

영화검색

검색

## NLP를 활용한 거짓 리뷰 검출을 위한 네이버 영화 평점 anticheat

TEAM 리얼리뷰 ★★★★★ 10

박정은, 김민아, 이재인

# Content

1. 문제 정의
2. 프로젝트 목표
3. 분석 기획 및 사용 스택
4. 데이터 수집
5. EDA 및 전처리
6. Data Mining
7. 결과 및 한계

# 1. 문제 정의

“네이버 영화 리뷰는 믿을 수 없다!”

- 콘텐츠 관련 서비스에 대한 **고관여 소비자** 증가 추세
- BUT 채널에서 광고, 홍보 목적 혹은 악의적으로 **거짓된 리뷰를 생산하는 사람들이** 잔존
- 그 때문에 **네이버 영화의 경우 신뢰도가 매우 낮음**
- 영화 데이터베이스가 넓은 편임에도 **영화 관련 정보 통합 플랫폼으로서의 역할을 못하며** 영화 선택에 있어서 판단요소로 작용을 하지 못하고 있음
- 추가적으로 네이버 **시리즈온 유입 및 매출 증대의 기회를 잃고 있음**

## 2. 프로젝트 목표



네이버 영화 리뷰 속 거짓리뷰 판별을 통한 조작 평점을 제외한  
‘진짜’ 평점 도출 과정 설계

## 2. 프로젝트 목표

prot****님의 평점 리스트			총 41개의 평점이 있습니다.
번호	감상평	글쓴이·날짜	
17831466	디어 에반 헨슨 ★★★★★ 10 재미있게 잘 봤습니다. *^^* 신고	prot**** 21.12.04	
17831465	연애 빠진 로맨스 ★★★★★ 10 재미있게 잘 봤습니다. *^^* 신고	prot**** 21.12.04	
17831464	유체이탈자 ★★★★★ 10 재미있게 잘 봤습니다. *^^* 신고	prot**** 21.12.04	
17810639	장르만 로맨스 ★★★★★ 9 재미있게 잘 봤습니다. *^^* 신고	prot**** 21.11.21	
17807215	베놈 2: 렛 데어 비 카니지 ★★★★★ 10 재미있게 잘 봤습니다. 신고	prot**** 21.11.20	
17807211	상치와 텐 링즈의 전설 ★★★★★ 10 재미있게 잘 봤습니다. 신고	prot**** 21.11.20	
17807208	고장난 론 ★★★★★ 10 재미있게 잘 봤습니다. 신고	prot**** 21.11.20	
17807206	강릉 ★★★★★ 10 재미있게 잘 봤습니다. 신고	prot**** 21.11.20	
17807200	라스트 듀얼: 최후의 결투 ★★★★★ 10 재미있게 잘 봤습니다. 신고	prot**** 21.11.20	
17807196	007 노 타임 투 다이 ★★★★★ 10 재미있게 잘 봤습니다. 신고	prot**** 21.11.20	

네이버 영화 리뷰 속 조작 평점을 제외한 ‘진짜’ 평점 도출 과정 설계



왜 ‘진짜’ 평점?

평점은 영화를 평가하는 지표이기 때문.

따라서 가짜리뷰와 함께 평점이 과대/과소 평가된 영화가 없도록 하여,  
영화의 평가지표인 평점의 신뢰도를 높이는 것이 플랫폼 자체의 신뢰도를 높일 수  
있다고 생각하였음.

### 3. 분석 기획 및 사용 스택

#### ‘가짜 리뷰’의 조작적 정의

영화나 배우를 광고, 홍보하는 목적으로 보이는 댓글을 선정

- 악의적 리뷰 / 광고성 리뷰
- 악의적 리뷰의 경우, 특정 커뮤니티에서 사용되는 어휘 등
- 특정 영화에만 집중되어 나오는 어휘를 전체 데이터에서 활용하기에 어려움이 있기 때문

### 3. 분석 기획 및 사용 스택

#### 데이터 수집 및 적재

네이버 영화 플랫폼 크롤링



#### 샘플 추출

장르, 기간 별 언어차이 검증

편의 표본 추출



#### 전처리 및 EDA

결측치 제거 및 이상치 확인  
크롤링 데이터 탐색을 통한,  
최종컬럼 도출

불용어 제거 및 언어 교정  
가짜리뷰 선정

NLP



Okt 형태소 분석기를 통한 토큰화

Word2vec 모델을 통한 워드 임베딩  
Sent2vec을 통한 문장 임베딩



#### 클러스터링 분석(준지도)

K\_means 모델 이용한 군집화  
PCA, TSNE 통한 시각화

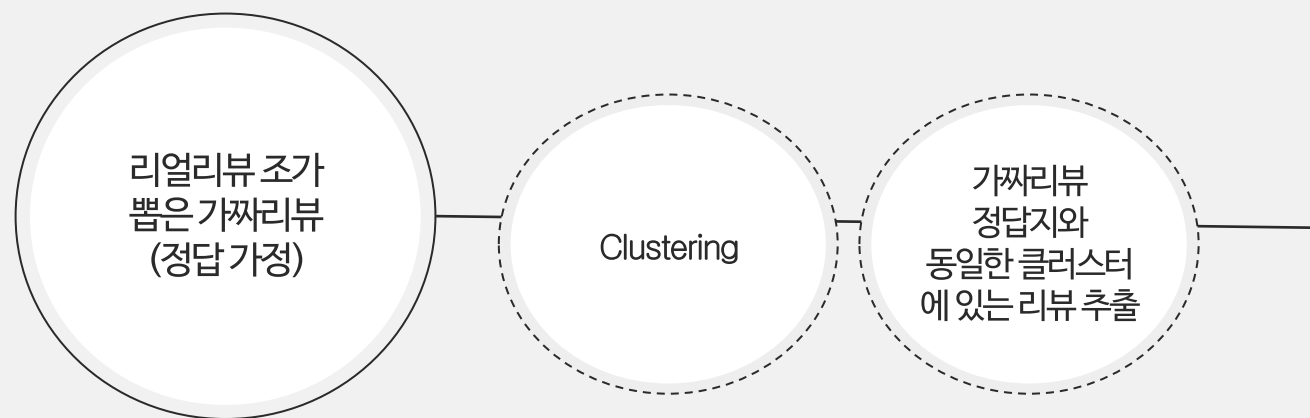


### 3. 분석 기획 및 사용 스택

#### [ 준지도 학습 ]

: 레이블 된 데이터 + 레이블 되지 않은 데이터

소수의 데이터를 예측확률을 기반으로  
재학습시키며 모델의 정확도를 높여 나가는 방식



	영화이름	댓글	평점
178	강릉	정말재밌네요 역시 스튜디오산타클로스 믿고봅니다	10
513	목격자	무더운 한여름밤을 시원하게 해주는 영화 최고	10
446	고질라:킹오브몬스터	하하 이 영화 정말로 기대됩니다	10
522	마약왕	배우들이 너무 연기를 너무 잘함ㅠㅠ 특히 송강호 미쳤더라ㅠㅠ	10
184	강릉	올해 하반기 최고의 기대작	10

- ① 가짜리뷰에 대한 명확한 정답이 존재하지 않는다.
- ② 비슷한 부류의 문장들이 군집 되겠지만, 댓글자체의 가짜/진짜에 대한 진위여부를 분류한 것이라고 보기는 어려울 것이다.
- ③ 따라서, 우리는 **준지도학습의 클러스터 가정** 차용



## 4. 데이터 수집

### ① 네이버 영화 플랫폼 크롤링 및 적재

추출 data

영화코드	영화명	작성일	작성자	평점	관람여부	공감/비공감 수	작성일자	개봉일	
1	러브빅수영리	2003-11-28 00:00:00	1138 2003-12-04 00:00:00	다양하게 펼쳐지는 각양각색 사람들이기-1 교편이가 너무 귀여움-1*	4	0	0	0	6nabi***
2	러브빅수영리	2003-11-28 00:00:00	1145 2003-12-05 00:00:00	물드보이 대신에 선택했지만 정말 감동을 받았습니 이 영화 감추려고 할말,	10	0	0	0	7socc***
3	러브빅수영리	2003-11-28 00:00:00	1207 2003-12-06 00:00:00	12월에 볼 최고의 영화- 연인들이나 가족끼리 꼭보세요	10	0	0	0	8dack***
4	러브빅수영리	2003-11-28 00:00:00	1214 2003-12-06 00:00:00	어느 누구에게도 사랑을 시작하게 하는 영화라고 생각해요	5	0	0	0	8seer***
5	러브빅수영리	2003-11-28 00:00:00	1221 2003-12-06 00:00:00	강력추천합니다!!	10	0	0	0	8sill***
6	러브빅수영리	2003-11-28 00:00:00	1226 2003-12-06 00:00:00	스토리가 지겹고 어수선하지만 여럿까지 사랑을 표현한거에대해서는 불만입니다	1	0	2	2	8joi1***
7	러브빅수영리	2003-11-28 00:00:00	1228 2003-12-06 00:00:00	영화보면서 이번성탄절에 사랑이 눈처럼내리길**	10	0	0	0	8sac***
8	러브빅수영리	2003-11-28 00:00:00	1236 2003-12-06 00:00:00	사랑하는 사람과 또는 가족과 함께 꼭 보세요!!! 돈 아깝지 않아요~!	10	0	0	0	8iron***
9	러브빅수영리	2003-11-28 00:00:00	1242 2003-12-06 00:00:00	올 겨울 이 영화 한편이면 마음이 따뜻해집니다.	10	0	0	0	8kivi***
10	러브빅수영리	2003-11-28 00:00:00	1254 2003-12-06 00:00:00	연시대를 끌어안고 싶어지는영화 -	10	0	0	0	8sois***
11	러브빅수영리	2003-11-28 00:00:00	1258 2003-12-06 00:00:00	이 영화에 출연한 모든 연기자들 좋아하게 된다.*	10	0	0	0	8yma***
12	러브빅수영리	2003-11-28 00:00:00	1267 2003-12-06 00:00:00	크리스마스 전에 잊지말고 꼭보시길...*	10	0	0	0	8maku***
13	러브빅수영리	2003-11-28 00:00:00	1276 2003-12-06 00:00:00	너무 재미있었어요.. 최고치고..*	10	0	0	0	8prom***
14	러브빅수영리	2003-11-28 00:00:00	1295 2003-12-07 00:00:00	진정한 감동	1	0	2	2	9fkdu***
15	러브빅수영리	2003-11-28 00:00:00	1297 2003-12-07 00:00:00	사랑 대박 외롭게 만드는 영화_t_t 그지만 무지 재밌다--;	10	0	0	0	9coot***

