

신한AI, 보다 나은 금융 생활을 위한 AI 서비스 아이디어 경진대회

# 우리들의 일상을 투자로! STO 조각투자 시작하자

인공지능을 활용한 콘텐츠 토큰증권 투자 플랫폼 제안

러닝파이브



# INDEX

01. 아이디어 배경

02. 아이디어 소개

03. 데이터 수집 및 EDA / 전처리

04. 모델링

05. 성능 검증

06. 결론

## STO(Security Token offering)란?

부동산, 미술품 등 실물 및 금융 자산을 블록체인의 기반의 토큰에 연동한 디지털 자산인 증권토큰으로 발행하는 것

암호화폐 자금 조달 방식(ICO)의 새로운 대안으로,  
실제 자산가치를 담보로 발행하며  
각 국가의 제도권 법규 즉 증권법이나 자금관련 법의 통제를 따름

투명성, 안정성, 유동성이 장점

# STO

## 2023년 2월5일 STO 가이드라인 제시 이후 변화하는 증권 시장

### 금융위가 내놨다 'STO 가이드라인'...상반기 제도화

토큰증권(STO)에 관한 구체적인 가이드라인이 나왔다. 금융위원회는 5일 “자본시장법 규율 내에서 STO를 허용하겠다”라며 이를 위한 토큰 발행·유통·...

2023. 2. 5.

### STO 허용하자 증권사들 바빠졌다...시장 선점 속도

금융위원회가 STO(증권형 토큰)를 전면 허용하면서 새로운 시장이 열리게 됐다. STO는 블록체인 기술(스마트 계약)을 적용해 실물 자산과 연동...

2023. 1. 23.

### “집·한우·음악·그림 쪼개 팝니다”...STO 르네상스 열린다

코인보다 안전하고, 주식·부동산보다 새롭다.” '토큰 증권(STO·Security Token Offering)'이 여의도 증권가의 관심 키워드로 부상하고 있다.

2023. 2. 6.

### STO 가이드라인 직후 발 빠르게 움직이는 증권사들

머니투데이 | 이두리 기자

### 누가 선점할까...증권사 STO 동맹 '속도'

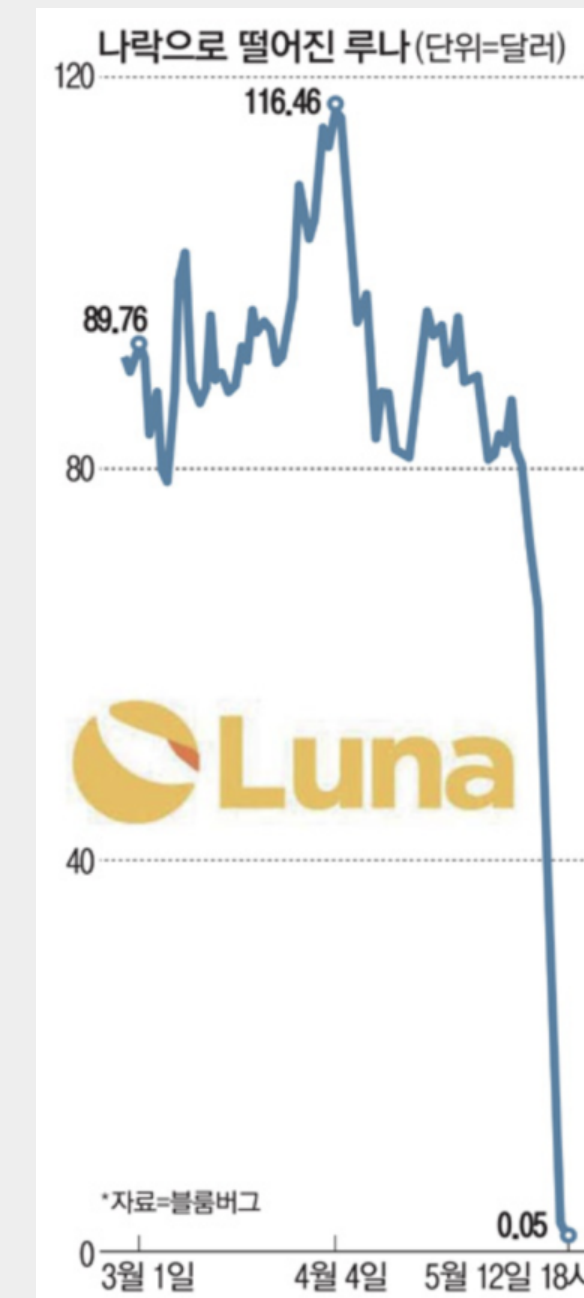
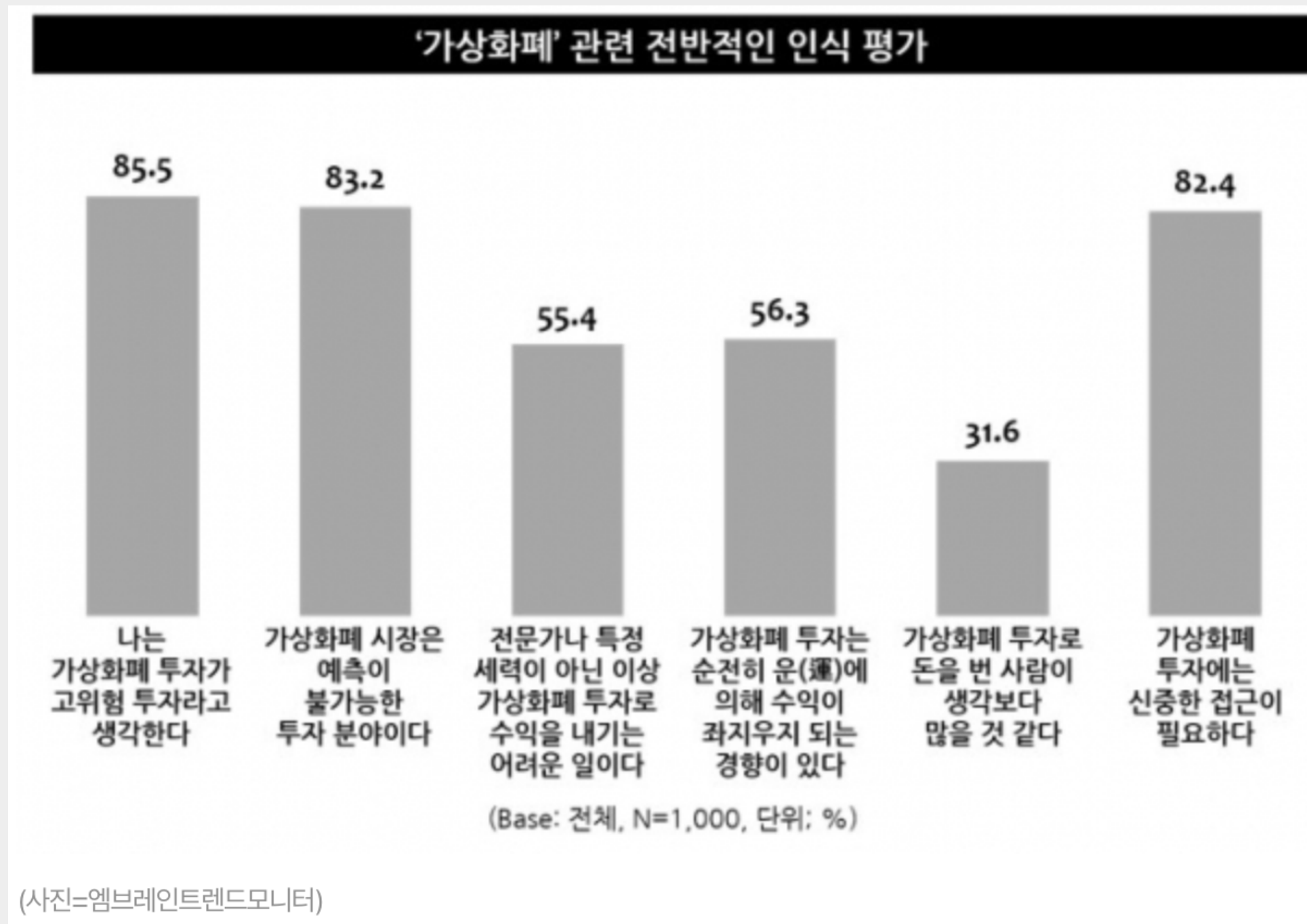
에너지경제신문 | 입력 2023.04.23 10:26

