
피파 온라인 댓글 데이터 감성 분석 및 관련 선수 추천 모델 개발

AI 11기 김현승

목차

01 배경

02 데이터 전처리

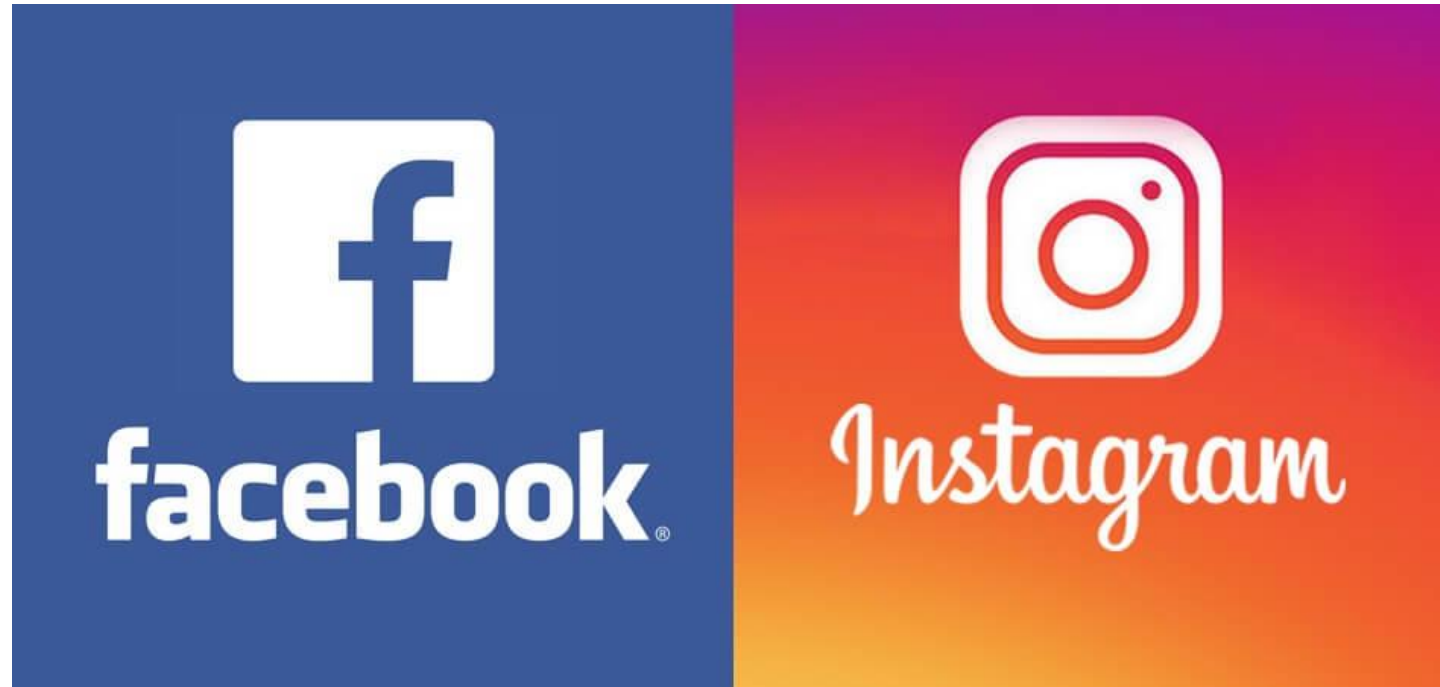
03 모델링 및 관련 선수 추천

04 결론 및 한계

01

배경

배경



coupang

MUSINSA

댓글 감성 분석



소비자 선호도 파악

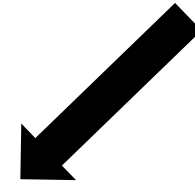


상품 or 기능 보완

배경



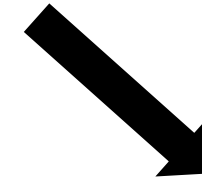
댓글 감성 분석(긍정 or 부정)