- 작성자ID
- 댓글ID
- 댓글
- 평점
- 관람여부
- 스포여부
- 공감/비공감 수
- 작성일자
- 개봉일

\*데이터 출처: 네이버 영화  
<https://movie.naver.com/>

## 4. 데이터 수집

### ② 표본 추출을 위한 요인 검정

#### 어휘사용 현황을 고려한 표본 추출

: 특정 언어의 편향을 막기 위하여 표본 추출 시 고려해야할 요인 선정

**장르** 장르 별로 특정 어휘의 빈도수나 어감이 다를 것이다 : 각 5만개씩 추출하여 검증

〈 6개 장르 〉  
공포, 드라마, 액션, 코미디, 판타지, 애니메이션

**기간** 세대 별로 다른 은어나 문장 패턴들이 있을 것이다 : 각 5만개씩 추출하여 검증

(1) 스마트폰 보급 확대 전 : ~2009

(2) 스마트폰 보급 확대 시기 : 2010~2015

(3) OTT 서비스 출시 및 이용자수 증가 시기 : 2016~

## 4. 데이터 수집

② 표본 추출을 위한 요인 검정 -- 장르 별 워드클라우드 결과

공포



## 드라마



## 액션



코미디



## 애니메이션



# 판타지





## 4. 데이터 수집

### ② 표본 추출을 위한 요인 검정 -- 기간 별 워드클라우드 결과

~2009



2010 ~ 2015



2016~



## 4. 데이터 수집

“기간“ 과 “ 장르“ 별로 어휘에 차이가 없다는 것을 확인



네이버 영화 리뷰는 요인 간 특성이 동질하다고 판단



조사 편의를 위해 “가짜 리뷰가 있을 법한” 영화 리뷰 선정



크롤링을 통해 요인에 상관 없이 **약 250만개의 표본 추출**

## 4. 데이터 수집

origin_index	영화 이름	개봉일	댓글 번호	작성일자	댓글 내용	평 점	관 람 여 부	스 포 여 부	공 감 수	비 공 감 수	공 감 차 이	날 짜 차 이	유저ID	영화코 드	댓글 길이
0	러브 엑추 얼리	2003-11-28 00:00:00	1138.0	2003-12-04 00:00:00	다양하게 펼쳐지는 각양 각색 사랑이야기코맹이가 너무 귀여ㅋ	4	0	0	0	0	0	6	nabi****	36843	31
1	러브 엑추 얼리	2003-11-28 00:00:00	1145.0	2003-12-05 00:00:00	올드보이 대신에 선택했 지만 정말 감동을 받았습 니다 이영화 강추예요 정 말	10	0	0	0	0	0	7	socc****	36843	39
2	러브 엑추 얼리	2003-11-28 00:00:00	1207.0	2003-12-06 00:00:00	월에 볼 최고의 영화 연인 들이나 가족끼리 꼭보세 요	10	0	0	0	0	0	8	dark****	36843	27

언어라는 특성 상 사람도 AI도 ‘댓글’ 자체로는 가짜 리뷰에 대한 판단을 하기 어렵다고 생각하였음.

따라서, 댓글만이 아니라 여러 **정황적 정보**들을 같이 학습시키고자 최종 데이터를 선정

## 5. EDA 및 전처리

### ✓ 댓글 자연어 활용을 위한 전처리 과정

1. 정규표현식을 통한 전처리 과정

2. 형태소 분석기 중 Okt(morphs)를 활용한 토큰화

: 자연어처리에서 일반적으로 자주 사용되는 MeCab과 비교하였으나, 경우 의미를 잃을 수 있는 수준까지 단어를 자르는 것으로 확인되어 Okt 선택

3. 불용어처리

: 토큰화 후 확인하여 의미를 잃은 단어 제외

4. 분석가능한 글이 남아있지 않은 로우 삭제

## 5. EDA 및 전처리

### (1) 개봉일로부터 댓글 작성일까지의 기간

**가정** ) 평점을 조작하는 리뷰는 영화가 상영하는 기간동안 작성될 것

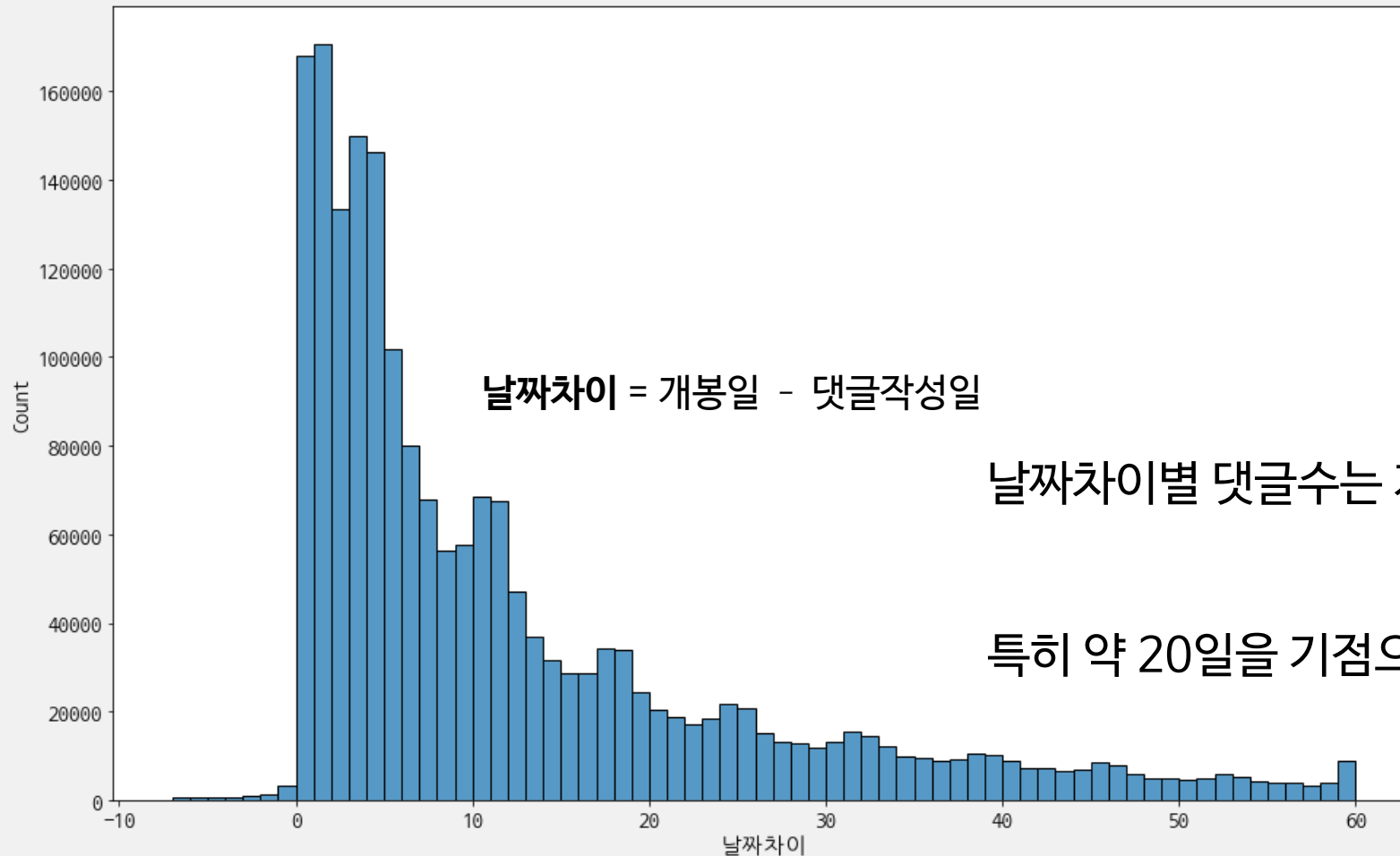
\* 단, 이후에 재개봉을 하거나, OTT와 같은 매체를 통한 상영으로 인한 재평가까지는 생각하지 않음.