음악이나 영화도 STO로 조각투자 가능!



# STO

규제 밖의 가상화폐, 암호화폐의 인식과 문제



**'영끌' 실패한 2030의 잇단 극단선택...전문가 "이건 '사회적 타살'"**

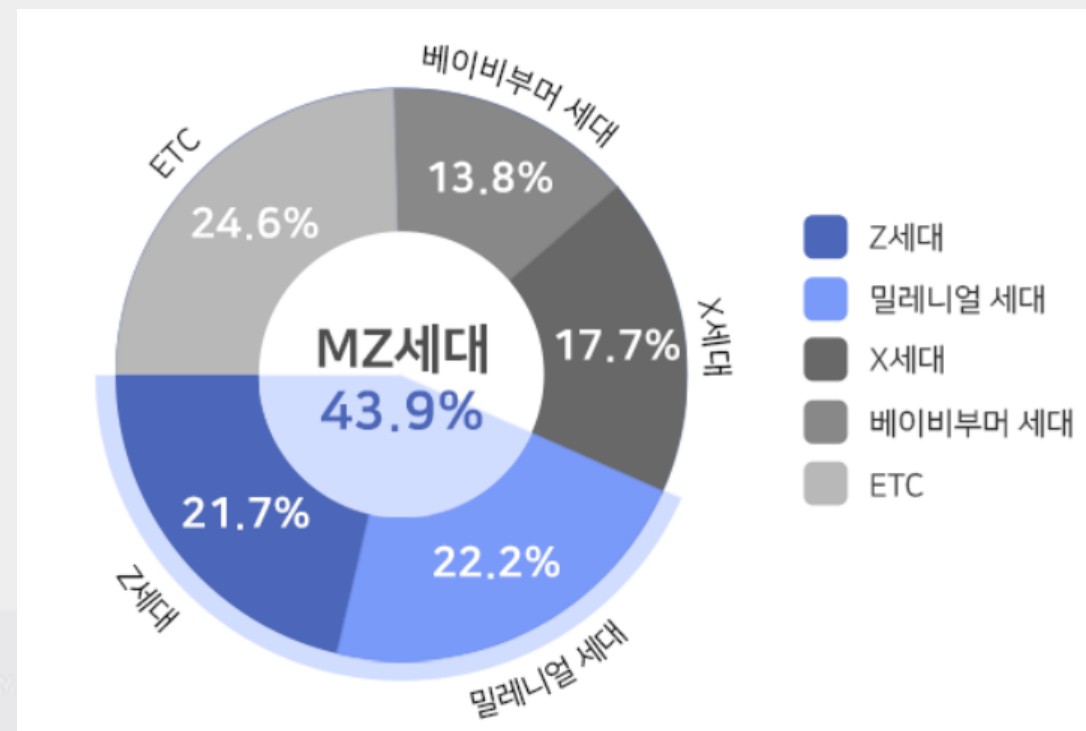
가상화폐와 주식에 투자했다가 실패해 극단적 선택을 하는 사례가 이어지고 있다. 전문가들은 개인회생 등 주식·코인 투자 실패의 아픔을 딛고 재기...

**무너지는 비트코인, 암호화폐 제국의 위기**

MZ

밀레니얼 세대(1980~2004 년 출생)와 Z세대(1995~2004 년 출생)를 통칭하는 용어  
이전 세대에 비해 재테크와 금융 서비스에 높은 관심

신규 고객 확보를 위한  
10대~MZ세대 고객 타케팅 필요



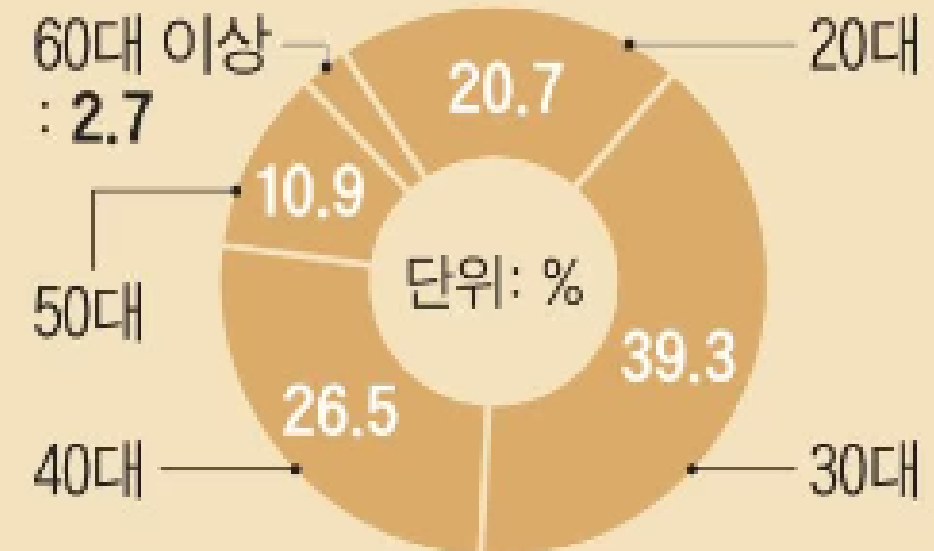
### MZ세대 모시는 금융사

캠퍼스 공략	신한은행, 대학생 전용 모바일 통합 앱 '헤이영 스마트 캠퍼스'
Z세대 집중 타킷	카카오뱅크 '미니', 하나은행 '아이부자 앱', 국민은행 '리브 넥스트'
취향 저격 마케팅	우리은행, MZ마케팅팀 신설...MZ 세대 특화 상품, 콘텐츠 개발
	신한라이프, 금융권 처음으로 가상인간 '로지' 모델로 기용
금수저 MZ 유입	신한금융그룹, KB금융그룹 '패밀리오피스' 서비스 본격화

MZ세대 중심으로  
디지털 플랫폼 기반의 재테크가 확산

### 조각투자 연령대별 비중

강남빌딩 (카사)

