



3조_기획서

≡ 속성 1

1. 문제 배경

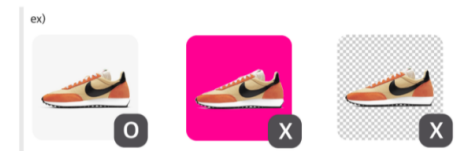
- 번개장터에서 제공하는 판매자 가이드북은 존재한다.
- 다만, 기초적인 내용과 제한사항을 제시하고 있으며 이조차도 현실적으로 지켜지지 않는 것이 문제이다.
- 이처럼 기초적인 가이드를 따르지 않는다는 것은 가이드북의 필요성이 느껴지지 않기 때문이다.
- 기업의 이윤 극대화를 위해서는 판매자들의 광고가 유저에게 전달되어야 하고 이를 위해 판매자용 가이드북을 개선할 필요가 있다.

광고 텍스트 가이드

1. 상점명 : 최소 2자 이상 ~ 최대 10자 / 띄어쓰기 불가
2. 상품명 : 띄어쓰기를 포함한 최소 2자 이상 ~ 최대 40 자
3. 홍보문구 : 띄어쓰기를 포함한 최대 2,000자
4. 광고에 관련된 내용을 객관적으로 작성해야 합니다.
5. 텍스트에는 전화번호, 주소, 이메일, 아이디 등의 광고주 정보를 기재할 수 없습니다.
6. 이용자 유인을 위해 지나치게 자극적으로 문구를 기재한 경우 광고 집행이 불가합니다.
7. 상점명, 상품명, 홍보문구(상품설명)의 경우 소재 구성 > 광고 텍스트 참고

1. 깔끔한 단색 배경으로 상품에만 집중

내 상품의 주목도를 높이기 위해서는 채도가 낮은 컬러의 배경색상 사용을 추천합니다.



2. 이미지 내 텍스트 사용 자양

이미지 내 텍스트는 가독성이 떨어질 뿐 아니라 이미지 퀄리티를 낮게 만들기 때문에 가급적 사용하지 말아주세요.



3. 상품과 관련된 이미지 사용

클릭을 유도하기 위해 상품과 무관한 이미지 혹은 선정적인 이미지를 사용하지 말아주세요.
판매까지 이어질 확률이 낮은 무의미한 클릭이 발생되어 광고비만 많이 부과되는 비효율적인 광고를 집행하게 될 수 있습니다.

2. 문제 정의

목적 : 번개장터 광고 수익 증가(eCPM)

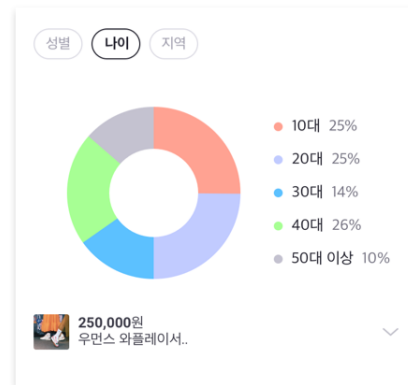
- 실제 유저가 물건을 구매했는지 판단할 수 있는 데이터가 부족하기 때문에 CTR을 높여서 번개장터의 광고 수익(eCPM)이 증가하는 것을 목표로 프로젝트 진행
- 기대수익(eCPM)이 높은 광고들의 특징을 분석
 - 판매자에게 광고 가이드를 제공하여 CTR이 증가하도록 함
 - 그에 따른 기업의 광고판매 수익 확대 예상

추천시스템 입/출력

- 입력 : 가상의 유저를 설정, 정보 입력(나이, 성별, 지금까지의 구매내역 등)
- 출력 : 그 유저가 특정 광고를 클릭할 확률(CTR)

결과물

- 유저의 특성에 따라(나이, 성별, 지금까지의 구매내역 등) 어떤 광고를 추천해줬을때 기대수익(eCPM)이 높을지 예측
- 또한 유저 특징(나이, 성별), 광고 카테고리(전자기기, 의류), 광고 시간, 광고에 들어가는 이미지, 텍스트 등에 따른 예상 클릭율(CTR)을 분석하여 광고주(판매자)에게 분석자료로 제공



광고 성과를 확인하면 생각과 다른 결과를 확인할 수
있어요. 30대 여성이 관심있을 거라고 생각했는데,
실제로는 40대 여성이 더 관심을 보일 수 있죠.