일정기간 이후에 달린 리뷰 : 거짓 리뷰와는 거리가 멀다고 생각하여 **제거**  
그 기점을 구하기 위해 개봉일로부터의 일수별 댓글 수와 누적 댓글 수를 확인



## 5. EDA 및 전처리

### (1) 개봉일로부터 댓글 작성일까지의 기간



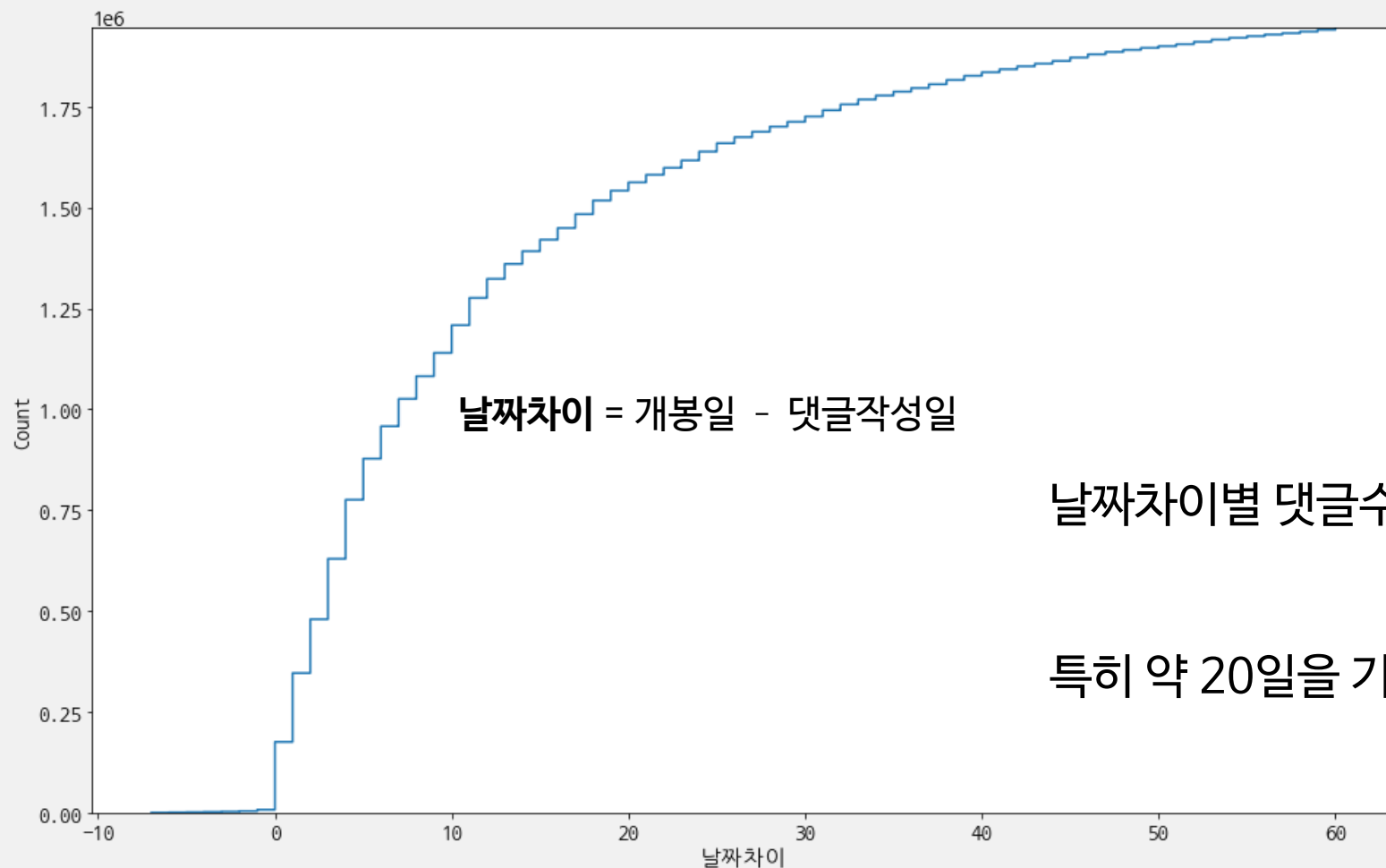
날짜차이 = 개봉일 - 댓글작성일

날짜차이별 댓글수는 개봉일을 시점으로 꾸준히 하락

특히 약 20일을 기점으로 하락하는 폭이 많이 줄어듦

## 5. EDA 및 전처리

### (1) 개봉일로부터 댓글 작성일까지의 기간

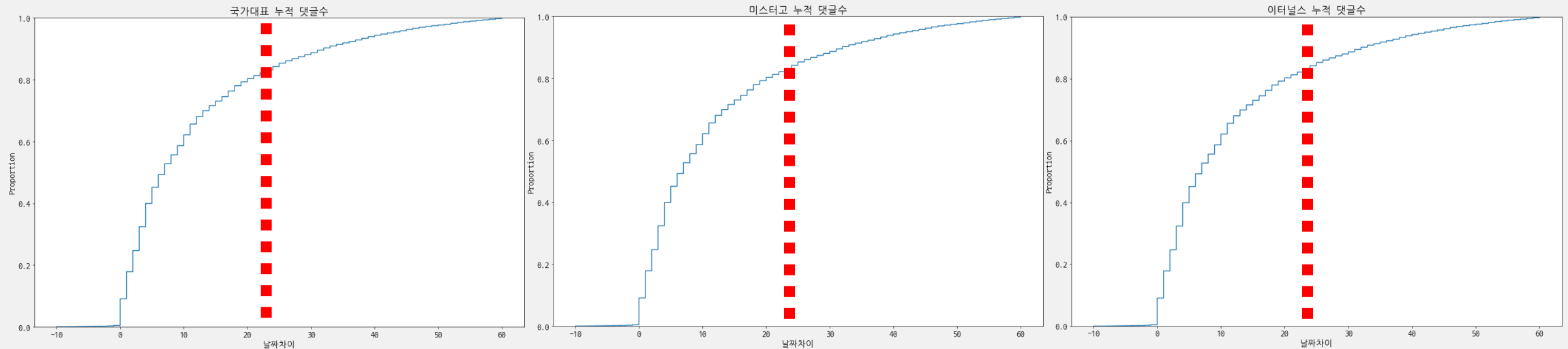


날짜차이별 댓글수는 개봉일을 시점으로 꾸준히 하락

특히 약 20일을 기점으로 상승하는 폭이 많이 줄어들음