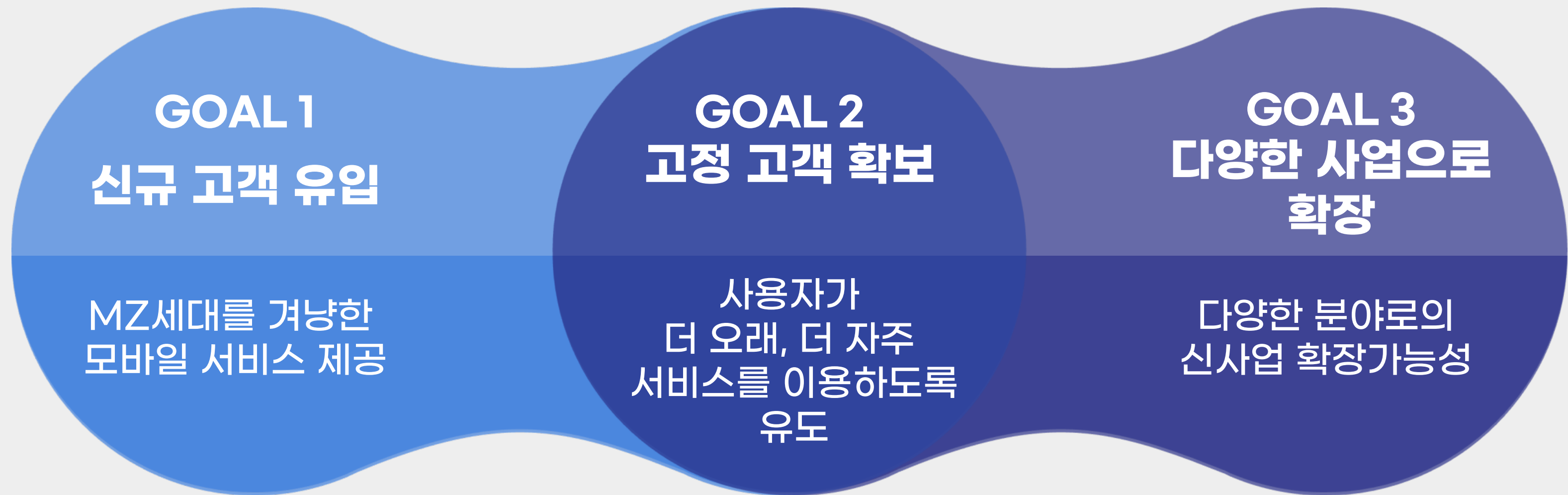


※1, 2호 공모건물 투자자 1만112명 기준

## 우리들의 일상을 투자로! STO로 조각투자하자

딥러닝을 활용한 토큰증권 콘텐츠 투자 플랫폼 제안

은행과 본격적인 거래를 시작하는 MZ세대가 특정 은행에서 금융·비금융 콘텐츠로 재미를 경험하고  
좋은 인상을 받으면서 자연스럽게 주거래 고객으로 남게 될 것





# 기대효과

## 블록체인 기반 토큰형 증권의 기대효과

- ✓ 해당 자산의 매매과정속 여러 거래 비용 절감 가능
- ✓ 위조 변조 불가능, 거래정보 실시간 투명하게
- ✓ 가상화폐와 같은 위험한 투자에서 보다 안전한 투자로
- ✓ 부동산이나 미술품 등 다양한 분야도 거래 가능
- ✓ 투자금액이 적은 투자자들도 저렴한 비용으로 조각 투자 가능



## 기대효과

트렌드에 민감한 MZ를 타겟으로!

연령층이 낮은 투자자가 늘어남에 따라 STO를 통해 투자 접근성을 높여  
시장 전체를 활성화 시킴



MZ 세대는 보유자산은 적고 투자에 관심이 많아 분할투자가 적합

재미있는 투자 방식으로 신한투자증권의 발빠른 시장 선점

부동산, 미술품등 다양한 방면으로 확장 가능

# 데이터 수집 : 데이터 선정이유

### 가온차트

### 가온차트

국내 주요 음악 플랫폼과 이동통신사에서 제공하는 데이터에  
가중치를 두고 순위를 계산한 차트  
스트리밍, 다운로드, BGM 판매량 등에 가중치를 부여한  
가온 디지털 차트를 종속 변수로 사용

### 멜론

### 멜론

국내 음원 플랫폼 중 가장 많은 월간 활성 이용자수를 보유 \* 23년 1월 기준  
멜론의 음악 좋아요 수, 댓글 수, 장르, 가수, 발매일 데이터 수집

# 데이터 수집 : 데이터 선정이유



KOBIS

## KOBIS 박스오피스 영화정보

매월 업데이트되는 영화진흥위원회가 운영하는 KOBIS 영화관 입장권통합전산망 데이터로 영화명, 감독, 제작사, 수입사, 배급사, 개봉일 등 정보 제공하는 신뢰성 있는 데이터일 것이라고 판단



KOBIS

## KOBIS 개봉일람

KOBIS 박스오피스 영화정보와 동일하게 KOBIS 영화관입장권 통합전산망 데이터로 동일한 컬럼으로 구성되어 있으며 1971년부터 2023년까지의 데이터로 KOBIS 박스오피스 영화정보보다 많은 양의 데이터

## 종속 변수 선정 - 음악

	singer-music	rank_count
0	방탄소년단-봄날	324
1	폴킴-모든 날, 모든 순간	268
2	akmu-어떻게 이별까지 사랑하겠어, 널 사랑하는 거지	171
3	아이유-밤편지	169
4	임영웅-이제 나만 믿어요	161
5	방탄소년단-작은 것들을 위한 시	160
6	마크툽-오늘도 빛나는 너에게	148
7	annemarie-2002	145
8	아이유-blueming	144
9	폴킴-너를 만나	143
10	방탄소년단-dynamite	142