선수 선호도 파악



선수 보완



평점 + 감성 점수 도출



관련 선수 추천

가설

1. 긍정적인 댓글, 부정적인 댓글을 이루는 단어들은 차이가 있을 것이다.
 2. 선수의 포지션에 따라 자주 사용되는 단어들은 차이가 있을 것이다.
-

02

데이터 전처리

데이터 전처리

피파 인벤 – Crawling

선수		평가	작성자
	LOL Z. 이브라히모비치 너무 좋다 중거리, 감차, 헤더되고 침투도 구린것도아님 ㅋㅋ 센백에 놓여진 윙백들이 달라붙어도 밀쳐버리는 모오오오습	  107 ★ 3.92	 팡목천왕 19:04
	17 톰 슈타르케 9카 매물 있을려나	  66 ★ 5.00	 누안칸지만 19:01

userId	player_name	position	comment	rating
클름광팬	박지수	df	선수체감 자체가 엄청 가볍고 민첩함	4.00
클름광팬	박지수	df	주력이랑 방향전환 모두 빨라가지고 수비커버 범위도 넓게 가져갈수 있음.	4.00
클름광팬	박지수	df	5카 기준 2천만bp 정도로 대장급 공격수들을 스피드로 제압가능.. 우리집 센터백은...	4.00

중복, 결측치 제거

20만 행 x 5개 열

데이터 전처리

KNU 한국어 감성사전

ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ 특수문자

단어: 기쁘다

어근: 기쁘

감성: 2 (매우 긍정)

주관적인 점수 X

객관적인 점수 0

이 선수 체감 너무 좋은데 골결이 아쉽다.



[이, 선수, 체감, 너무, 좋은데, 골결, 아쉽다]



[0, 0, 0, 1, 2, 0, -2]