## 5. EDA 및 전처리

### (1) 개봉일로부터 댓글 작성일까지의 기간



영화를 랜덤으로 3개를 뽑아 개봉일 이후 누적 댓글수를 확인

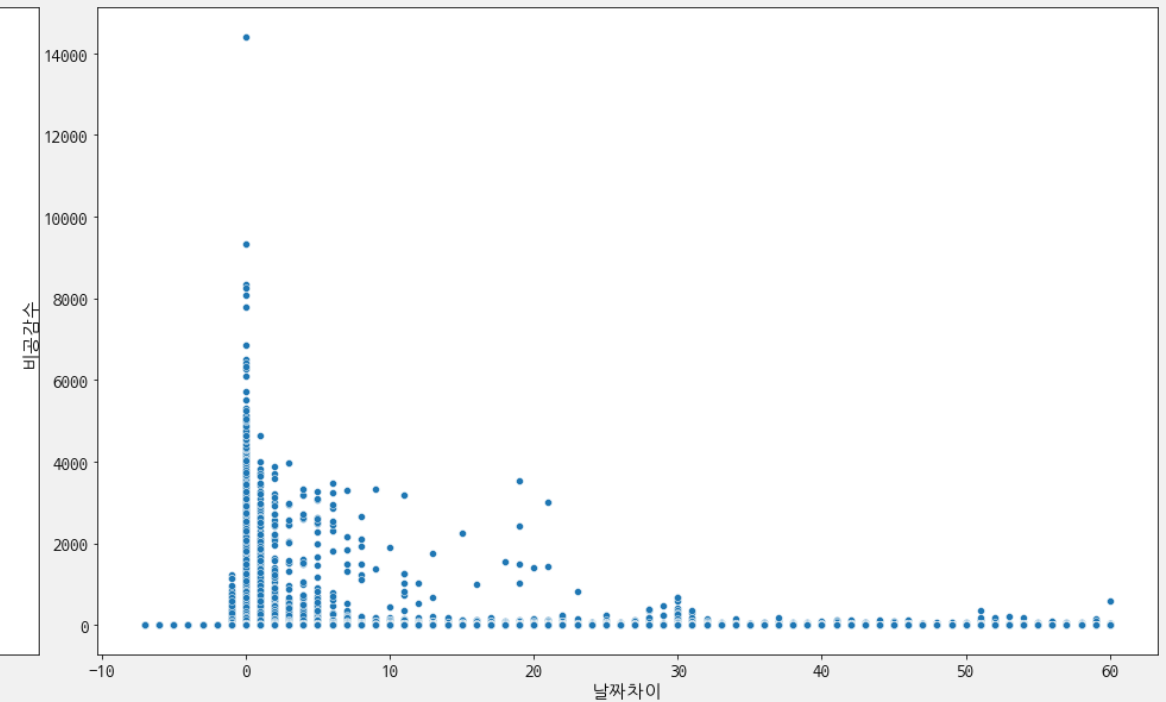
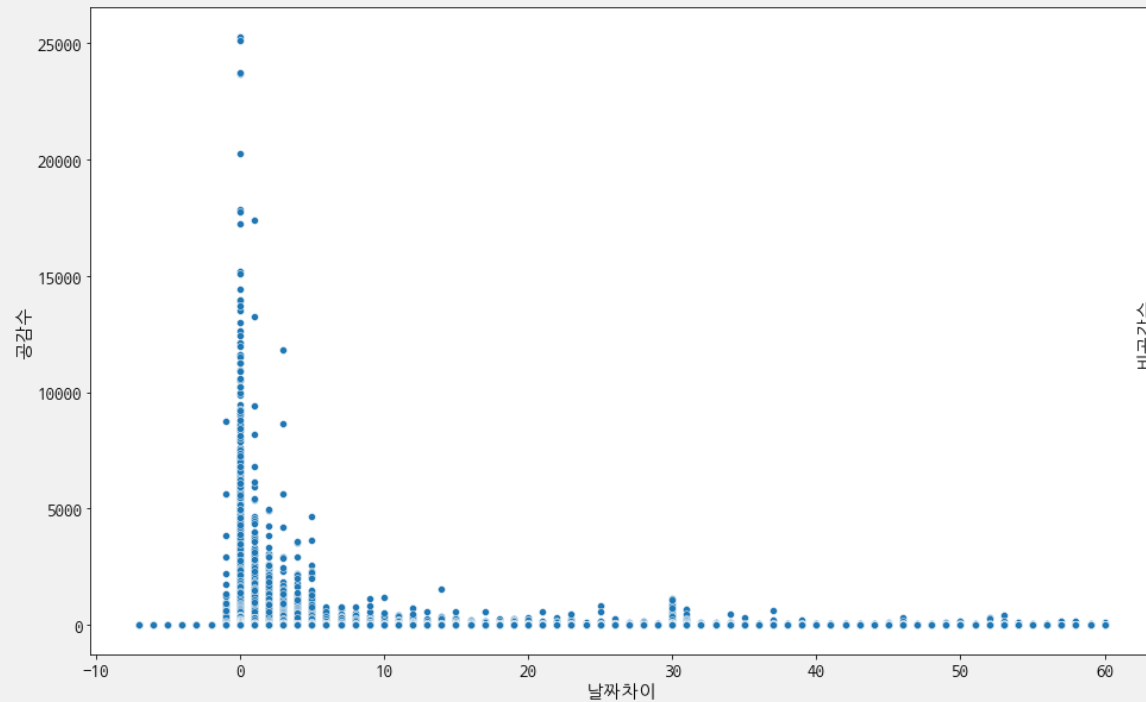
그 결과 전체 영화로 확인했을 때와 마찬가지로 20일 부근에서 기울기가 완만해짐을 알 수 있음

👉 본 분석에서는 개봉 후 3주인 21일 이전까지의 댓글만을 사용하기로 결정

## 5. EDA 및 전처리

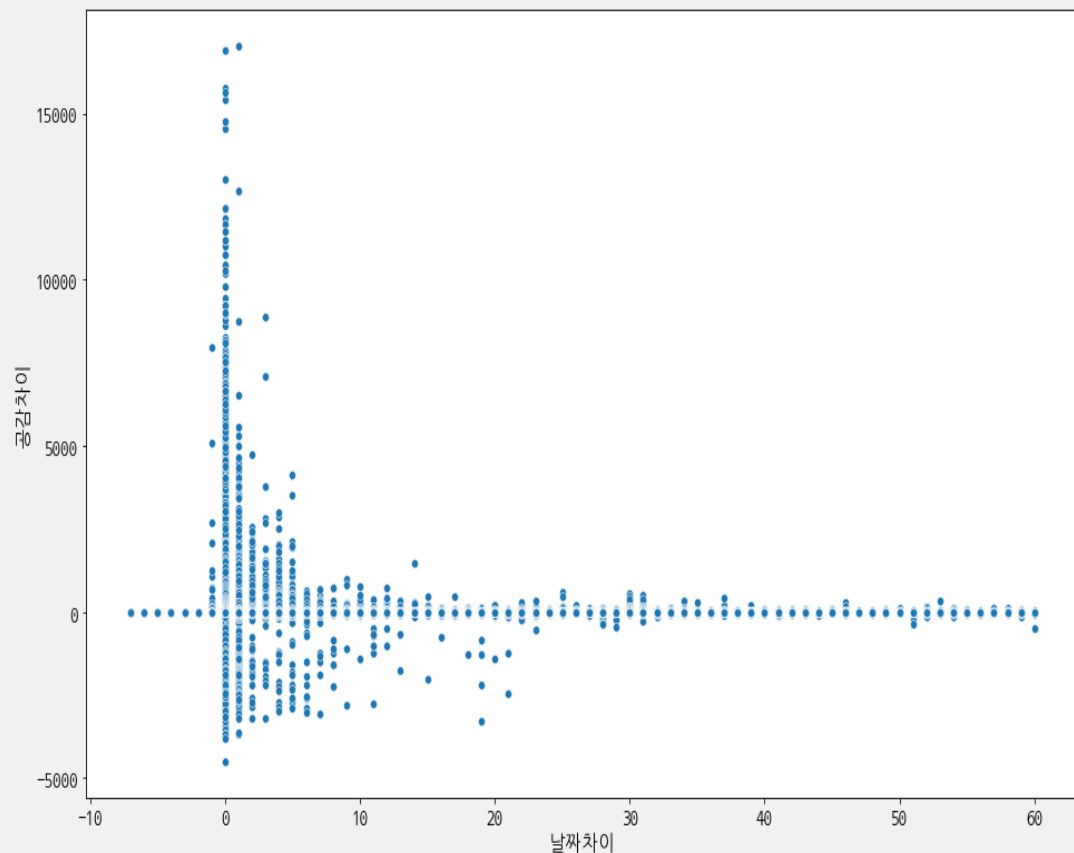
### (2) 공감수 / 비공감수

날짜 차이별 댓글의 공감수와 비공감수



## 5. EDA 및 전처리

### (2) 공감수 / 비공감수

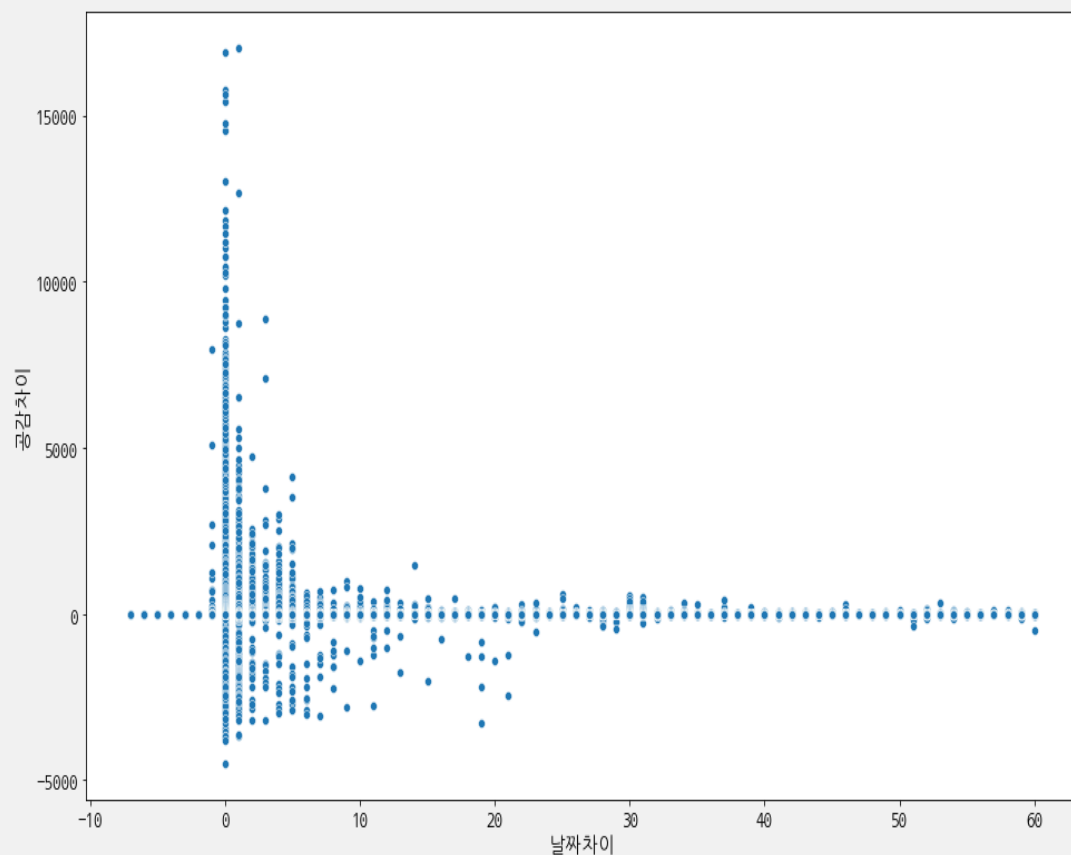


개봉 당일에 공감수와 비공감수가 많은 댓글이 몰려 있음을 알 수 있다.