스트리밍, 다운로드, BGM 판매  
량 등에 가중치를 부여한  
가온 디지털 차트에서

해당 곡이 차트에 진입한 횟수를  
선정

차트 진입 횟수가 많을 수록  
흥행한 곡이라 판단

종속 변수 선정 - 영화

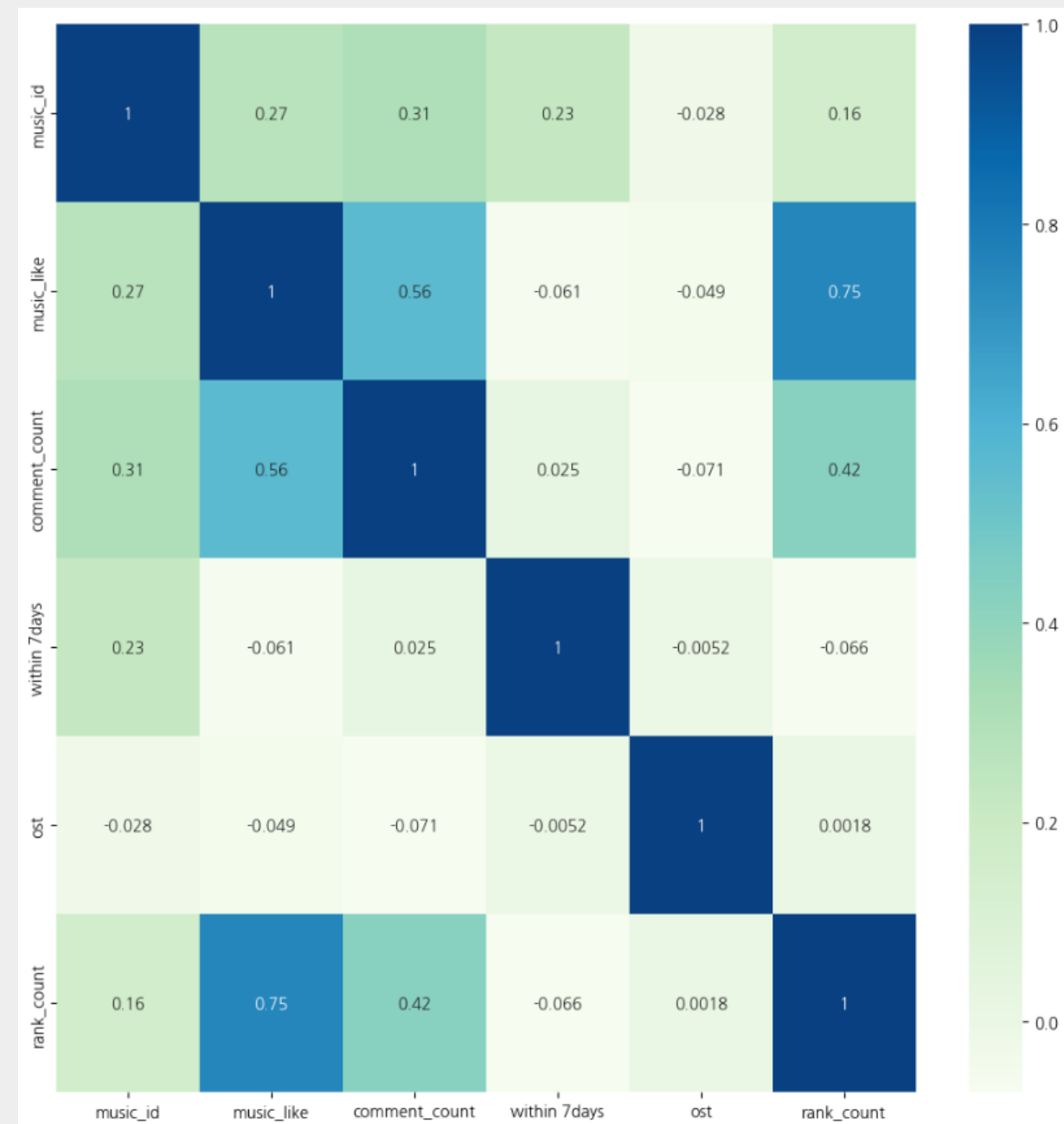
순위	영화명	개봉일	매출액 ▲ ▼	관객수 ▲ ▼
1	명량	2014-07-30	135,758,658,810	17,615,919
2	극한직업	2019-01-23	139,655,543,516	16,266,338
3	신과함께-죄와 벌	2017-12-20	115,727,528,087	14,414,658
4	국제시장	2014-12-17	110,947,621,230	14,264,478
5	어벤져스: 엔드게임	2019-04-24	122,492,181,020	13,977,602
6	겨울왕국 2	2019-11-21	114,843,932,550	13,750,668
7	베테랑	2015-08-05	105,170,856,250	13,414,484
8	아바타	2009-12-17	125,304,346,000	13,338,863
9	도둑들	2012-07-25	93,672,497,700	12,984,701
10	7번방의 선물	2013-01-23	91,437,368,170	12,812,144

1. 우리나라의 경우 박스오피스 순위를 매길 때 기준은 관객수

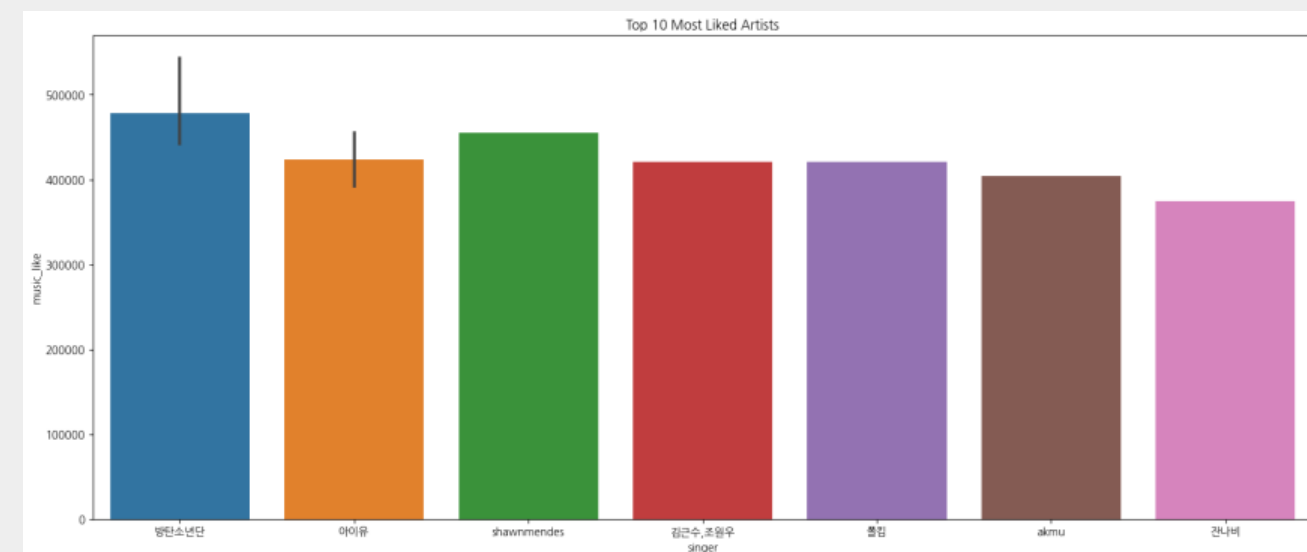
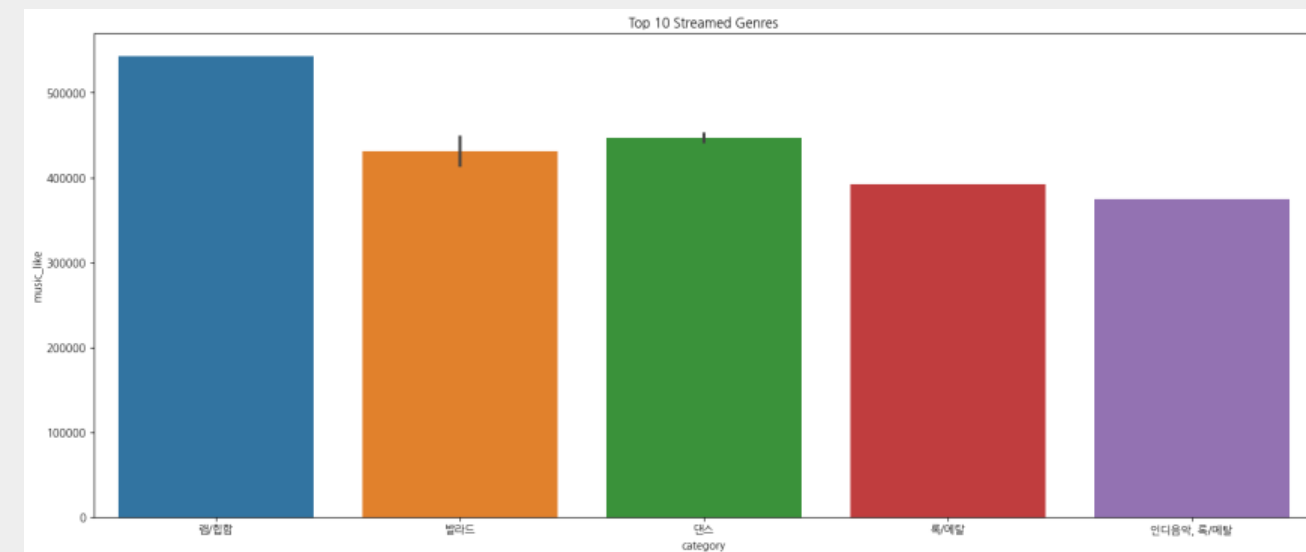
2. 영화진흥위원회 영화관입장권 통합전산망(KOBIS)에선 관객 수와 매출액을 동시에 표시하고 있으나 주된 기준은 관객수