감성 점수 = 1

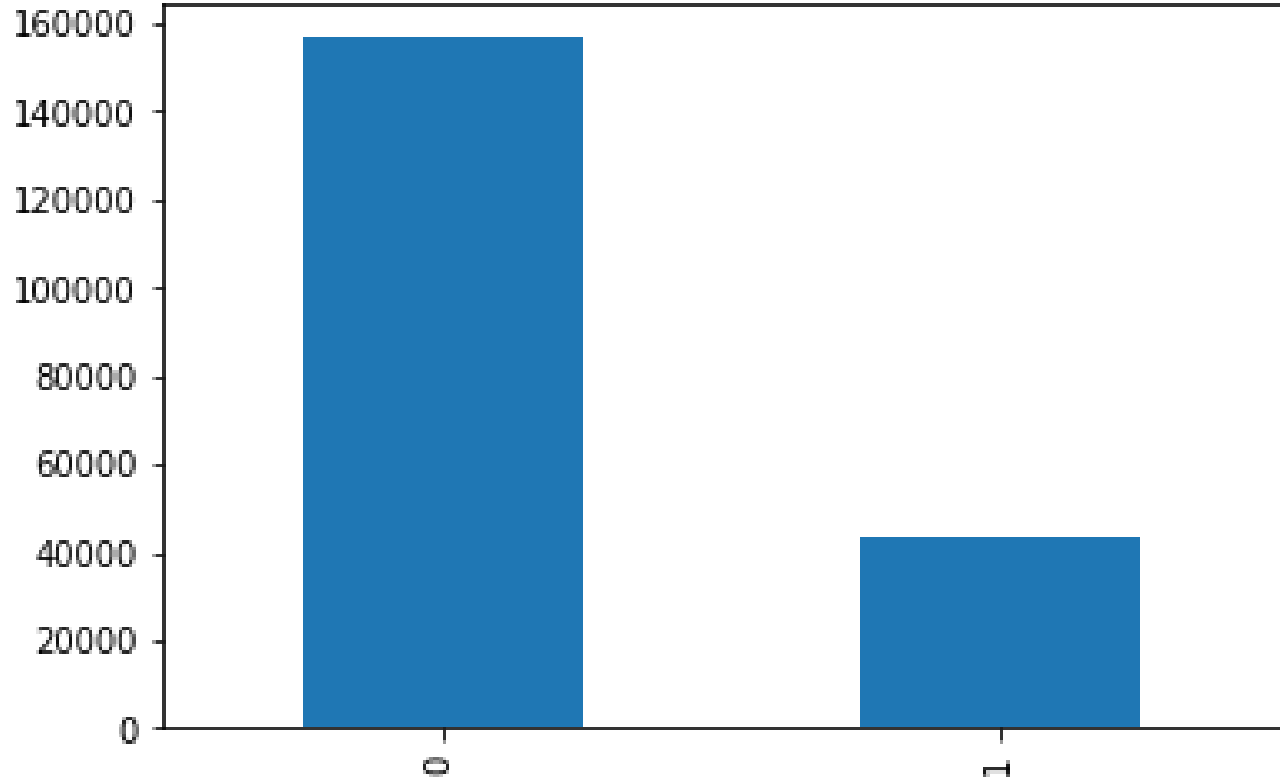
데이터 전처리

감정 점수 > 0 → 1, 감정 점수 ≤ 0 → 0

userId	player_name	position	comment	rating
클름광팬	박지수	df	선수체감 자체가 엄청 가볍고 민첩함	4.00
클름광팬	박지수	df	주력이랑 방향전환 모두 빨라가지고 수비커버 범위도 넓게 가져갈수 있음.	4.00
클름광팬	박지수	df	5카 기준 2천만bp 정도로 대장급 공격수들을 스피드로 제압가능.. 우리집 센터백은...	4.00



userId	player_name	position	comment	rating	score	label
클름광팬	박지수	df	선수체감 자체가 엄청 가볍고 민첩함	4.00	1	1
클름광팬	박지수	df	주력이랑 방향전환 모두 빨라가지고 수비커버 범위도 넓게 가져갈수 있음.	4.00	0	0
클름광팬	박지수	df	5카 기준 2천만bp 정도로 대장급 공격수들을 스피드로 제압가능.. 우리집 센터백은...	4.00	-1	0



0 : 78%

1 : 22%

데이터 전처리

1) 특수문자 및 불용어 제거

stopwords = ['을', '를', '.....', '구매합니다', '올렸습니다', '팔렸나요', '.....']

2) 정수 인코딩

맨시티 구단에 고급여 풀백으로 쓸라고 은카만 샀는데 걔 바로 LAM RAM번갈아 가...

부캐뮌헨에3카사용중종긴함 1티어는 너무보유자욕심인것같고슬직히 15티어가적당 블랑보다...

```
[[243, 717, 956, 937, 375, 205, 11, 6344, 2, 24, 327, 7, 194, 287, 1761, 1822, 2430, 28, 99, 530, 103, 106, 165, 521, 333, 143, 53, 750, 45, 3324, 9001, 10, 117, 29, 263, 45, 8, 6, 242, 98, 26, 1312, 529, 223, 372, 4, 2, 52, 25, 3, 93, 32, 33, 167, 59, 2098, 3508, 87, 27, 242, 662, 167, 1160, 1126, 40, 6956, 1022, 77]]
```

3) 패딩

```
array([[ 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0, 243, 717, 956, 937, 375]), dtype=int32)
```

03

모델링 및 관련 선수 추천

모델링

문장 = 여러 단어들의 순서로 이루어진 시퀀스 데이터(RNN)

1) Basic



2) Randomized Search CV



관련 선수 추천



1) Min Max Scaling

userId	player_name	position	comment	rating	score
클롭광팬	박지수	df	선수체감 자체가 엄청 가볍고 민첩함	4.00	1
클롭광팬	박지수	df	주력이랑 방향전환 모두 빨라가지고 수비커버 범위도 넓게 가져갈수 있음.	4.00	0
클롭광팬	박지수	df	5카 기준 2천만bp 정도로 대장급 공격수들을 스피드로 제압가능.. 우리집 센터백은...	4.00	-1

userId	player_name	position	comment	rating	score	new_score
클롭광팬	박지수	df	선수체감 자체가 엄청 가볍고 민첩함	4.00	1	0.218182
클롭광팬	박지수	df	주력이랑 방향전환 모두 빨라가지고 수비커버 범위도 넓게 가져갈수 있음.	4.00	0	0.200000
클롭광팬	박지수	df	5카 기준 2천만bp 정도로 대장급 공격수들을 스피드로 제압가능.. 우리집 센터백은...	4.00	-1	0.181818