이것은 개봉 당일 빠른 시간에 작성된 글들이 공감을 받기 시작하면 베스트 댓글로 수많은 사람들에게 노출 되기때문에 공감, 비공감을 많이 얻을 수 있다는 것으로 예측

## 5. EDA 및 전처리

### (2) 공감수 / 비공감수

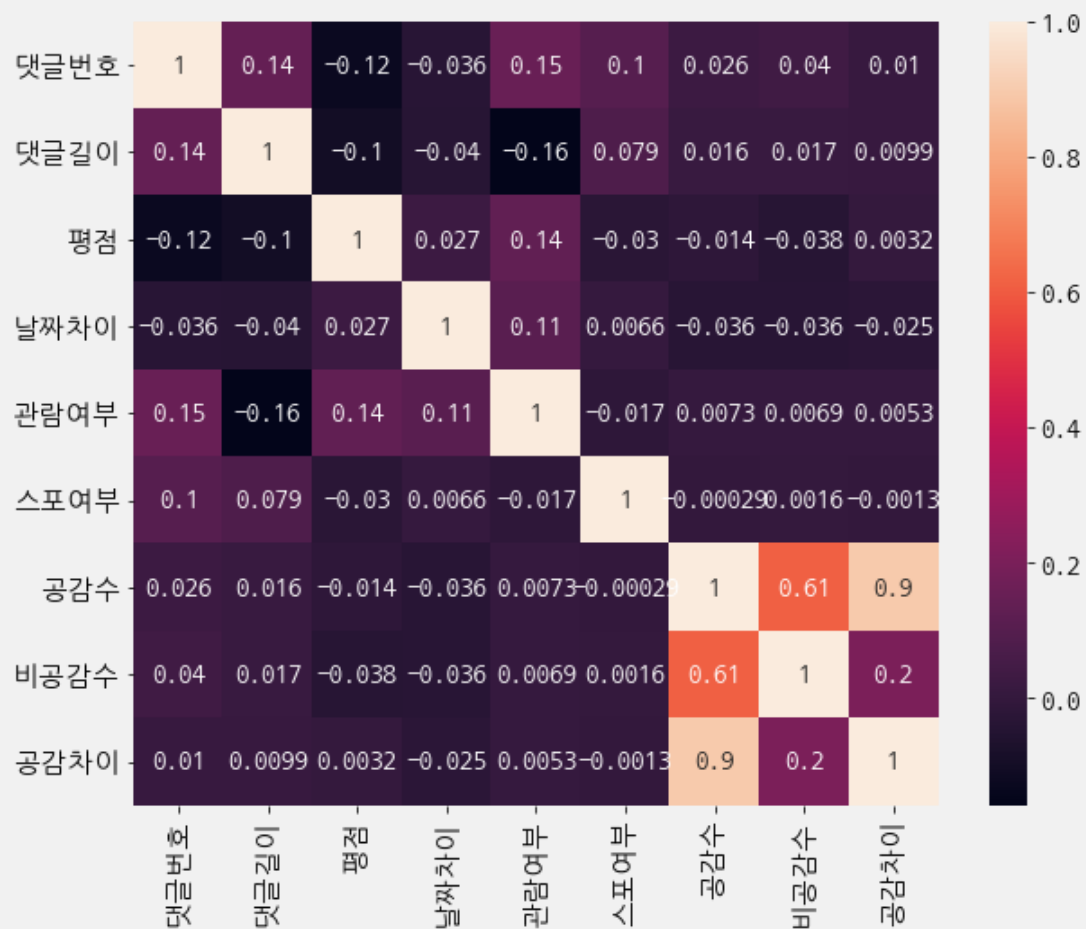


따라서 공감수가 많다고 진실된 댓글이고, 비공감수가 많다고 거짓된 댓글이라고 판단하기에는 무리가 있다고 판단.

따라서 공감수, 비공감수는 본 분석에서 사용하지 않기로 결정.

## 5. EDA 및 전처리

### (3) 모든 컬럼에 대한 상관관계



공감수 부분을 제외한 나머지 컬럼은 모두 상관관계가 0에 가까웠다.

공감수에 관련한 컬럼은 앞에서 제외하기로 결정했으므로 나머지 5개 컬럼은 서로가 관련이 없음을 확인하고 이후 분석을 진행함

## 6. Data Mining

### 1. Word2Vec(Skip-gram)으로 워드 임베딩 진행

Cbow의 경우 비슷한 위치나 의미를 가진 단어 위주로 높은 값을 보여서 조작 댓글로 보이는 댓글들의 문장 자체를 보는 것에 적합하지 않음

Skip-gram 활용 벡터값 조회 결과

```
[25]: 1 print(real_all_vec_sg.wv.most_similar("사랑"))

[('결혼', 0.6609616875648499), ('애뜻', 0.6461261510848999), ('진실한', 0.6445134878158569), ('운명', 0.610365629196167), ('우정', 0.607973575592041), ('부러워', 0.6044792532920837), ('애틀한', 0.5997061729431152), ('녹일', 0.5881887078285217), ('싫게끔', 0.5868324041366577), ('인연', 0.5859013199806213)]

[26]: 1 print(real_all_vec_sg.wv.most_similar("알바"))

[('댓글알바', 0.9397057890892029), ('알바생', 0.9285981059074402), ('댓글', 0.818755030632019), ('조작', 0.7206826210021973), ('남기는거냐', 0.6914980411529541), ('주작', 0.6830329895019531), ('댓글', 0.6819416284561157), ('풀었나보다', 0.6796232461929321), ('특실하네', 0.6678712964057922), ('덧글', 0.6636623740196228)]

[27]: 1 print(real_all_vec_sg.wv.most_similar("평점"))

[('평점', 0.8382205963134766), ('덧글', 0.8159210681915283), ('평점', 0.7665284276008606), ('후기', 0.7594587802886963), ('리뷰', 0.7488888502120972), ('평', 0.7467483878135681), ('네이버', 0.7305018305778503), ('평점', 0.7205923199653625), ('네티즌', 0.7041584253311157), ('평이', 0.7022728323936462)]

[28]: 1 print(real_all_vec_sg.wv.most_similar("아이맥스"))

[('포디', 0.7658892273902893), ('돌비', 0.7100581526756287), ('디지털로', 0.7048077583312988), ('용산', 0.7000710368156433), ('용아', 0.674085795879364), ('리얼디', 0.6637886762619019), ('쓰리', 0.6636122465133667), ('리움', 0.658818244934082), ('엑스', 0.6581565737724304), ('아이맥스', 0.6543728113174438)]

[29]: 1 print(real_all_vec_sg.wv.most_similar("가족"))

[('끼리', 0.7801409363746643), ('연인', 0.7572812438011169), ('부모님', 0.6905289888381958), (' 쌤쌤', 0.6671674251556396), ('함께', 0.6629760265350342), ('가법', 0.6594142913818359), ('애인', 0.6565884351730347), ('친척', 0.6561362147331238), ('보러가기에', 0.6460480690002441), ('친구', 0.6376146674156189)]

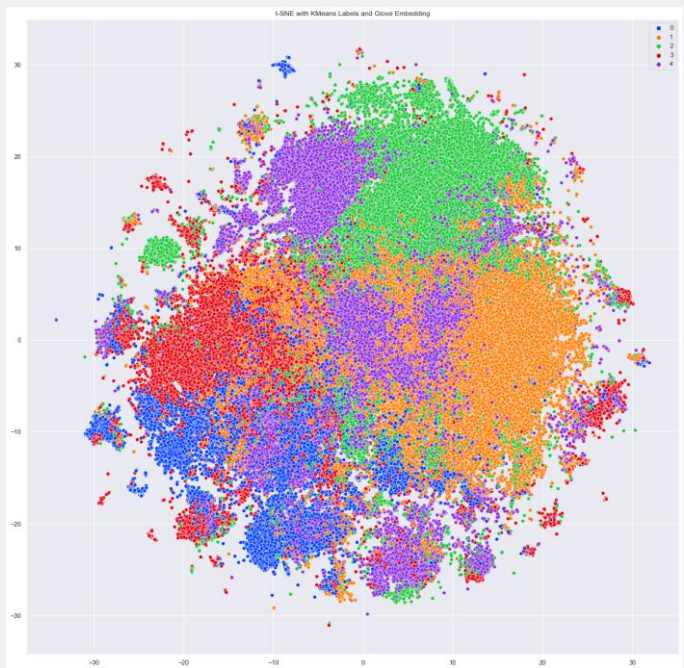
[30]: 1 print(real_all_vec_sg.wv.most_similar("조작"))

[('주작', 0.7572006583213806), ('알바', 0.720682680606842), ('댓글알바', 0.6886276602745056), ('덧글', 0.6620368957519531), ('뺑뺑기', 0.6598663926124573), ('네이버', 0.6596469283103943), ('질종', 0.6549936532974243), ('알바생', 0.6533792018890381), ('매크로', 0.6436693668365479), ('통계', 0.643543243408
```