## 모델 간의 상관관계 및 값 확인 - 음악

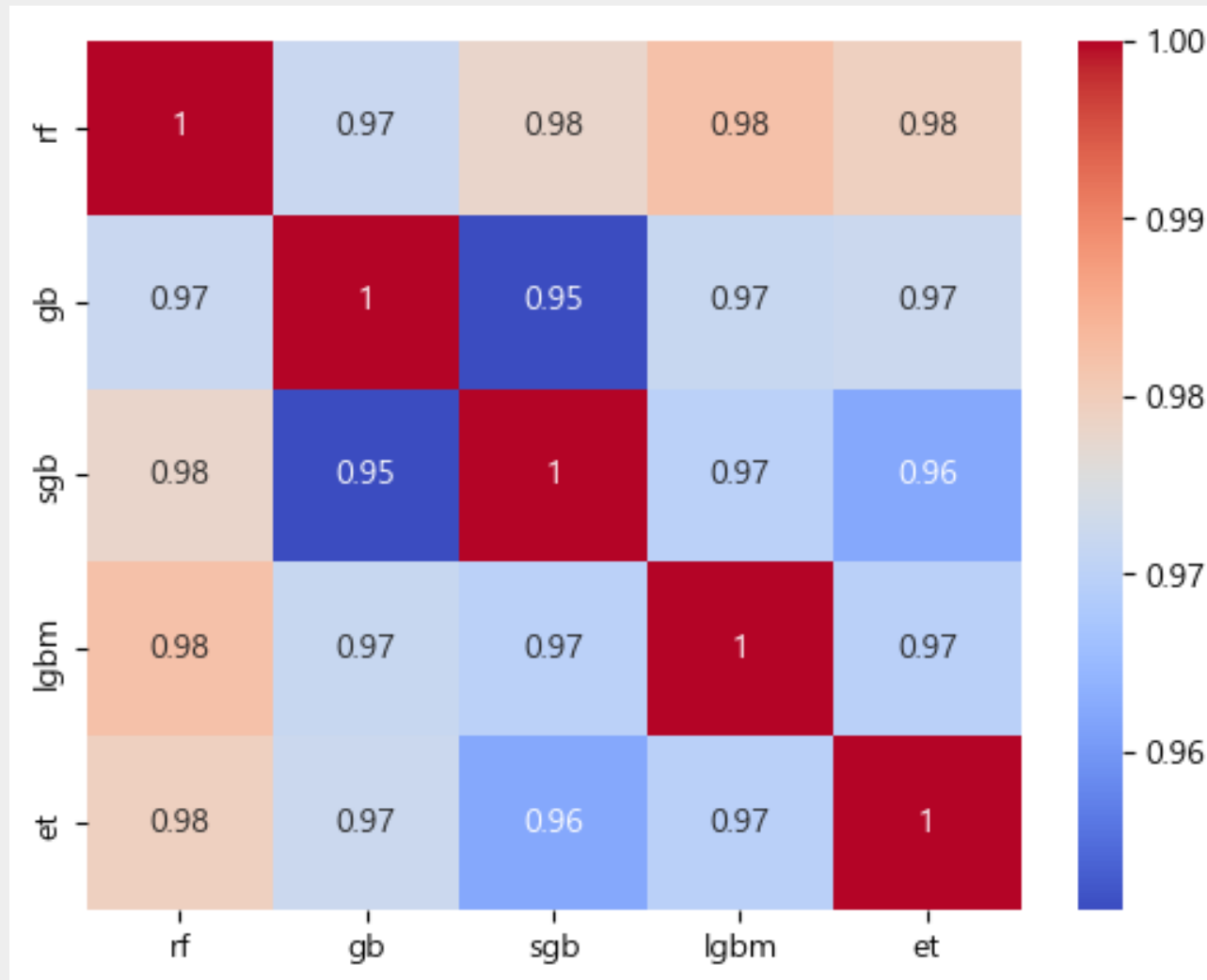
### 데이터 피처별 상관관계 시각화



### 상위 데이터 값 확인



### 모델 간의 상관관계 - 영화



01 과적합

02 높은 상관관계



## 피쳐 엔지니어링 - 음악

### ✓ 피쳐 삭제

- 불필요한 데이터 삭제 (유니크한 값)

### ✓ 데이터 정제

- 음악 데이터 정규표현식을 통해 특수문자, 공백 제거

### ✓ 데이터 벡터화

- 가수 이름 'singer' 벡터화 - word2vec 사용

### ✓ 데이터 인코딩

- 'category' 데이터 원핫 인코딩

### ✓ 타겟 변수 생성

가온차트 랭킹에 든 횟수(rank\_count)를 종속변수로 설정

	count	mean	std	min	25%	50%	75%	max
music_id	4385.0	2.033301e+07	1.336602e+07	38426.0	5546828.0	30351580.0	32034629.0	36414423.0
music_like	4385.0	5.355663e+04	6.289172e+04	3.0	12176.0	30131.0	70763.0	542966.0
comment_count	4385.0	4.450392e+02	1.044317e+03	0.0	30.0	127.0	408.0	19203.0
within 7days	4385.0	8.424173e-01	3.643908e-01	0.0	1.0	1.0	1.0	1.0
ost	4385.0	9.486887e-02	2.930672e-01	0.0	0.0	0.0	0.0	1.0
rank_count	4385.0	9.903079e+00	1.789958e+01	1.0	1.0	3.0	10.0	324.0