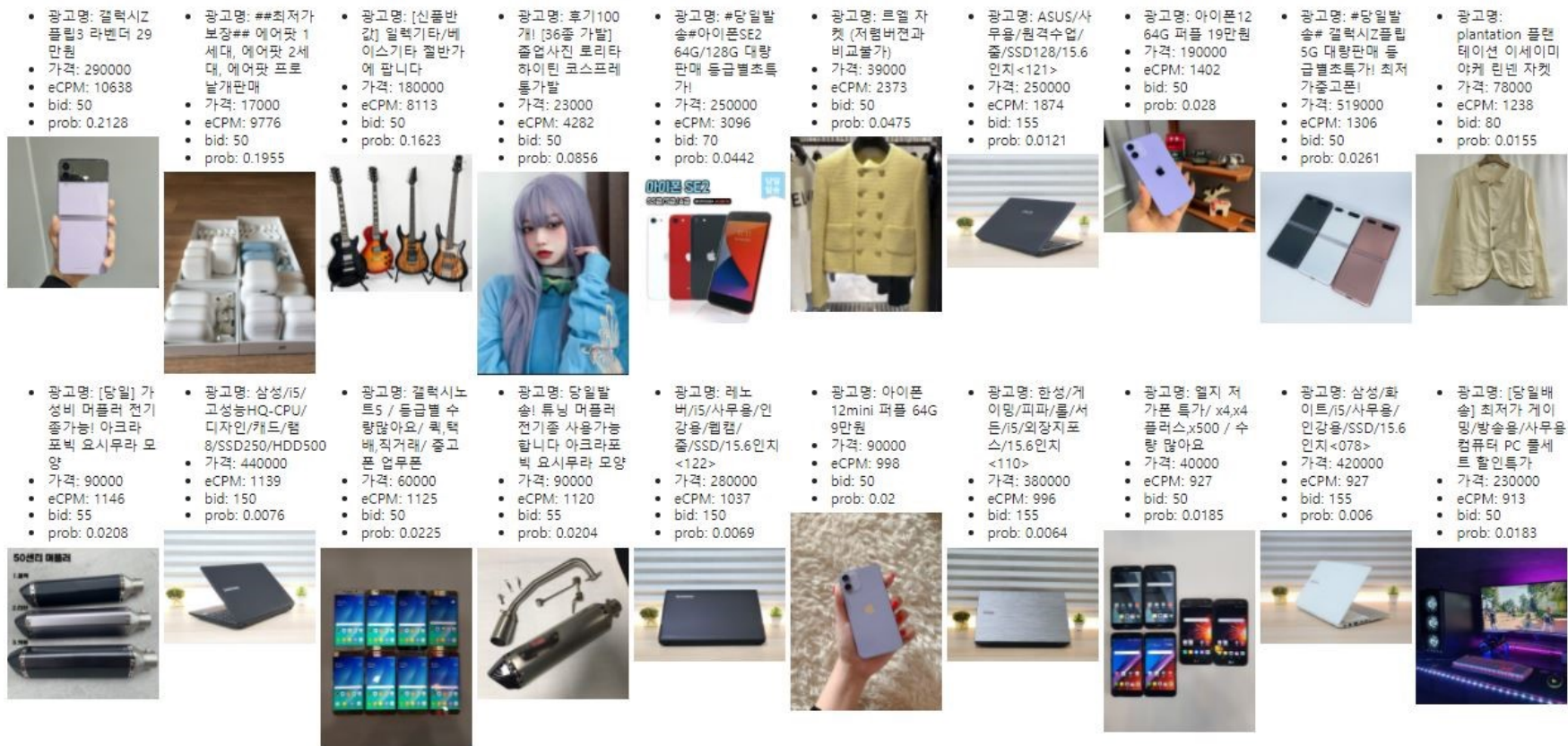
예상 결과물

1. 광고 로그 데이터를 학습하여 CTR을 예측하는 모델 개발 (LightGBM)
2. 가상의 유저(30세 여성, 30세 남성)를 설정하여 그 유저에게 어떤 광고를 추천해주었을때 기대수익(eCPM)이 높을지 예측해 봄
 - 30세 여성