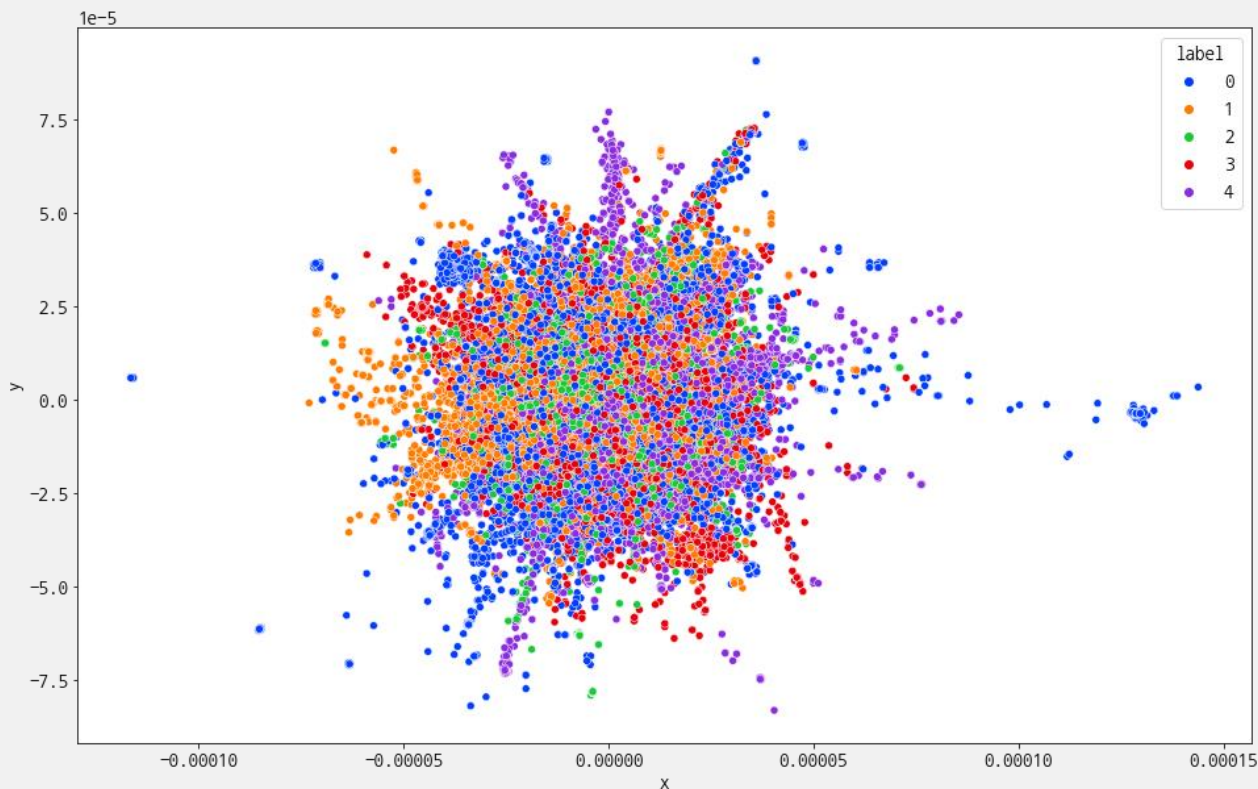


## 6. Data Mining

2. 문장 단위 벡터화: Word2Vec 값을 이용하여 Sent2Vec 진행, K-Means로 클러스터링  
댓글 문장들을 벡터화 하여 시각화 및 확인 결과 가짜평점의 분포가 눈에 띄게 나타나지 않음



20만개 샘플 데이터 클러스터링 시각화 결과

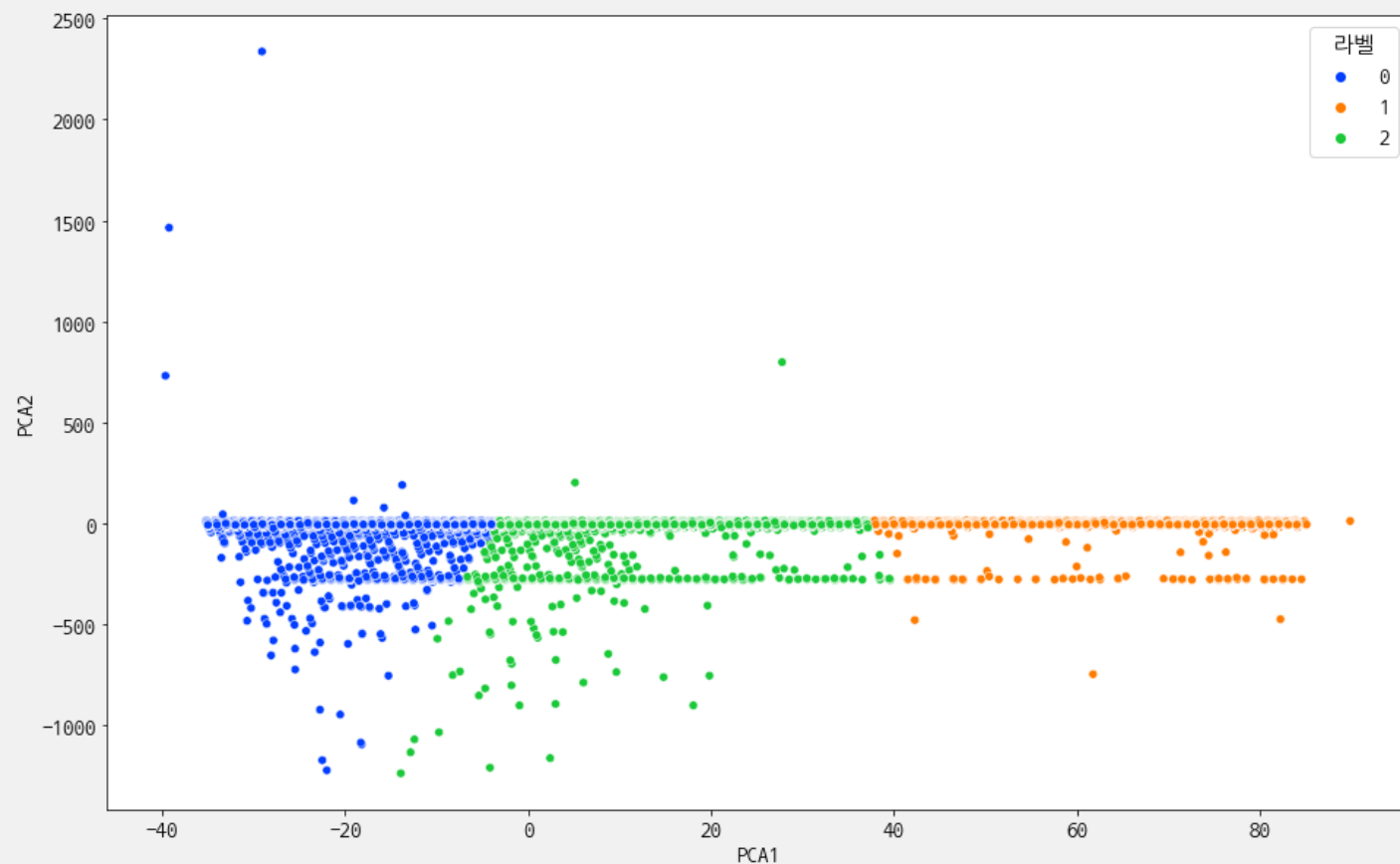
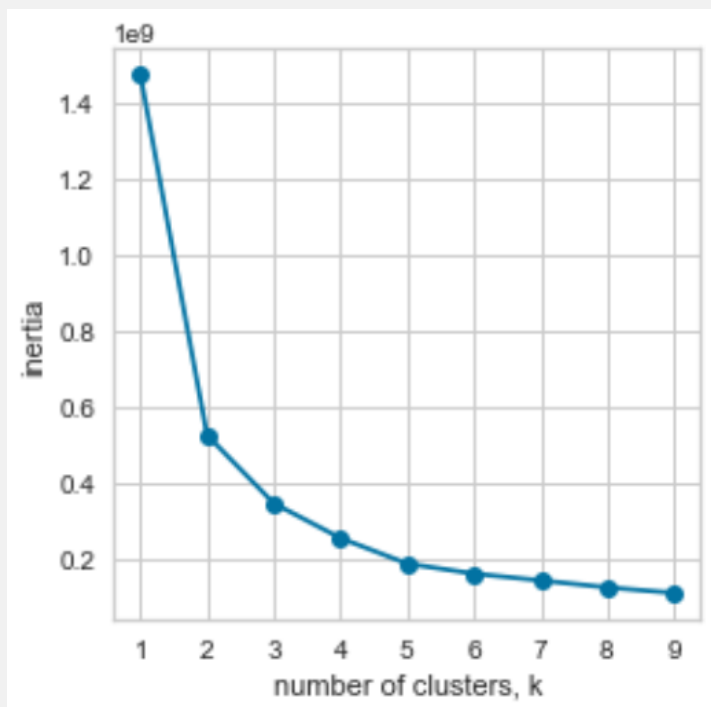