### ✓ 설명변수 생성 : 발매일 기준 순위권 유무

- 멜론의 발매일 데이터와 가온차트의 주간차트 시작 일자를 비교하여, 발매일 기준 7일 이내에 순위에 들었으면 1, 아니면 0

### ✓ 설명변수 생성 : 국내드라마 OST

- 'category'에서 '국내드라마' 키워드를 포함하고 있을 경우 1, 아니면 0

## 피처 엔지니어링 과정 - 영화

### 중복값, 표준화

중복값 제거, 표준화를 통해 같은 내용인데  
다르게 표현된 값 통일

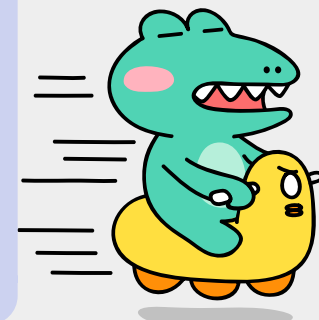
### 결측치, 이상치

크롤링 후 결측치 제거,  
이상치는 정확한 값이기에 보류

### 수치화, VIF

관객수 평균값으로 수치화 및 더미변수로  
변환, VIF 및 조건수로 다중공선성 확인

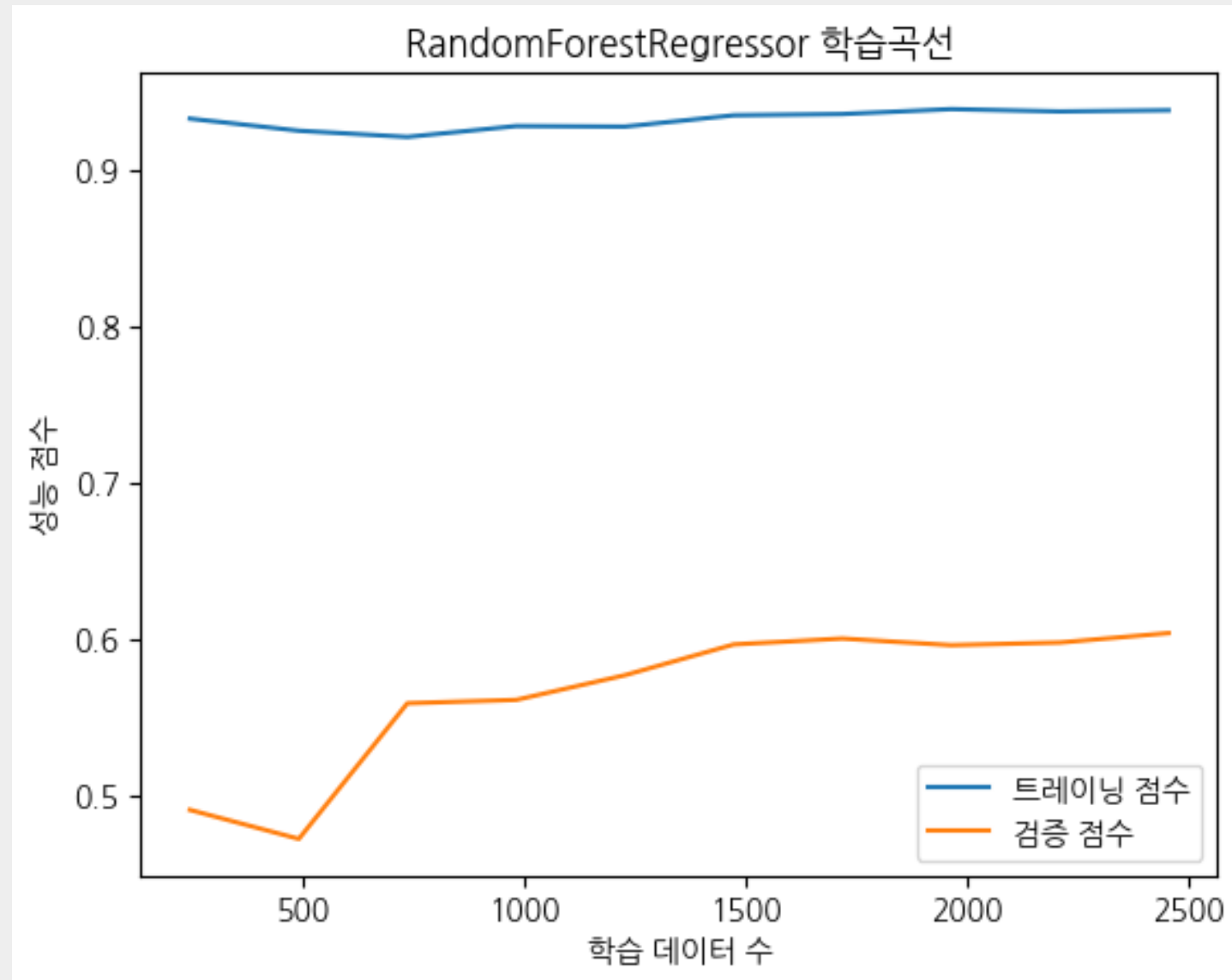
간드아...



