

올리브영 후기의 색조화장 만족 및 불만족
요인 Topic과 피부타입의 연관성 연구
: 토픽모델과 머신러닝을 활용하여

EAT Sight

김예준

오유성

오현진

조영서 (발표자)

참고논문

참고한 점

Journal of the Korean Data &
Information Science Society
2021, 32(5), 1035–1047

http://dx.doi.org/10.7465/jkdi.2021.32.5.1035
한국데이터정보과학회지

IT 기업 직원의 만족 및 불만족 요인에 따른 이직률 예측:
토픽모델링과 머신러닝을 활용하여[†]

최진옥¹ · 신동원² · 이한준³

¹고려대학교 스마트미디어 서비스 연구센터 · ²명지대학교 산학협력단 · ³명지대학교 경영정보학과
접수 2021년 8월 27일, 수정 2021년 9월 7일, 게재확정 2021년 9월 7일

요 약

IT 산업에서의 이직은 기술 축적과 개발 및 관리에 있어 기업들에게 큰 어려움이 되고 있어 이에 대한 연구가 긴요한 실정이다. 하지만 아직까지 IT 분야의 이직 및 인력 관리와 관련된 연구는 많지 않다. 이에 본 연구에서는 이직률에 영향을 미치는 직원들의 만족 요인과 불만족 요인들을 살펴보고 이를 활용한 IT 기업의 이직률 예측 모델을 제안하고자 한다. 이를 위해 온라인 기업 리뷰 사이트인 잡플래닛에서 국내 유가증권시장에 상장된 129개 IT 기업에 종사하는 직원들의 리뷰 21,589건을 수집하고 토픽모델링을 실시하였다. 추출된 토픽들을 활용하여 이직률에 대한 머신러닝 기반의 예측 모델을 제안하였다. 또한 변수 중요도 평가를 통하여 이직률에 영향을 미치는 직원들의 만족 및 불만족 요인별 영향 정도를 분석하여 제시하였다.

주요용어: 기업 리뷰, 머신러닝, 이직률, 직원 만족, 토픽모델링.

텍스트 데이터 활용

- 리뷰 데이터 활용

: 사용자의 주관적 의견 리뷰로 만족/불만족 요인Topic 분석

토픽 모델링 적용

- 직원 리뷰에서 이직률 영향 미치는 만족/불만족 요인

: 올영 고객 리뷰에서 색조화장품에 대한 만족/불만족 요인

머신러닝 활용 예측 모델

: 참고논문 : 만족/불 요인 바탕으로 이직률 예측

: 올영 : 만족/불만족 요인 바탕으로 고객 피부 타입 예측

데이터 수집 가공 절차

리뷰 데이터 수집 -> NLP 전처리 -> topic modeling적용 -> ML

NLP전처리

: 토픽 모델링,머신러닝 전에,

TEXT를 정리,분석 쉽게 변환하는 가공 과정 EX)토큰화

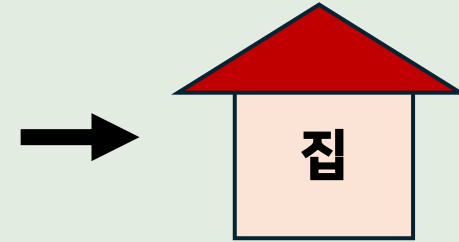


중복 주제 없음

배경(1/3)



→ 발색 , 구경



리뷰 찾기

구매

리뷰 작성

배경(2/3)

PINK OLIVE 조*서 로그아웃 | 마이페이지 | 장바구니(1) | 주문배송 | 고객센터 | 올영매장 | Global

OLIVE YOUNG

상품, 브랜드, 성분 검색

오늘드림 | 올영매장찾기 | 최근 본 상품

카테고리

스킨케어
메이크업/헤어
뷰티소품
더모 코스메틱
맨즈케어
향수/디퓨저
헤어케어
바디케어
건강식품
푸드
구강/건강용품
여성/위생용품
라이프/케아프
AWARDS

상품설명 | 구매정보 | 리뷰 (12,181) | Q&A (182)