거짓 리뷰를 포함한 60만개 샘플 데이터 클러스터링 시각화 결과

## 6. Data Mining

### 3. 정황정보 컬럼: PCA 진행 후 K-Means로 클러스터링

실제로 가짜리뷰로 정의한 560여개의 댓글 중 75%에 해당하는 댓글이 같은 군집으로 나타남



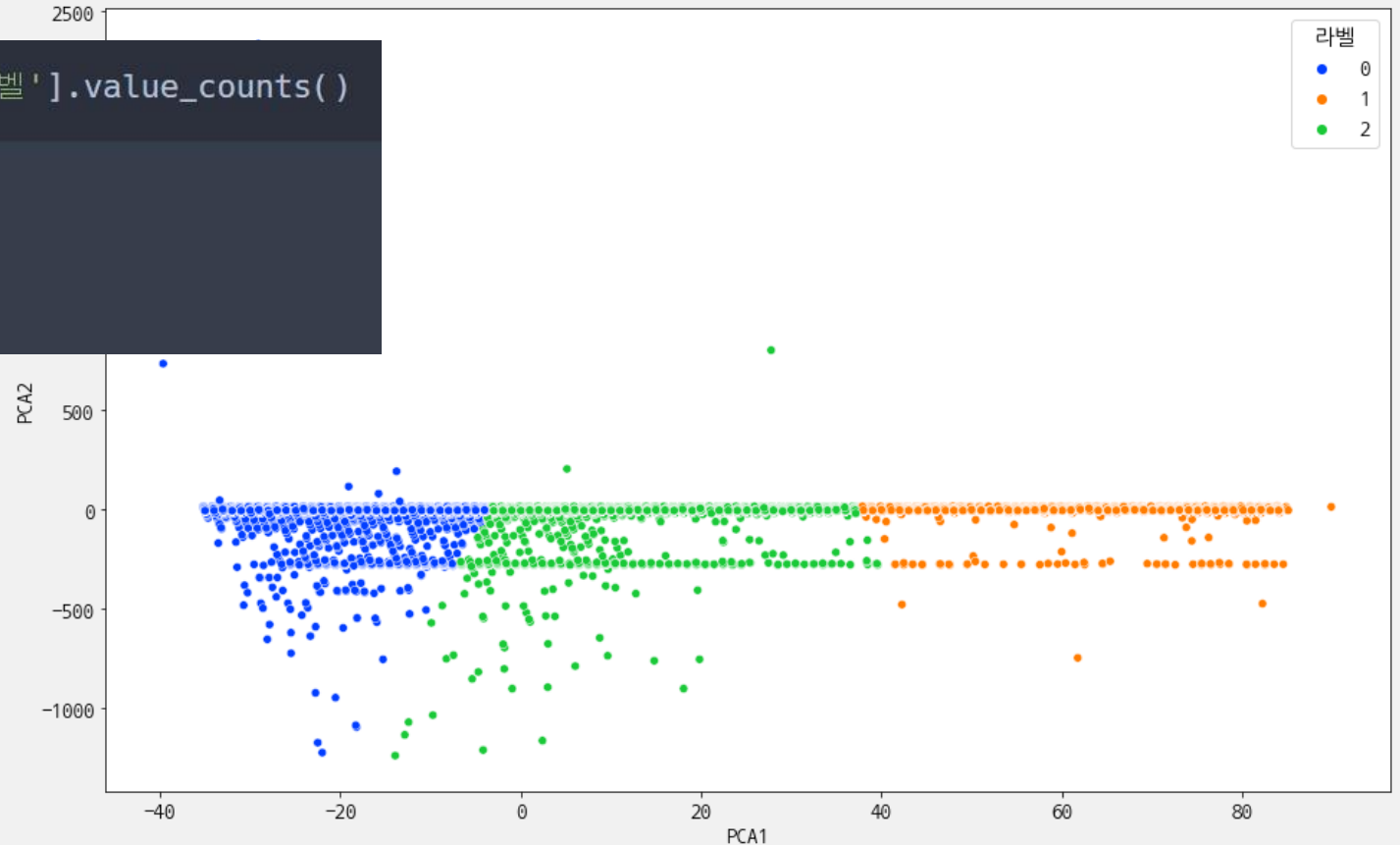
## 6. Data Mining

### 3. 정황정보 컬럼: PCA 진행 후 K-Means로 클러스터링

실제로 가짜리뷰로 정의한 560여개의 댓글 중 75%에 해당하는 댓글이 같은 군집으로 나타남

```
1 df[df['origin_index'] == 3000000]['라벨'].value_counts()
```

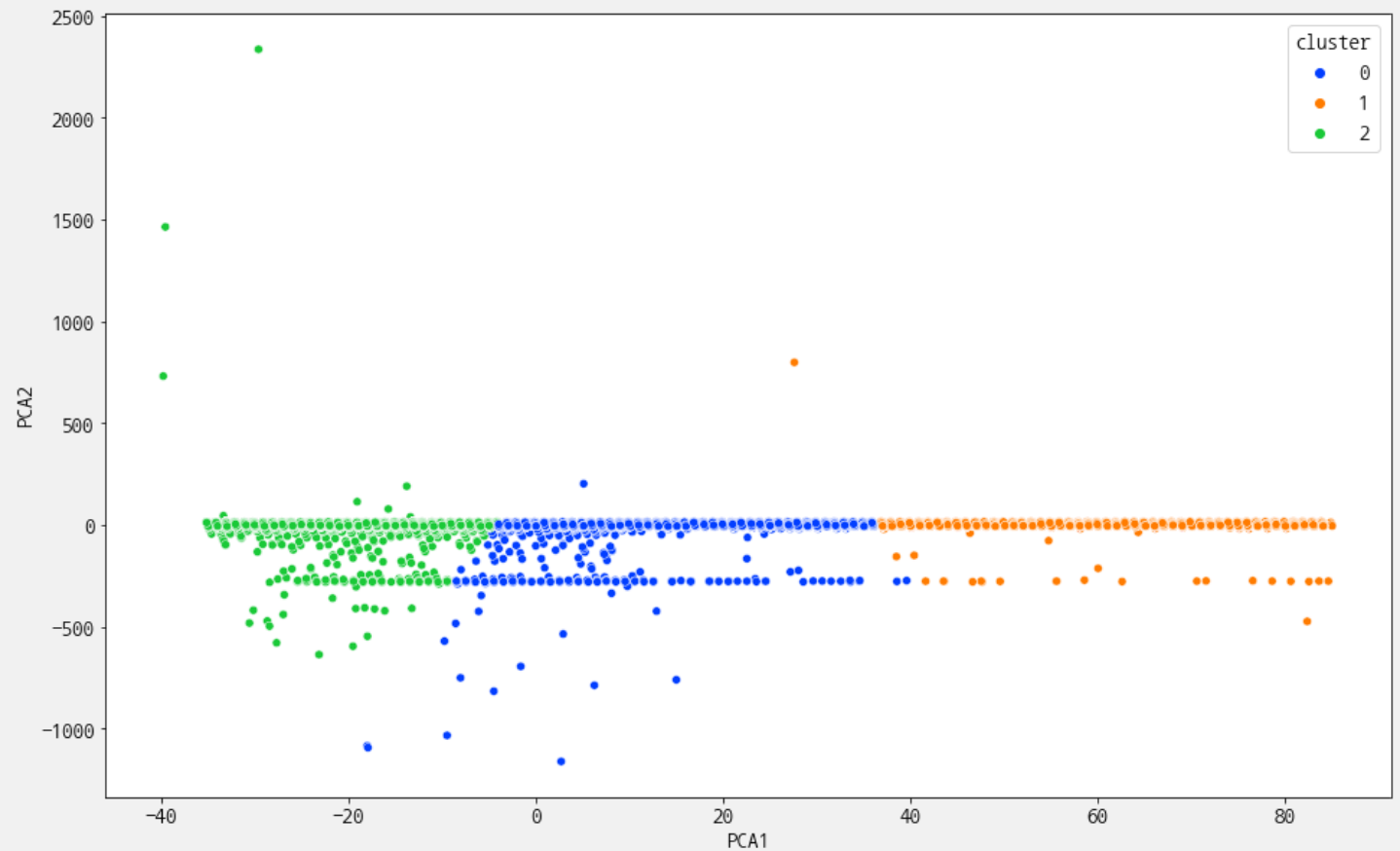
```
0    464
2     89
1       8
Name: 라벨, dtype: int64
```