## 모델링 과정



## 모델 : RandomForest - 음악 데이터



Training - 0.94

Test - RMSE : 8.87

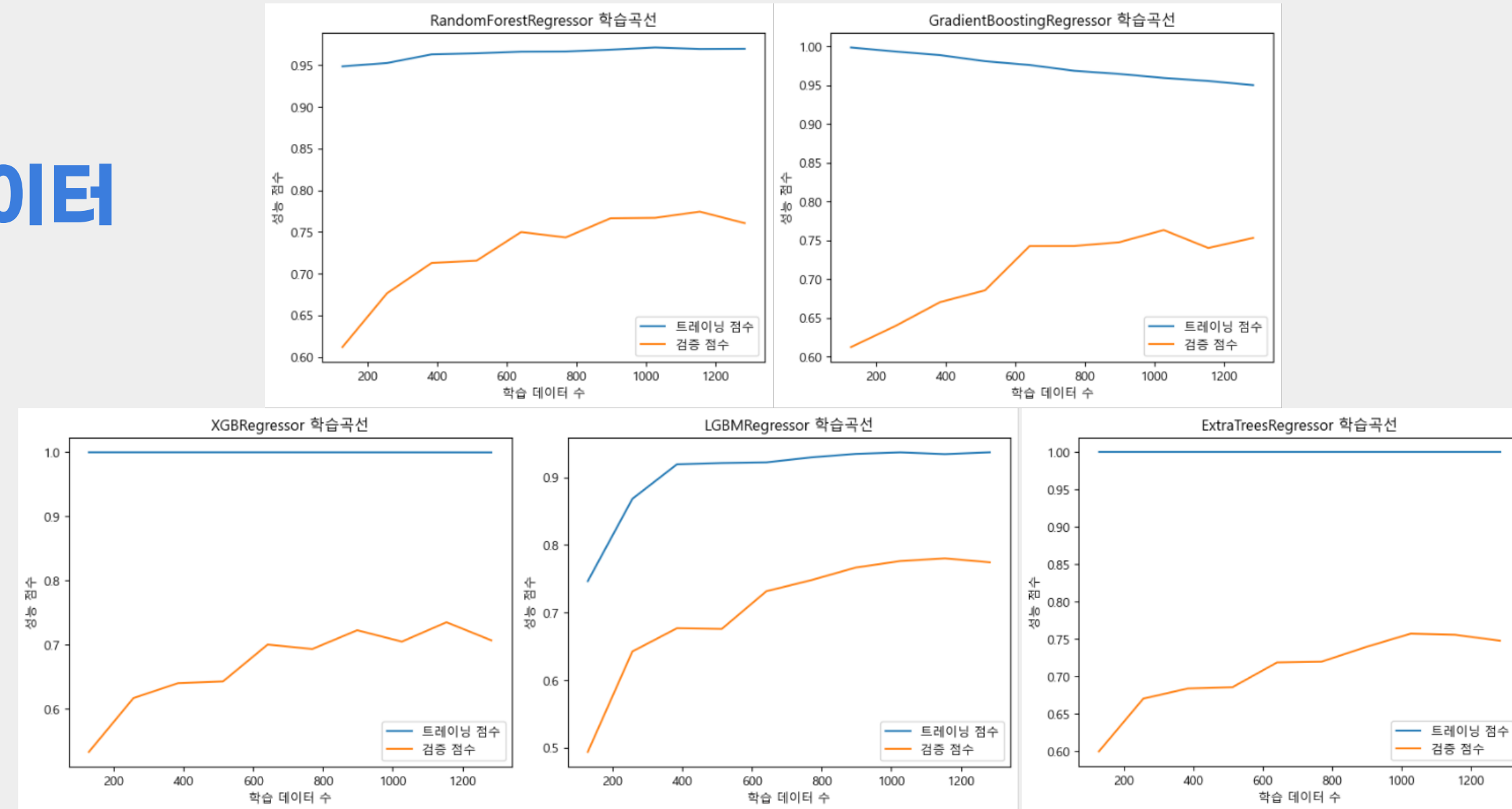
MAE 4.41

R - squared : 0.71

성능 검증표 - 영화 데이터

	RandomForest	GradientBoosting	XGB	LGBM	ExtraTrees	StackingCV
선택 이유	좋은 예측 성능, 과 적합 방지	이상치에 안정적인 성능	좋은 예측 성능	빠른 속도, 좋은 예 측 성능	높은 정확도와 일반 화 성능	모델 간의 높은 상관 관계
Optuna 전	RMSE: 714387 MAE: 282635 r^2: 0.8269	RMSE: 789027 MAE: 370771 r^2: 0.7889	RMSE: 746791 MAE: 299637 r^2: 0.8109	RMSE: 729075 MAE: 299760 r^2: 0.8197	RMSE: 761752 MAE: 324064 r^2: 0.8032	RMSE: 810882 MAE: 314642 r^2: 0.7770
Optuna 후	RMSE: 713014 MAE: 280513 r^2: 0.8276	RMSE: 770063 MAE: 304605 r^2: 0.7989	RMSE: 754912 MAE: 346434 r^2: 0.8067	RMSE: 722162 MAE: 324167 r^2: 0.8231	RMSE: 750173 MAE: 329888 r^2: 0.8091	RMSE: 726141 MAE: 287707 r^2: 0.8212