<ul style="list-style-type: none"> 광고명: 허리리본 블루 원피스 F n8979 애슬리상점 가격: 12000 eCPM: 9855 bid: 50 prob: 0.1971 	<ul style="list-style-type: none"> 광고명: ##최저가 보장## 에어팟 1세대, 에어팟 2세대, 에어팟 프로 날개판매 가격: 17000 eCPM: 9315 bid: 50 prob: 0.1863 	<ul style="list-style-type: none"> 광고명: [신품반값] 일렉기타/베이스기타 절반가에 팝니다 가격: 180000 eCPM: 7654 bid: 50 prob: 0.1531 	<ul style="list-style-type: none"> 광고명: 오브제 스팅클장식 실크 원피스 드레스 66 n8991 애슬리상점 가격: 30000 eCPM: 7137 bid: 50 prob: 0.1427 	<ul style="list-style-type: none"> 광고명: 갤럭시Z 플립3 라벤더 29만원 가격: 290000 eCPM: 6024 bid: 50 prob: 0.1205 	<ul style="list-style-type: none"> 광고명: plantation 플랜테이션 이세이미 아케 린넨 자켓 가격: 78000 eCPM: 4144 bid: 80 prob: 0.0518 	<ul style="list-style-type: none"> 광고명: 트릴 자켓 (저렴버전과 비교물가) 가격: 39000 eCPM: 3161 bid: 50 prob: 0.0632 	<ul style="list-style-type: none"> 광고명: 정장12종 모음(사계절) 가격: 65000 eCPM: 2476 bid: 50 prob: 0.0495 	<ul style="list-style-type: none"> 광고명: 이중카라 배색 원피스 FF n8369 애슬리상점 가격: 30000 eCPM: 2013 bid: 50 prob: 0.0403 	<ul style="list-style-type: none"> 광고명: 나이키 맨투맨 티셔츠 S n8971 애슬리상점 가격: 10000 eCPM: 1934 bid: 50 prob: 0.0387 
<ul style="list-style-type: none"> 광고명: ASUS/사무음/원격수업/줄/SSD128/15.6인치<121> 가격: 250000 eCPM: 1644 bid: 155 prob: 0.0106 	<ul style="list-style-type: none"> 광고명: 레니본 화이트 원피스 54 n7151 애슬리상점 가격: 65000 eCPM: 1372 bid: 50 prob: 0.0274 	<ul style="list-style-type: none"> 광고명: #당일발송#아이폰SE2 64G/128G 대량 판매 특가별초특가! 가격: 250000 eCPM: 1349 bid: 70 prob: 0.0193 	<ul style="list-style-type: none"> 광고명: 럭키슈에드 원피스 36 n5162 애슬리상점 가격: 120000 eCPM: 1327 bid: 50 prob: 0.0265 	<ul style="list-style-type: none"> 광고명: 은은한플라워 플린피스 F n8986 애슬리상점 가격: 20000 eCPM: 1305 bid: 50 prob: 0.0261 	<ul style="list-style-type: none"> 광고명: COS 코스 그린색 니트 원피스 XS n8825 애슬리상점 가격: 45000 eCPM: 1282 bid: 50 prob: 0.0256 	<ul style="list-style-type: none"> 광고명: 당일발송! 튜닝 머플러 전기줄 사용가능합니다 아크라포빅 요시무라 모양 가격: 90000 eCPM: 1222 bid: 55 prob: 0.0222 	<ul style="list-style-type: none"> 광고명: 타임 브이덕 니트 F 90 n8009 애슬리상점 가격: 30000 eCPM: 1222 bid: 50 prob: 0.0244 	<ul style="list-style-type: none"> 광고명: 코스 A라인 원피스 S n8267 애슬리상점 가격: 30000 eCPM: 1181 bid: 50 prob: 0.0236 	<ul style="list-style-type: none"> 광고명: 255-260(40)정품 글든구스 슈퍼스타 가격: 99000 eCPM: 1159 bid: 115 prob: 0.0101 