## 6. Data Mining

### 4. 정황정보 컬럼과 문장 벡터 동시 활용 시도1

문장 벡터를 PCA로 차원축소하여 정황정보 컬럼 한 개와 동등한 상태로 클러스터링 진행



## 6. Data Mining

### 4. 정황정보 컬럼과 문장 벡터 동시 활용 시도1

문장 벡터를 PCA로 차원축소하여 정황정보 컬럼 한 개와 동등한 상태로 클러스터링 진행

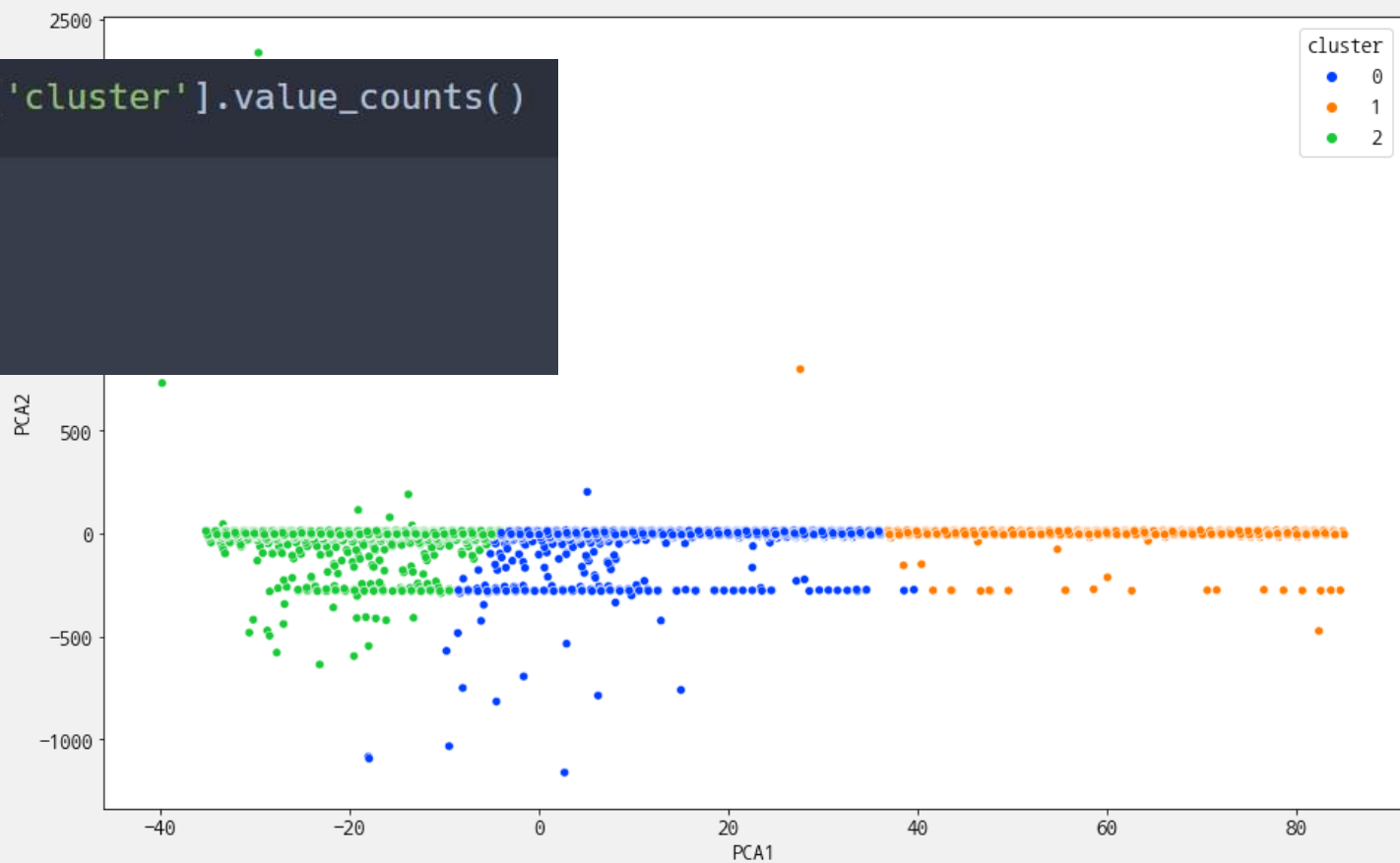
```
1 df[df['origin_index'] == 3000000]['cluster'].value_counts()
```

```
2      461
```

```
0       90
```

```
1       10
```

```
Name: cluster, dtype: int64
```



## 6. Data Mining

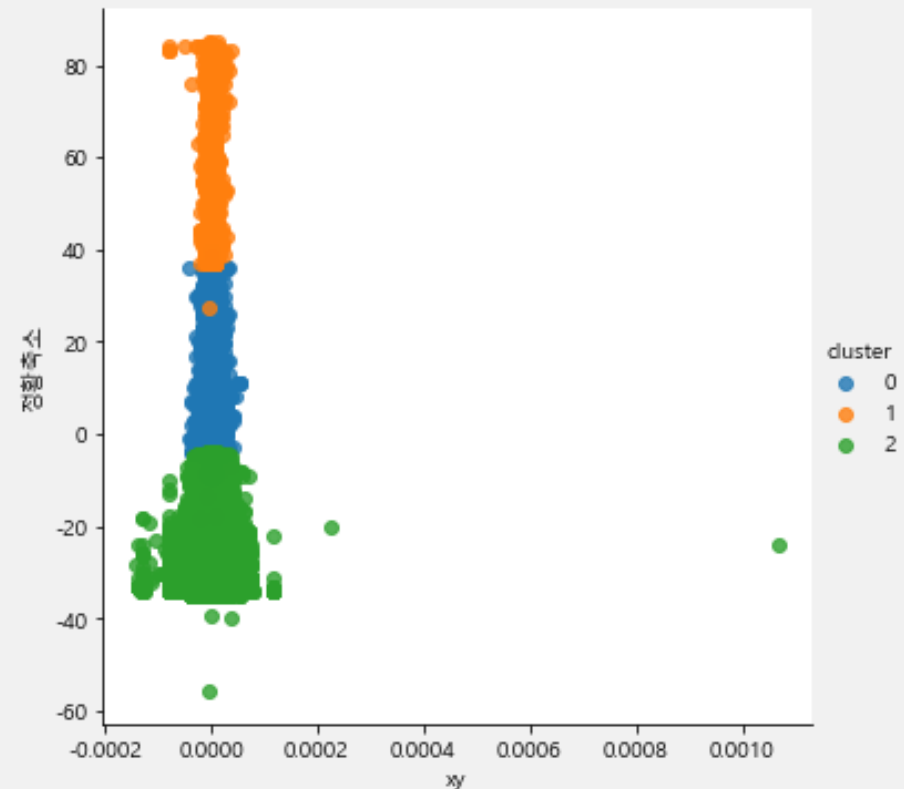
### 5. 정황정보 컬럼과 문장 벡터 동시 활용 시도2

문장 벡터와 정황정보 컬럼 모두를 각각 PCA로 차원축소하여 두 가지 속성으로 클러스터링 진행

(정황정보 컬럼 개수로 인한 영향을 줄이기 위함)

실제로 정황정보만 클러스터링 했을 시와 비슷한 형태

결국 정황정보를 빼고 가짜 판별은 어렵다고 판단



## 7. 결과 및 한계점

### 결과

1. 문장은 정황정보와 함께 두어야 유의미해 질 수 있다.

: 워드 임베딩을 통한 문장 단위 임베딩만으로는 가짜에 가까운 데이터를 가짜로 확정 짓기 어려운 형태

164973	걸캅스	재밌어요시사회로봤었는데 또보려구요	10	0	0	18	2
350832	오블리비언	오랜만에 볼만한 대작이 나왔다 년 들어 본 영화중에 최고	10	0	0	31	2
94332	건축학개론	토요일 보고왔는데 여운이 사라지질않는다 그래서또봐야 겠다	10	0	0	31	2

문장과 정황정보를 함께 두고 클러스터링 후 분류된 댓글들이 뽑았던 거짓 리뷰와 유사하게 나온 상태

3. 관람객이라고 믿을 수 있는 댓글은 아니었다.

## 7. 결과 및 한계점

### 결과

1. 문장은 정황정보와 함께 두어야 유의미해 질 수 있다.

: 워드 임베딩을 통한 문장 단위 임베딩만으로는 가짜에 가까운 데이터를 가짜로 확정 짓기 어려운 형태

2. 댓글길이 짧을 수록 거짓 댓글과 비슷한 정황일 확률이 높았다.

	댓글길이 평균	댓글길이 최대	댓글길이 최소
가짜 댓글이 많은 클러스터	18.21	32	1
그렇지 않은 군집1	99.81	126	74
그렇지 않은 군집2	46.74	73	11



## 7. 결과 및 한계점

### 결과

1. 문장은 정황정보와 함께 두어야 유의미해 질 수 있다.

:워드 임베딩을 통한 문장 단위 임베딩만으로는 가짜에 가까운 데이터를 가짜로 확정 짓기 어려운 형태

2. 댓글길이가 짧을 수록 거짓 댓글과 비슷한 정황일 확률이 높았다.

3. 관람객이라고 믿을 수 있는 댓글은 아니었다.

	관람여부
가짜 댓글이 많은 클러스터	242565
그렇지 않은 군집1	19187
그렇지 않은 군집2	68954

## 7. 결과 및 한계점

### 한계점

#### 1. 한국어 형태소 분석기의 한계

: 사소한 뉘앙스 파악할 수 있을 만큼 고도화된 형태소 분석이 가능한 모델이 부재

#### 2. 유저 정보 수집의 한계

: 아이디의 생성일시나, 특정 유저의 다계정 여부 등을 알아볼 수 있는 데이터가 있었다면 조작을 위해 만들어진 아이디임을 판별하기 쉬웠을 것

#### 3. 반지도학습 연구 정보 부족의 한계

: 자연어 반지도학습을 통한 모델링 관련 연구 정보가 부족하여 분석 진행에 어려움이 존재하였음

## 8. 분석 활용

### 프로젝트 활용 시 기대효과

리뷰의 댓글과 정황 함께 분석 : 네이버 영화 리뷰 속 조작 평점을 어뷰징

- ▶ 조작평점 제외 후 실제 평점을 도출하여 네이버 영화 평점의 신뢰도 상승
- ▶ 네이버 영화 이용자수 증가
- ▶ 시리즈온 유입 가능성 증가 및 매출 상승 기대



추후 준지도학습을 통한 라벨링을 이용하여  
KNN을 활용한 조작 리뷰 분류 모델 설계 기획 중

Any Question?

감사합니다:)