전체 상품 옵션

총 12,181 건

4.7 점

★★★★★

81% 13% 5% 1% 1%

5점 4점 3점 2점 1점

발색력 | 지속력 | 발림성 | 수분감

아주 만족해요 71% | 지속이 오래돼요 45% | 아주 만족해요 74% | 아주 촉촉해요 70%

보통이에요 26% | 보통이에요 47% | 보통이에요 26% | 보통이에요 30%

다소 아쉬워요 3% | 예상보다 짧아요 8% | 다소 아쉬워요 1% | 매트해요 1%

유용한순 | 도움순 | 최신순 | 포토리뷰 | 일반리뷰 | 체크다 | 리뷰 검색 필터

추천 키워드 전체 | #사용감 | #사용부위 | #사용효과 | #상품만족도 | #색상

젠니 TOP187

복합성 · 물론 · 모공 · 민감성

★★★★★ 2025.03.07 · 매장

[옵션] 02 로즈

#text 680.05 × 38.59 | 지속이 오래돼요 | 발림성 아주 만족해요 | 수분감 아주 촉촉해요

... 글로우 톤이 유명해서 샀었는데 만족합니다. 케이스도 미쁘고 컬러도 이쁘고, 그리고 광택도 오래가고 착색도 좋아서 좋아용!! 근데 용량이 적다는게 유일한 단점이 아닐까...

리뷰가 도움이 돼요! 15 신고하기

1. 배민 등에 비해 점수 다양함
2. ⇒ 2점 : 12,000건 중 1% = 120건
3. 리뷰이벤트 없음
4. 상세 리뷰 조회 가능
5. HTML 가능

```
<div class="poll_type1">@</div>
<div class="poll_type1">@</div>
<div class="poll_type1">@</div>
::after
</div>
<div class="txt_inner">== $0
"힌스 글로우 톤이 유명해서 샀었는데 만족합니다. 케이스도 미쁘고 컬러도 이쁘고, 그리고 광택도 오래가고 착색도 좋아서 좋아용!! 근데 용량이 적다는게 유일한 단점이 아닐까..."
<!-- ## 리뷰 고도화 1차 ## : 태그 추가 -->
<!-- ## 리뷰 고도화 2차 ## 연관상품 -->
```

배경(3/3)

피부 타입 별
스킨케어 제품이 아닌, 색조화장 제품에서도 피부 타입
별로 중시하는 TOPIC이 있을까?

-배경지식-

피부타입 별로 스킨케어 볼 때 TOPIC 다름

건성(건조 피부타입) : 수분이 많은 제품 선호

유분감 있는 제품 선호

지성(기름 많은 타입) : 수분 가벼운 제품 선호

유분감 있는 제품 불호

복합성(중간 혹은 혼합) : 중간지점 어딘가 선호

연구과정 (변동가능)

1. 별점 별 리뷰 데이터 수집(HTML)
2. 수집 데이터 피부타입으로 분류
3. NLP전처리(토큰화 등)
4. 별점 1~5점 별로 **TOPIC** 추출
5. 만족,불만족 요인 분석
6. 만족 : 4,5점 / 중립 : 3 / 불만족 : 2,1 요인 분석
7. 고객 피부타입 별로 요인 분류
8. 피부타입과 요인의 연관 , 피부타입 별로 요인 비교

연구과정(1/4)

추천 키워드 전체

#사용감

#사용부위

#사용효과

#상품만족도

#색상

겔니

TOP187

★★★★★ 2025.03.07 · 매장

[옵션] 02 로 로즈

#text 680.05 × 38.59

지속이 오래돼요 | 발림성 아주 만족해요 | 수분감 아주 촉촉해요

“헐스 글로우 젤 틴트가 유명해서 사봤는데 만족합니다. 케이스도 이쁘고 컬러도 이쁨, 그리고 광택도 오래가고 착색도 좋아서 좋아용!! 근데 용량이 적다는게 유일한 단점이 아닐까...”