관련 선수 추천

2) 감성 점수 반영

$$\text{new_rating} = \begin{pmatrix} \text{긍정 단어면 } 1 + \text{new_score} \\ \text{부정 단어면 } 1 - \text{new_score} \end{pmatrix} \times \text{rating}$$

userId	player_name	rating	score
클롭광팬	박지수	4.00	1
클롭광팬	박지수	4.00	0
클롭광팬	박지수	4.00	-1

userId	player_name	new_rating
클롭광팬	박지수	4.872727
클롭광팬	박지수	3.200000
클롭광팬	박지수	3.272727

관련 선수 추천

3) KNN (cosine similarity)

	userId	0228	09캐릭7카 팜	1002lmw	100이 니	10u드 로
player_name						
A. 그랑돌리에		0.0	0.0	0.0	0.0	0.0
A. 그랑크비스 트		0.0	0.0	0.0	0.0	0.0

아이템 기반 협업 필터링

박지성 선수와 가장 연관있는 선수리스트:

- 1. 기성용
- 2. 홍명보
- 3. 손흥민
- 4. 유상철

리오넬 메시 선수와 가장 연관있는 선수리스트:

- 1. 네이마르 Jr.
- 2. 킬리안 음바페
- 3. 루이스 수아레스
- 4. 피케

본인 포함 유사도 Top 5

04

결론 및 한계

가설 검증

1. 긍정적인 댓글, 부정적인 댓글을 이루는 단어들은 차이가 있을 것이다.

긍정



('패스', 13086), ('체감', 12470), ('진짜', 12313), ('보다', 11553), ('침투', 10948)

부정



('진짜', 16629), ('그냥', 12512), ('급여', 12323), ('인대', 11323), ('체감', 10933)

가설 검증

2. 포지션에 따라 자주 사용되는 단어들은 차이가 있을 것이다.

FW



('침투', 14884), ('진짜', 10632), ('체감', 9232)

MF



('패스', 15332), ('체감', 10395), ('진짜', 10115),

DF



('급여', 11361), ('수비', 9590), ('진짜', 8195)

한계 및 해결방안

한계

01 같은 선수라도 시즌별로 스탯이 다른데 선수를 추천할 때 시즌을 고려하지 못했다.

02 감성 점수를 기반으로 새로운 rating 점수를 만들 때 통계적인 지식이 부족해서 타당성을 설명하기 힘들었다.

03 한국어 감성사전에 없는 신조어들은 감성 점수에 반영할 수 없었다.

해결방안

.....● 시간이 더 주어진다면, 시즌별로 선수를 세분화하여 더 정확한 추천이 가능할 것 같다.

.....● 감성 분석과 추천 시스템이 결합된 관련 연구를 찾아본다면 더 설득력을 갖출 것 같다.

.....● 빈도수가 높은 신조어들만이라도 사전에 직접 추가한다면 신조어의 특징을 더 살릴 수 있을 것 같다.

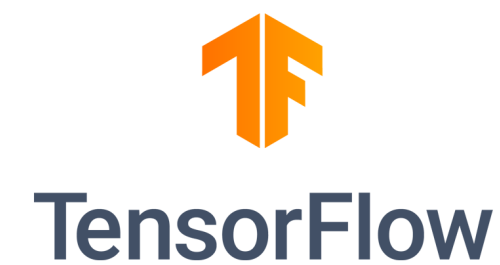
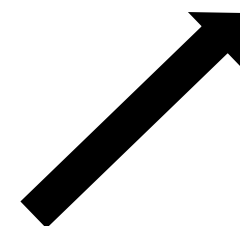
Thank you



BeautifulSoup



KNU 한국어 감성사전





colab