남성에 비해 원피스등 의류 광고의 기대수익이 높은 것을 확인

- 30세 남성



여성에 비해 노트북, 핸드폰 등 전자기기의 비중이 높음

인사이트

- 남성에게는 전자기기 광고의 비중을, 여성에게는 의류 광고의 비중을 늘린다
- 또한 예측한 CTR을 바탕으로 판매자(광고주)에게 가이드라인 제시
 - 예1) 여성의 경우 제목이 짧은 광고의 클릭율이 높다
 - 원피스 판매자에게는 제목을 짧고 간결하게 쓰도록 가이드
 - 예2) 남성의 경우 이미지의 밝기가 밝을수록 클릭율이 높더라
 - 전자기기 판매자에게는 밝기가 높은 이미지를 사용하도록 가이드

3. 데이터 정의

데이터 설명

- 유형을 묶어 로그특성, 광고, 판매자, 유저로 구분
- 유형마다 content, adv, viewer를 앞에 넣어줌
- 크롤링을 통해 필요한 데이터 수집

유형별 설명

1. 로그특성

	imp_id	content_id	user_id	advertiser_id	label	bid_price	server_time_kst
0	99d1612d39112085014f	162892746	1310312	12851816	0	50	2021-08-31T05:02:38.277+09:00
1	971a612d3e480cf3d178	162892746	1310312	12851816	0	50	2021-08-31T05:24:14.109+09:00
2	971a612d3e480cf3d15a	162892746	1310312	12851816	0	50	2021-08-31T05:23:48.340+09:00
3	99d1612d3e6b208a041a	162892746	1310312	12851816	0	50	2021-08-31T05:24:20.169+09:00
4	99d1612d39112085012c	162892746	1310312	12851816	0	50	2021-08-31T05:01:30.062+09:00
...
806932	998c612e283901e36f4a	162990599	3800924	6874411	0	50	2021-08-31T22:01:53.052+09:00
806933	998c612e283901e36f68	162990599	3800924	6874411	0	50	2021-08-31T22:02:49.103+09:00
806934	9919612e39f723e66b02	162990599	7881402	6874411	0	50	2021-08-31T23:17:30.388+09:00
806935	99d1612e3dd6240ba416	162990599	11374311	6874411	0	50	2021-08-31T23:34:01.483+09:00
806936	97cf612e43f60bdce723	162336696	2797729	548278	0	50	2021-09-01T00:00:42.694+09:00

2. 광고

3. 크롤링

content_status	: 게시글 상태(판매중, 거래완료, 삭제)
content_b_pay	: 번개페이 여부(0: 없음, 1: 있음)
content_place	: 거래지역
content_text	: 광고 상품 설명
content_delivery_fee	: 배송비 여부(배송비 포함, 배송비 별도)
content_img_url	: 이미지 링크

	adv_follower_count	adv_grade	adv_item_count	adv_views	adv_review_count	adv_comment_count	adv_pay_count	adv_parcel_post_count	adv_tran
0	2131	5285	247	4525	542	739	496	0	
1	2131	5285	247	4525	542	739	496	0	
2	2131	5285	247	4525	542	739	496	0	
3	2131	5285	247	4525	542	739	496	0	
4	2131	5285	247	4525	542	739	496	0	
...
806932	49	349	165	1111	35	77	7	7	
806933	49	349	165	1111	35	77	7	7	
806934	49	349	165	1111	35	77	7	7	
806935	49	349	165	1111	35	77	7	7	
806936	73	30	25	648	4	45	0	3	

	viewer_gender	viewer_age	viewer_following_count	viewer_pay_count	viewer_parcel_post_count	viewer_transfer_count	viewer_chat_count
0	1	24	3	4	0	0	0
1	1	24	3	4	0	0	0
2	1	24	3	4	0	0	0
3	1	24	3	4	0	0	0
4	1	24	3	4	0	0	0
...
806932	2	29	23	0	25	0	0
806933	2	29	23	0	25	0	0
806934	0	0	0	0	0	0	0
806935	2	42	6	0	0	0	0
806936	0	0	3	0	0	0	0