리뷰가 도움이 돼요! 15 신고하기

```
<div class="poll_type">@@</div>
<div class="poll_type">@@</div>
<div class="poll_type">@@</div>
::after
</div>
<div class="txt_inner">== $0
“헐스 글로우 젤 틴트가 유명해서 사봤는데 만족합니다. 케이스도 이쁘고 컬러도 이쁨, 그리고 광택도 오래가고 착색도 좋아서 좋아용!! 근데 용량이 적다는게 유일한 단점이 아닐까...”
<!-- ## 리뷰 고도화 1차 ## : 태그 추가 -->
<!--## 리뷰 고도화 2차 ## 연관상품 -->
```

1. 별점 별 리뷰 데이터 수집(HTML)

2. 수집 데이터 피부타입으로 분류

3. NLP전처리(토큰화 등)

4. 별점 1~5점 별로 TOPIC 추출

5. 만족,불만족 요인 분석

6. 만족 : 4,5점 / 중립 : 3 / 불만족 : 2,1 요인 분석

7. 고객 피부타입 별로 요인 분류

8. 피부타입과 요인의 연관, 피부타입 별로 요인 비교

OLIVE YOUNG

상품, 브랜드, 성분 검색

카테고리

오덕

랭킹

헬스+

LUXE EDIT

뷰티

스킨케어 >

스킨/토너

에센스/세럼/앰플

크림

로션

미스트/오일

스킨케어세트

마스크팩 >

시트팩

패드

페이셜팩

코팩

패치

클렌징 >

클렌징폼/젤

오일/밤

워터/밀크

필링&스크럽

타슈/패드

립&아이리무버

선크어 >

선크림

선크스틱

선크션

선크스프레이/선크패드

타닝/에프터션

메이크업 >

립메이크업

메이크업메이크업

아이메이크업

뷰티소품 >

메이크업소품

스킨케어소품

아이소품

헤어/바디소품

팔사/네일소품

뷰티디바이스

뷰티잡화

더모 코스메틱 >

스킨케어

바디케어

클렌징

선크어

마스크팩

맨즈케어 >

스킨케어

메이크업

쉐이빙

헤어케어

바디케어

프래그런스/라이프

헤어케어 >

샴푸/린스

트리트먼트/팩

헤어에센스

염색약/펌

헤어기/브러시

스타일링

바디케어 >

샤워/입욕

립케어

핸드케어

바디미스트

제모/악상

데오드란트

선물세트

베이비

로션/오일

향수/디퓨저 >

여성향수

남성향수

유니섹스향수

미니/고체향수

홈프래그런스

네일 >

일반네일

젤네일

네일팁/스티커

네일케어

연구과정(2/4)

목표 : 피부 타입 별 만족 및 불만족 토픽 분석
만족도 정의 방법 : 2가지

<<< 1~5 별점

<<< 상세 만족도
발색력, 지속력 등 3척도 상세리뷰



연구과정(3/4)

1. 별점 별 리뷰 데이터 수집(HTML)
2. 수집 데이터 피부타입으로 분류
3. NLP전처리(토큰화 등)
4. 별점 1~5점 별로 TOPIC 추출
5. 만족,불만족 요인 분석
6. 만족 : 4,5점 / 중립 : 3 / 불만족 : 2,1 요인 분석
7. 고객 피부타입 별로 요인 분류
8. 피부타입과 요인의 연관 , 피부타입 별로 요인 비교

**료타**
별점 4.5 · 겨울원톤 · 디크서클 · 중조
한달이상사용

★★★★☆ 2024.12.08
[옵션] N10 모브쇼킹
발색력 다소 아쉬워요 | 지속력 보통이에요 | 발림성 아주 만족해요 | 수분감 아주 촉촉해요
제형이랑 첫발색 다 좋는데 시간이 지나면 너무 원하게 빨개져서 손이 잘 안가네요..
이 리뷰가 도움이 됐어요! 05 0

**인려혜원쓰**
별점 4.5 · 겨울원톤 · 각질 · 모공
한달이상사용

★★★★☆ 2024.11.17 · 매장
[옵션] N10 모브쇼킹
발색력 보통이에요 | 지속력 보통이에요 | 발림성 아주 만족해요 | 수분감 보통이에요
지속력이 좋았던데 전 아니었어요.. 색도 별로.. 선택미스입니다
이 리뷰가 도움이 됐어요! 05 0

리뷰 중 자신의 피부 타입을 등록한 고객의 데이터를 사용

피부 타입 분류 (리뷰 필터 적용 가능)

