하세요.

마지막으로, 이 분석을 바탕으로 상품이 플랫폼에 업로드된 후 7일 이내에 판매될 것으로 예상되면 '0'을, 7일 이상 걸릴 것으로 예상되면 '1'을 답변으로 제시해주세요.

{eval_input}

예측 결과:

데이터 분석 - 2차 발표 이후의 결과들 (에러 사항 포함)

kullm-polyglot-12.8b-v2 + 1 epoch fine-tuning + 2-shot prompt

제공된 제목과 상품 설명을 분석하여, 이 상품이 플랫폼에 업로드된 후 7일 이상 지났는지 아니면 7일 미만인지 예측해주세요.
상품의 판매 가능성을 고려하여, 7일 이내에 판매될 것으로 예상되면 '0'을, 7일 이상 걸릴 것으로 예상되면 '1'을 선택하여 답변해주세요.

다른 설명이나 예시를 추가하지 말고, 직접적으로 '0' 또는 '1'로만 답변해주세요.

<2개의 예시>

{eval_input}

예측 결과:

- 예측 결과를 내놓지 않고 새로운 예시를 만들려는 경향이 있었음
 - 이러한 현상을 막고자 하는 프롬프트도 만들어 봤으나 실패
- repetition_penalty hyperparameter도 지정했으나 효과가 크지 않음

제목 :

데이터 분석 – 2차 발표 이후의 결과들 (에러 사항 포함)

다른 시도들

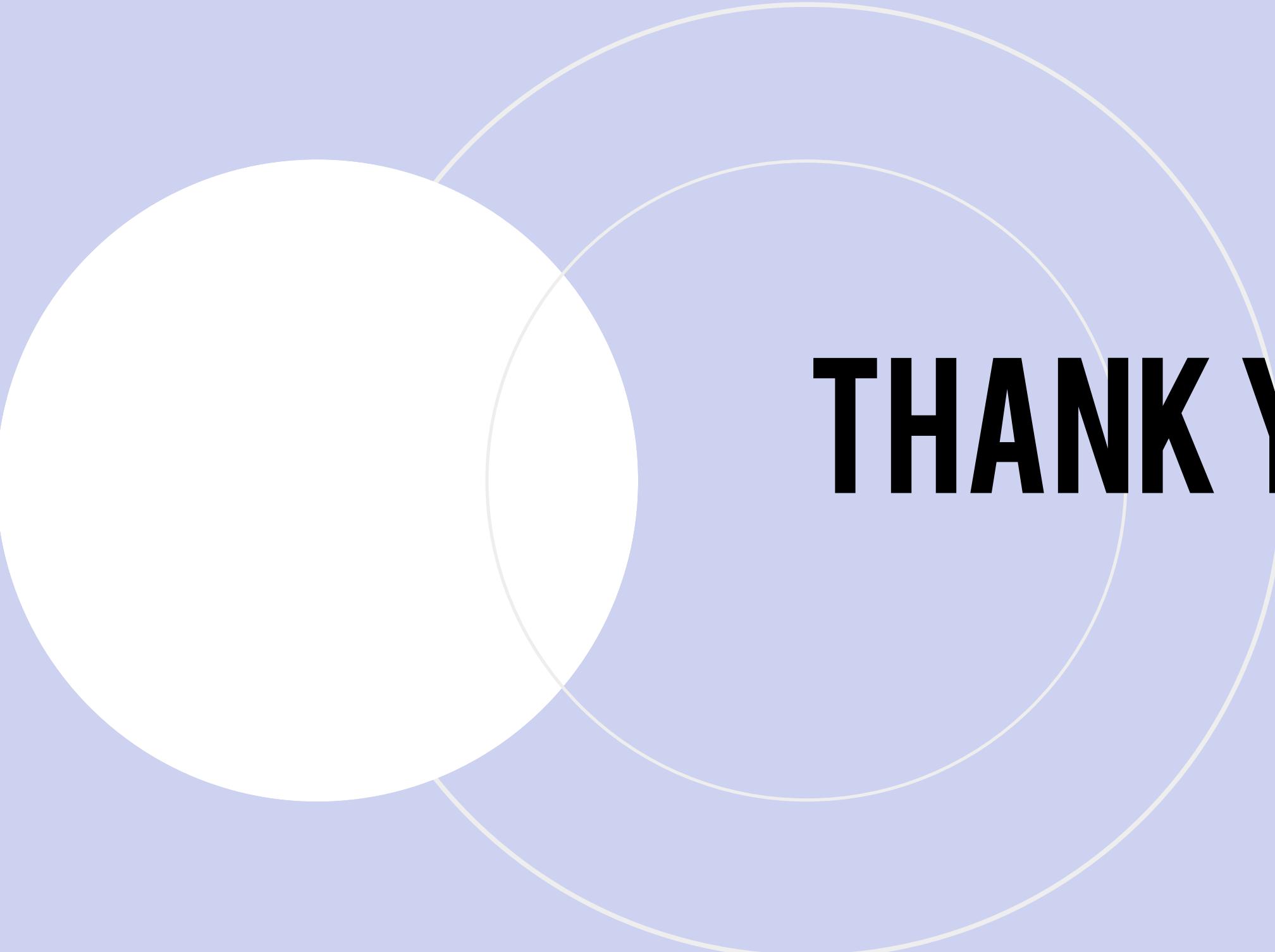
- 1-shot prompt
 - 제공된 예시의 예측 결과가 1이면 모델이 무조건 1을 출력하고, 0이면 모델이 무조건 0을 출력하는 문제 발생
- 더 많은 epoch, hyperparameter tuning 또는 더 큰 베이스라인 모델
 - 4-bit quantization, 16bit floating point, LoRA, PEFT 등의 기법들을 사용해도 학습 시간이나 추론 시간이 지나치게 오래 걸리는 문제 발생
- PLM(BERT 등) + Binary classification을 위한 reasoning layer 구조의 사용
 - 시도 가치가 높은 방법
 - 그러나 PLM은 일반적인 언어 이해에 강하지만, 특정 도메인(중고 거래 시장)의 미묘한 뉘앙스를 완전히 파악하기 어려움