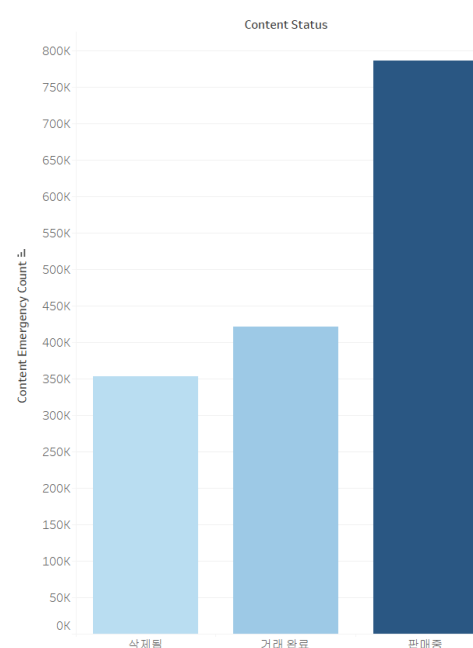
▼ merged data

imp_id	971a612ce6870be6fd33
content_id	116011108
user_id	12853149
adv_id	9698825
label	0
bid_price	50
server_time_kst	2021-08-31T01:17:37.287+09:00
content_name	갤럭시노트5 / 등급별 수량많아요/ 책,택배,직거래/ 중고폰 업무폰
content_keyword	갤럭시노트5,노트5,중고폰,업무폰,note5
content_price	60000
content_used	1
content_cat_1	전자제품
content_cat_2	600700
content_cat_3	600700001
content_emergency_count	1
content_comment_count	4
content_views	15503
content_likes	285
content_b_pay	1
content_delivery_fee	배송비 별도
content_place	경기도 성남시 분당구 서현1동
content_status	판매중
content_text	상품정보ㄴ✔️ 누보올레는 모든명품 최고가 매입ㄴ✔️타 업체 견적보다 만원 이라도...
content_img_url	https://media.bunjang.co.kr/product/116011108_...
adv_follower_count	719
adv_grade	2485
adv_item_count	830
adv_views	2068
adv_review_count	254
adv_comment_count	409
adv_pay_count	213
adv_parcel_post_count	0
adv_transfer_count	7
adv_chat_count	58
adv_title	중고폰요정
viewer_gender	1
viewer_age	39
viewer_following_count	0
viewer_pay_count	0
viewer_parcel_post_count	0
viewer_transfer_count	0
viewer_chat_count	0

