

MUSINSA

무신사에서 살아남기



CONTENTS

MUSINSA

CONTENTS 1

서론

CONTENTS 2

데이터수집

CONTENTS 3

전처리

CONTENTS 4

단어빈도

CONTENTS 5

감성분석

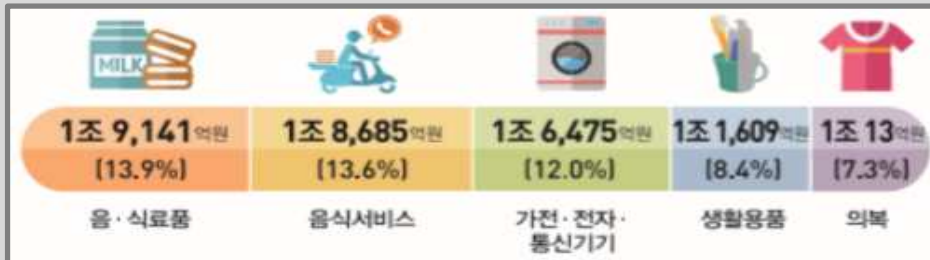
CONTENTS 6

주제분석

CONTENTS 7

결론

> 1) 코로나기간(2021년 2월) 온라인 쇼핑물 이용 동향



→ 의류쇼핑이 원활히 일어나는 가운데
2) 2021년 1월 기준 국내 온라인 패션몰 순위 1등이 무신사이다.

무신사는 여러 브랜드의 상품을 한 곳으로 모아 고객들이 손쉽게 어떤 상품이 인기가 있는지 알아볼 수 있는 쇼핑몰이다. 그로 인해 분류 별 랭킹에는 항상 **몇몇 브랜드가 독점**하는 현상이 발생했다. 랭킹이 높은 상품이 우선적으로 보이기 때문에 3) 부정적인 방법을 이용해 무신사 랭킹 순위를 조작하는 사례가 일어날 정도로 **랭킹이 중요**한 것이다. 그래서 순위가 낮은 브랜드나 신규 브랜드는 **상품별로 장·단점이 무엇인지, 어떤 점을 고려하고 보완하면 좋을지** 인지한다면 좋은 상품을 출시 할 수 있을 것으로 판단된다. 그리고 상품 별로 더 많은 정보를 표기한다면 고객 입장에서도

더욱 편할 것이다. 또한 이를 통해 브랜드 간의 경쟁이 생기면서 소비자 입장에서도 더욱 좋은 상품이 계속적으로 생겨나 고객만족도가 좋아질 것이다. 이렇게 되면 판매자, 구매자가 서로 **원원하게** 될 것이다.

1) http://kostat.go.kr/assist/synap/preview/skin/doc.html?fn=synapview388975_2&rs=/assist/synap/preview [2021년 2월 온라인쇼핑 동향 , 통계청, p 4]

2) <https://kkoongsbrothers.tistory.com/880> [2021년01월 기준 국내 온라인 패션쇼핑몰 순위 1위~100위 및 세계글로벌 랭킹]

3) http://m.apparelnews.co.kr/news/news_view/?idx=180175?cat=CAT127 [무신사 랭킹'이 뭐길래... 일부 입점사 순위 조작, 어페럴뉴스, 전종보기자]

100 페이지 중 1 페이지

1. 수집방법
requests로 상품의 URL을 저장 상품 별 셀레늄 접속

2. '무신사회원혜택' 광고
→ click()으로 닫기
긴 리뷰는 '더 보기'도 click!

3. '더 보기' 있는 리뷰는 태그가 하나씩 밀려남
원래 자리 텍스트는 빈칸
→ 더 보기 버튼 수 만큼 리뷰 텍스트를 추가 스크랩
빈칸인 리뷰는 제외

4. 별점 태그 class 자체의 텍스트에 별점 2배의 숫자
이를 'get_attribute' 로 추출하고 2로 나눔
> <https://fenderist.tistory.com/168>