## 학습 곡선 - 영화 데이터



**01** 높은 트레이닝 점수

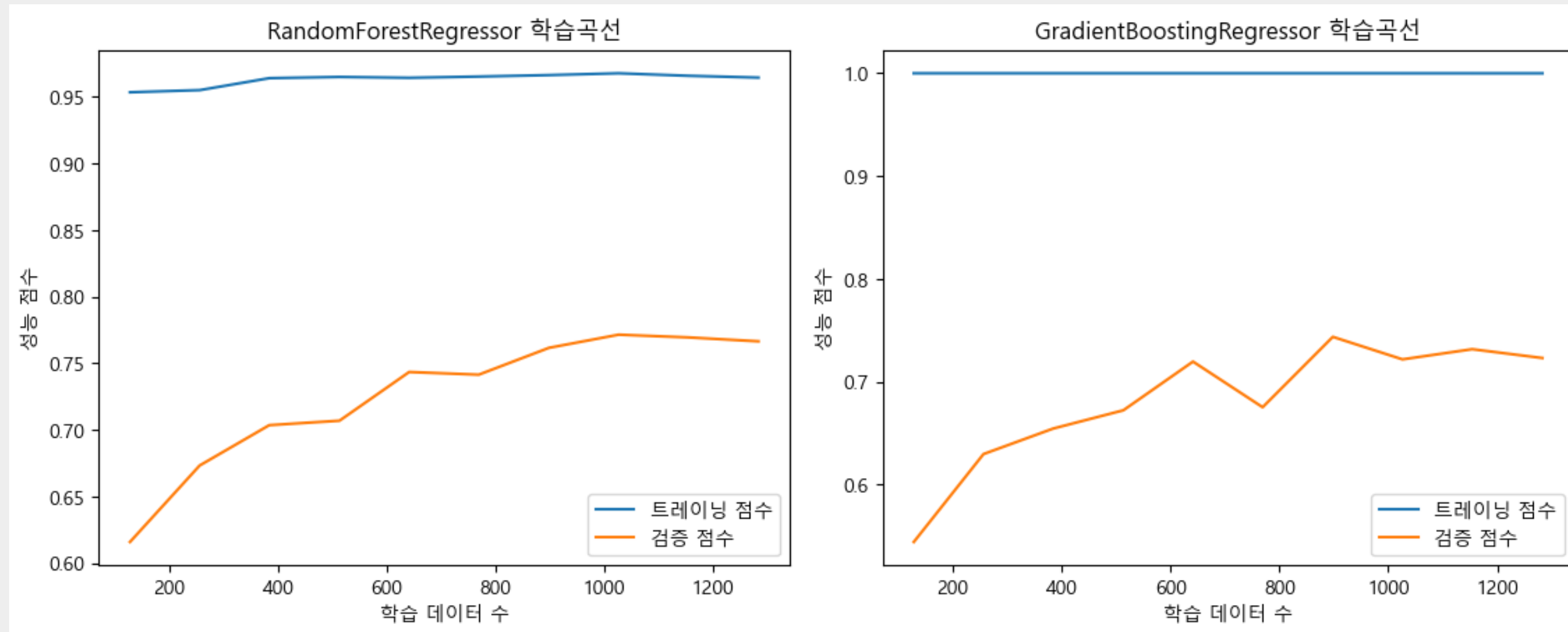
**02** 트레이닝 점수와 검증 점수의 차이

**03** 검증 점수 감소 구간 존재

**결론도출,**

**과적합**

## 튜닝 후 학습 곡선 - 영화 데이터



**01** 높은 트레이닝 점수

**02** 트레이닝 점수와 검증 점수의 차이

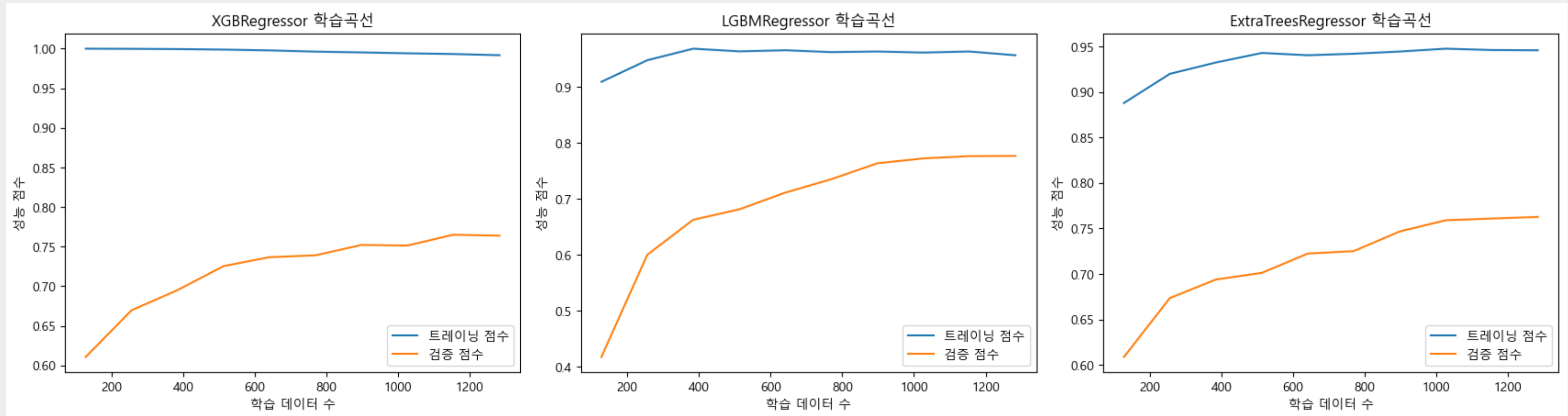
**03** 검증 점수 감소 구간 존재

**결론도출,**

**과적합**



## 튜닝 후 학습 곡선 - 영화 데이터



01 높은 트레이닝 점수

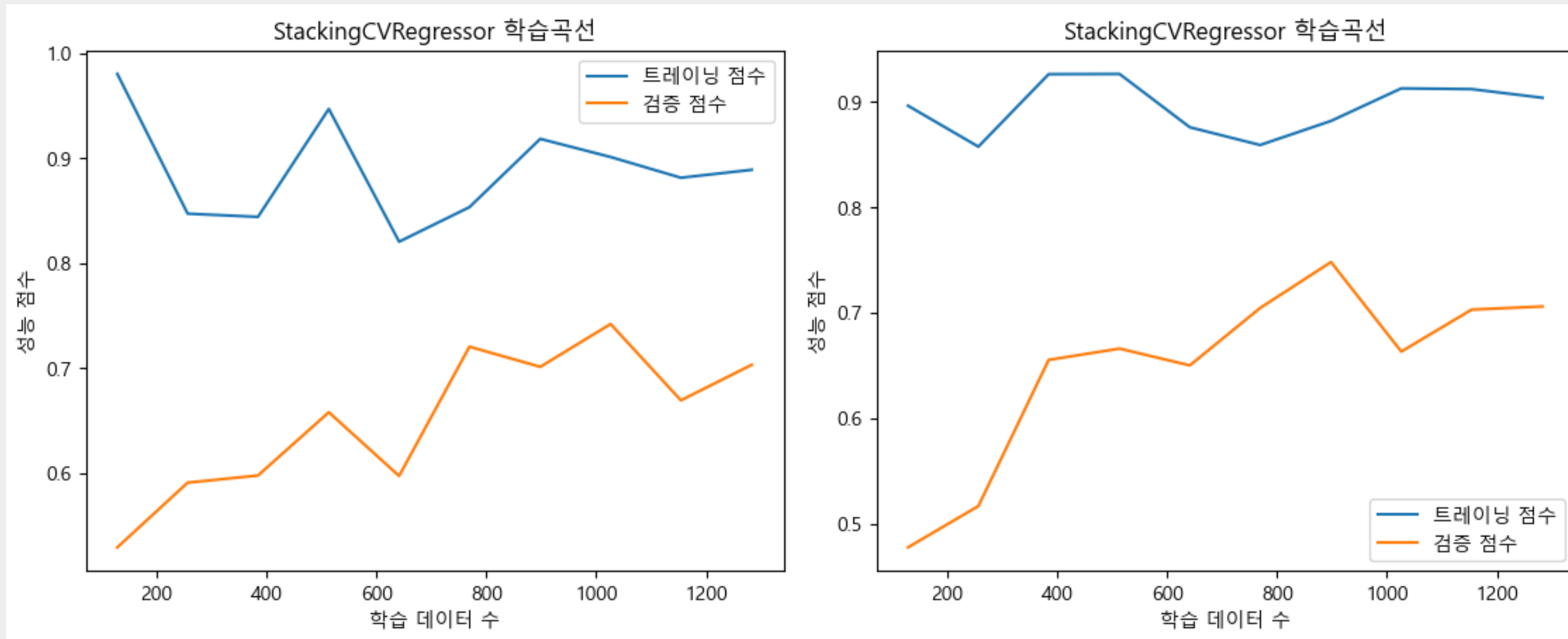
02 검증 점수 향상

03 트레이닝 점수와 검증 점수의 차이

결론도출,

과적합

## 튜닝 후 학습 곡선 - 영화 데이터



01 트레이닝 점수 감소 구간 존재

02 검증 점수 감소 구간 존재

03 트레이닝 점수와 검증 점수의 차이

결론도출,

과적합

# 결론

## 음악 STO 추천을 위한 음악 흥행 예측 모델 개발

정확도를 더 높이기 위해 다양하고 많은 양의 데이터를 통해  
새로운 앙상블 모델을 구현하거나 하이퍼파라미터를 조정하여 더 최적화  
된 모델을 구현해야 함

가수의 SNS 팔로우 수, Youtube 티저 영상 조회수, 포털사이트 노래 사  
전 홍보기사, 소속사나 작곡/작사가 역량 등  
더 다양한 데이터를 확보 할 수 있음

## 결론

### 영화 STO 추천을 위한 영화 관객 수 예측 모델 개발

다양한 모델과 스택킹 앙상블을 시도하였으나 랜덤 포레스트의 성능이 가장 높았다  
과적합 문제를 해결하려 하이퍼파라미터 튜닝과 앙상블을 시도하였으나 효과는 미미

## 한계점

데이터 전처리  
결측치가 있는 데이터를 대부분 삭제했음, 이상치 보류

모델의 활용  
모델이 가진 성능을 제대로 활용 못 함 ex)GB, XGB의 변수 중요도, LGBM의 조기종료

## 향후 방향

시장 선점, 다양한 분야로 확장

쉽게 다가갈 수 있는 분야로 MZ뿐만 아니라 다양한 연령층의 고객 확보

STO시장을 선점 할 수 있는 플랫폼 개발

음악이나 영화뿐 아니라 드라마, 예능, 부동산, 미술품, 한우 등

자금조달이 필요한 기업이나 스포츠팀

개인의 아이디어나 기술까지

다양한 분야의 증권토큰발행 가능

출처) <https://newsroom.koscom.co.kr/28812>

출처) <https://news.mt.co.kr/mtview.php?no=2022030714291436751>

출처) <https://www.chosun.com/economy/stock-finance/2021/09/07/7ZTV7XMG55F6FMJC6JS7D7R7W4/>

영화 관객 수 예측을 위한 기계학습 기법의 성능 평가 연구

<https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=JAKO202016863204479&dbt=NART>

영화 흥행 결정 요인과 흥행 성과 예측 연구

<https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=JAKO201105462029187>