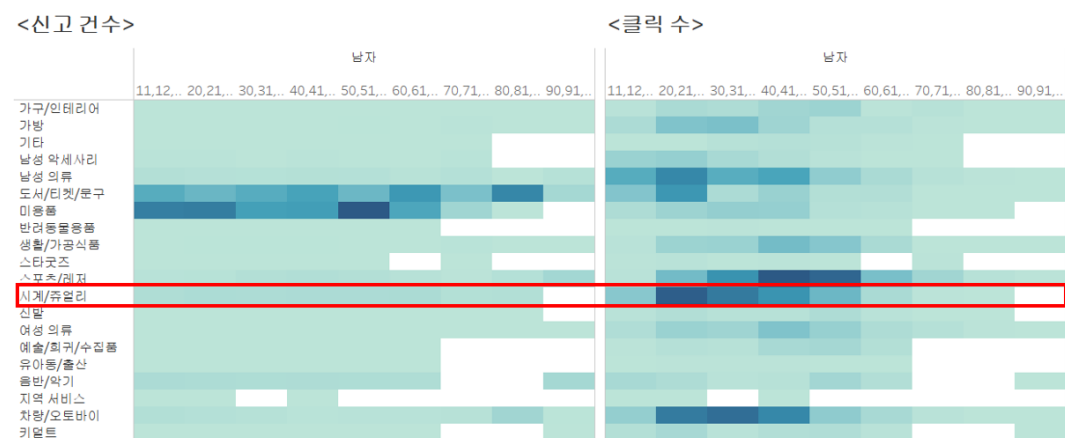
EDA

- Data 시각화를 통해 얻어낸 특이사항

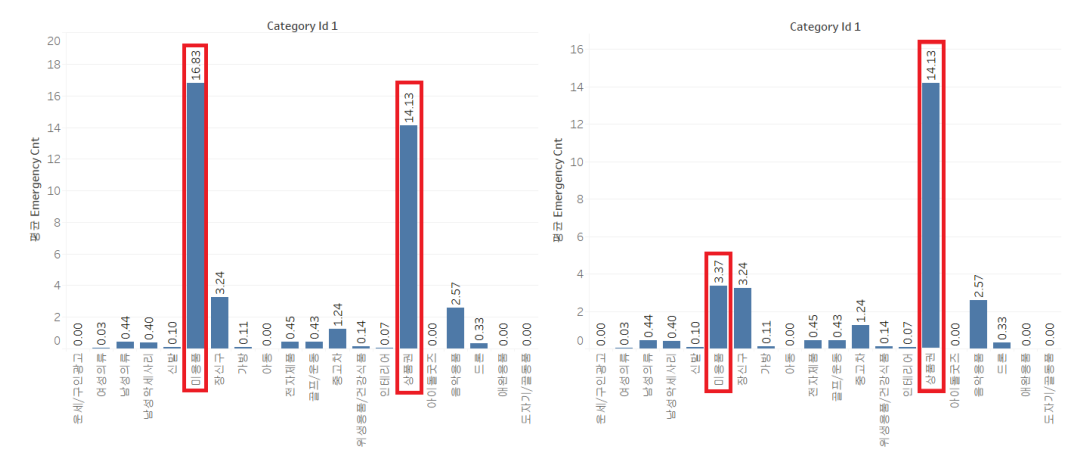
1. 신고건수가 거래완료인 상품보다 판매중인 상품의 수가 더 높음



2. 시계 / 주얼리 카테고리에 2~40대의 남성이 많은 관심을 보이지만, 신고의 수는 적음



3. 신고 수가 높은 카테고리는 즉각적인 사용(상품권) 또는 지속적인 효과(미용품)를 가진 제품이 많음



왼쪽 - 이상치 제거 전, 오른쪽 - 이상치 제거 후(신고 수가 압도적으로 높은 2건)

4. 기존 방법론 조사

번개장터의 광고 현황

- 번개장터의 광고체계는 외부광고, 내부광고로 나누어져 있음
- 내부광고는 광고 지면에 따른 **분류**, **과금방식**에 따른 분류로 나뉘짐
- 과금 방식에 따라 크게 두가지의 광고 상품이 있음
 1. 키워드 베이직
 - CPT : 시간 단위 과금, 고정된 광고비로 간편하게 집행
 - 노출 지면에서 키워드 프리미엄 다음에 노출
 2. 키워드 프리미엄
 - CPC : 1클릭당 광고비가 발생
 - 노출 지면 상단에 노출
 - 광고 타겟(연령, 성별, 지역) 선택, 관련 광고 키워드 관리 가능
- 노출 순위 : CTR, 키워드 적합도, 예산(bid)를 고려하여 결정

Click Prediction

- CTR(Click Through Rate), 광고를 클릭할 확률을 예측하는 것
- Click Prediction을 하는 이유?
 - 고객에게 매력적인 광고 상품을 적절하게 노출할 수 있음
 - eCPM(1000회 노출 당 광고비)를 추측해 광고의 효율을 명확하게 측정할 수 있음
- 번개장터에서는 LightGBM을 사용하여 CTR 예측 (2020년 여름 기준)
 - 종합적인 지표(성능, 속도)에서 Logistic Regression < 그외 앙상블(Xgboost, RandomForest) < DeepFM ≤ LightGBM 순이었다고 함

▼ Reference

5. 사용 모델 리스트

CTR 예측 모델

1. LightGBM

- Tree 기반(GBDT) 머신러닝 알고리즘
- 실제 번개장터에서 사용중인 모델이라고 함

2. DeepFM

- 딥러닝 기반으로 CTR을 예측하는 모델
- Kaggle의 CTR Prediction에서 좋은 결과를 보여줌
- 번개장터에서는 계산 속도 때문에 머신러닝 기반의 LightGBM을 사용하지만 성능면에서는 DeepFM이 더 좋을 것이라고 추정
- 두 모델을 모두 적용해보고 성능, 속도가 얼마나 차이 나는지 비교하여 우리에게 알맞는 모델 선택할 예정

평가 지표

- AUC (Area Under ROC)
- RIG (Relative Information Gain)

그 후 인사이트

- CNN : 광고 이미지 분석
 - RNN : 광고 제목, 설명 및 키워드 분석
 - 광고 노출 로그외의 데이터가 부족하기 때문에 주로 미리 학습된 모델을 가져와서 사용할 예정
-

6. 프로젝트 환경

- a. Git-hub을 작업 내용 백업
- b. Colab, Jupyter-Notebook을 통해 작업 내용 공유
- c. Notion을 통해 진행 상황 공유 및 기획서 작성
- d. data는 무겁거나, 보안 등의 문제 Google Drive로 공유