☐ 나의 피부 컨디션 정보 불러오기

피부타입 지성 건성 복합성 민감성 약건성 트러블성
중성

피부톤 쿨톤 원톤 봄원톤 여름쿨톤 가을원톤
겨울쿨톤

피부 고민 잡티 미백 주름 각질 트러블 블랙헤드
피지과다 민감성 모공 탄력 홍조 아토피
다크서클

연령대 10대 20대 30대 40대 이상

성별 여성 남성

기업 MBTI, 처럼 분류가 가능함.
피부타입으로 분류해서 리뷰
디벨롭 : 피부 고민까지 상세 분류
피부톤 별 키워드는 뻔한 결과일 것같음

연구과정(4/4)

올리브영 후기의 색조화장 만족 및 불만족 요인 Topic과 피부타입의 연관성 연구 : 토픽모델과 머신러닝을 활용하여

*Journal of the Korean Data &
Information Science Society*
2021, 32(5), 1035-1047

<http://dx.doi.org/10.7465/jkdi.2021.32.5.1035>
한국데이터정보과학회지

IT 기업 직원의 만족 및 불만족 요인에 따른 이직률 예측:
토픽모델링과 머신러닝을 활용하여[†]

최진옥¹ · 신동원² · 이한준³

¹고려대학교 스마트미디어 서비스 연구센터 · ²명지대학교 산학협력단 · ³명지대학교 경영정보학과
접수 2021년 8월 27일, 수정 2021년 9월 7일, 게재확정 2021년 9월 7일

요 약

IT 산업에서의 이직은 기술 축적과 개발 및 관리에 있어 기업들에게 큰 어려움이 되고 있어 이에 대한 연구가 필요한 실정이다. 하지만 아직까지 IT 분야의 이직 및 인력 관리와 관련된 연구는 많지 않다. 이에 본 연구에서는 이직률에 영향을 미치는 직원들의 만족 요인과 불만족 요인들을 살펴보고 이를 활용한 IT 기업의 이직률 예측 모델을 제안하고자 한다. 이를 위해 온라인 기업 리뷰 사이트인 잡플래닛에서 국내 유가증권시장에 상장된 129개 IT 기업에 종사하는 직원들의 리뷰 21,589건을 수집하고 토픽모델링을 실시하였다. 추출된 토픽들을 활용하여 이직률에 대한 머신러닝 기반의 예측 모델을 제안하였다. 또한 변수 중요도 평가를 통하여 이직률에 영향을 미치는 직원들의 만족 및 불만족 요인별 영향 정도를 분석하여 제시하였다.

주요용어: 기업 리뷰, 머신러닝, 이직률, 직원 만족, 토픽모델링.

DS 리뷰논문 4,5
연구 방식이랑 비슷할 듯

후보 주제들

주제: 게임 데이터 및 리뷰를 활용해 매출 예측

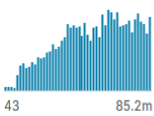
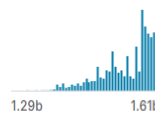
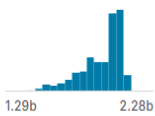
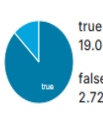
데이터:

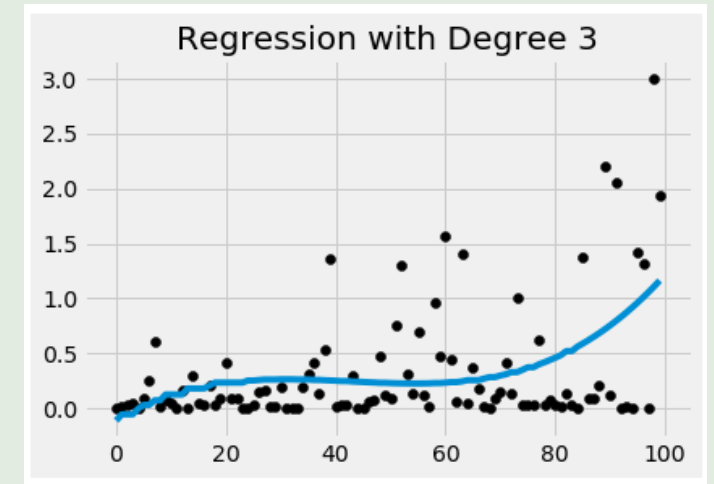
1. VGChartz에서 게임별 매출 정보를 웹 스크래핑 후

△ Name	△ Platform		△ Year		△ Genre		△ Publisher		# NA_Sales	# EU_Sales	# JP_Sales
games name	Platform of the games release		the game's release		Genre of the game		Publisher of the game		Sales in North America	Sales in Europe	Sales in Japan
11493 unique values	DS	13%	2009	9%	Action	20%	Electronic Arts	8%			
	PS2	13%	2008	9%	Sports	14%	Activision	6%			
	Other (12274)	74%	Other (13739)	83%	Other (10936)	66%	Other (14272)	86%			
Wii Sports	Wii		2006		Sports		Nintendo		41.49	29.02	3.77
Super Mario Bros.	NES		1985		Platform		Nintendo		29.08	3.58	6.81

모델: Regression model로 매출 예측

2. Steam API로 게임별 스팀 리뷰 데이터 추가

△ app_name	review_id	△ language	△ review	# timestamp_created	# timestamp_updated	✓ recommended
App name	Review id	Language of review	Review text	Review creation timestamp	Review latest update timestamp	Whether review recommends the ap
PLAYERUNKNOWN... 8%		english 44%	16657838 unique values			
Grand Theft Auto V 5%		schinese 17%				
Other (19084000) 88%		Other (8346967) 38%				
The Witcher 3: Wild Hunt	85181146	russian	Замечательная карточная игра в антураже фэнтези-рпг)	1611374238	1611374309	True
The Witcher 3: Wild Hunt	85181114	koreana	가성비 광	1611374195	1611374195	True



단점: VGChartz에선 오프라인 매출 데이터만 측정하고, 데이터가 오래됐거나 존재하지 않은 경우 있다.
구체적인 게임 총 매출 정보를 얻기가 힘들기에 해당 주제는 진행에 어려움이 있다.

주제: 트위터 피드 감성분석을 통한 국내 트렌드 예측 지표 제안

데이터:

1. 트위터 피드의 텍스트
2. 바이럴 되는 트윗의 리트윗/마음에 들어요/인용 수 등 수치데이터

데이터 수집 방법: Twitter API 사용 - snsrape 라이브러리, tweepy 라이브러리 등 활용.

목적

- 트렌드 변동 주기 확인
- 특정 감성이 트렌드 지속 기간 및 파급력에 미치는 영향 파악
- 트렌드가 지난 후 관심도 변동 측정

→ 트위터 감성분석을 통해 도출한 위의 정보들을 활용해, 트위터에서 트렌드를 예측할 수 있는 지표들을 제안한다.

단점

- 트렌드의 기준, 관심도 등 주요 변수를 정의하는 과정이 필요하다.
- 다소 광범위하기 때문에 데이터 수집 및 전처리 과정이 곤란할 수 있다.
- 연구 결과의 활용성이 모호하다.

주제 : 제로 칼로리 다이어트 식품의 소비자 기대와 불만 분석

1) 연구 목적 : 제로 칼로리 식품에 대한 소비자의 기대 요소와 부정적 반응을 체계적으로 분석하여, 향후 제품 개선 방향, 신제품 아이디어를 제시

2) 데이터 수집 : 쿠팡 웹 스크래핑

- '제로 칼로리', '제로음료', '제로간식' 등의 키워드를 활용한 웹 스크래핑
- 제품별 리뷰(텍스트 데이터) 및 평점(5점 척도) 수집

3) 분석 방법 :

감성분석(소비자 리뷰의 긍정/부정 감성 점수 분석)

토픽 모델링(기대 요소 및 불만 요인의 주요 주제 도출)

쿠팡 리뷰 데이터(5점 척도)에서 1점(불만족)과 5점(만족) 리뷰의 키워드 및 감성 차이 비교

4) 한계 : 쿠팡의 데이터 접근 제한 가능성, 가짜 리뷰 또는 일부 제품군에 리뷰가 부족할 가능성