데이터 분석 결과

Method	Accuracy	F1 Score
Logistic Regression (PCA, PoV: 80%)	54.69	68.75
Logistic Regression (PCA, PoV: 95%)	55.99	69.51
Random Forest Classifier	56.96	71.76
Random Forest Regressor	58.25	73.62
TF-IDF Vectorizer + Logistic Regression	64.52	77.55
TF-IDF Vectorizer + SVM	62.90	76.77
TF-IDF Vectorizer + Random Forest Classifier	54.84	66.67
KULLM-5.8B (baseline)	49.19	52.57
KULLM-5.8B (trained 2-epoch)	53.40	68.28
KULLM-5.8B (trained 2-epoch + 2-shot prompt)	58.25	73.62
KULLM-12.8B (larger baseline)	49.51	60.61
KULLM-12.8B (2-shot)	54.05	65.53
KULLM-12.8B (trained 1-epoch)	53.40	62.11
KULLM-12.8B (trained 1-epoch + prompt tuning)	54.37	68.31
KULLM-12.8B (trained 1-epoch + 2-shot prompt)	56.31	70.46
GPT-4 (prompt tuning)	<u>58.90</u>	<u>72.57</u>

결론

- 5.8B + 2 epoch보다 12.8B + 1 epoch의 성능이 떨어짐
 - 모델의 크기보다는 우리가 가지고 있는 데이터를 가지고 사전 학습 모델을 얼마나, 어떻게 미세 조정 하는지가 더 중요한 문제
- 그렇게 좋지 않은 accuracy를 가지고 있는 상황에서 accuracy보다 F1 score가 높다는 것의 이미
 - 모델이 편향을 가지거나 아무런 근거 없이 찍어서 추론했다는 것이 아니라 어느 정도 학습이 되었다는 증거
 - 즉, 완전히 텍스트에 기반해서 판매 여부 예측을 하는 것도 유효하지 않은 접근이 아님
→ 기업에서는 더 많은 컴퓨팅 자원으로 훨씬 큰 모델에 많은 학습을 할 것이기 때문에 발전 가능성이 큰 방법



THANK YOU!