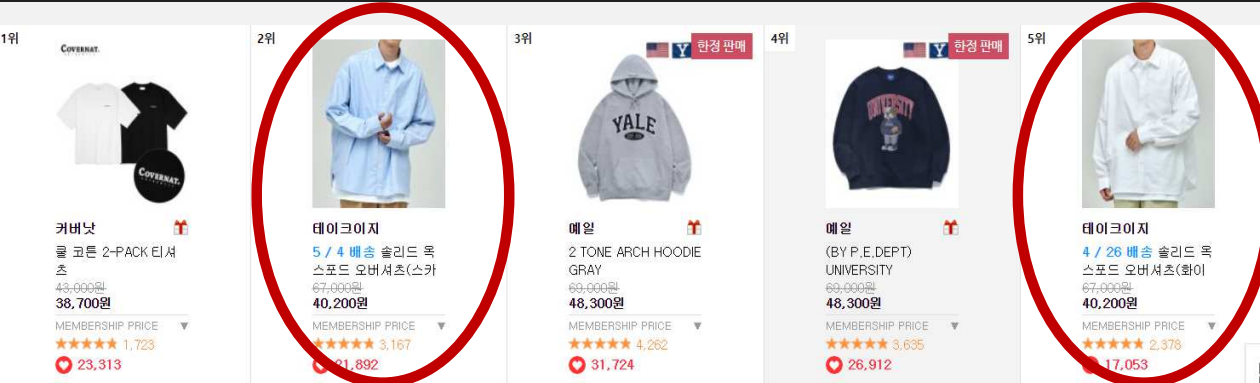
5. 여러 종류의 페이지 태그를 직접 할당
최대 20페이지 추출
→ 페이지 수가 20보다 작은 경우,
크거나 같은 경우 조건문으로 실행

6. 페이지 넘길 때 마다 같은 방법으로 반복 실행

7. 한 상품의 스크랩이 끝나면 상품의 대분류, 소분류, 가격을 스크랩 URL 번호를 이용하여 순위 저장

8. 스크랩 중간중간 얼마나 시간이 걸리고 몇 등 상품이 몇 개의 리뷰를 스크랩했는지 확인하는 문장도 구현

9. 가끔 오류 (element is not attached to the page document)가 발생
→ time.sleep으로 셀레늄 구동 중간중간 쉬는 시간을 줌



같은 브랜드의 같은 제품이지만 색상이 다를 경우 다른 제품으로 순위에 올라와 있는 상품이 있다. 이 때문에 중복된 리뷰는 제거

가격과 등수 데이터를 범위를 정해주어 가격 또는 등수 별로 상품들을 봤을 때 어떠한 장점과 단점, 문제점이 있는지 보면 좋을 것 같다고 생각하여 각각 4개의 범위로 라벨링 해주었다. 최대한 데이터 양을 균등하게 하기 위해 4분위수를 사용하여 범위를 정해주었다.

형태소 분석: Kiwi 사용 AND 명사만 추출, 단어 빈도분석, 감성분석, 주제분석에서 TF-IDF를 이용할 것이다.
 ※ **한국어 불용어 데이터**(<https://www.ranks.nl/stopwords/Korean>) 를 찾았고 추가적으로 필요 없다고 느껴지는 불용어를 추가하여 불용어 처리를 한 형태소 분석을 진행 하였다.
 그리고 '생지데님' 을 '생지데' 와 '님' 으로 분석하는 것과 비슷한 단어들을 따로 kiwi dictionary 에 추가해주었다.

	document	score	rank	big_select	small_select	brand	price
0	가볍고 핏이 낙낙해서 레이어로도 단일로도 잘 입을 수 있겠어요.	5.0	1	상의	반팔티셔츠	커버넌트	43000
1	m사려다 후기보고 s샀는데 s가 딱 맞네요 얇아서 여름에 입기 좋을거같고 가성비 좋...	4.0	1	상의	반팔티셔츠	커버넌트	43000
2	검은색 반팔이 없어서 샀는데 너무 잘 산거 같아요. 흰색도 이쁘고 살짝 비취지긴 한...	4.0	1	상의	반팔티셔츠	커버넌트	43000
3	깔끔하고 편해서 매일입네요 너무 좋아요 가격도 좋아요	5.0	1	상의	반팔티셔츠	커버넌트	43000
4	재질이 너무마음에들어요 싼값에 잘건져갑니다	5.0	1	상의	반팔티셔츠	커버넌트	43000

전처리 후 →

← 전처리 전

	document	score	rank	big_select	small_select	brand	price
	드로도 단일로도 잘 입을 수 있겠어요.	1	50등 이하	상의	반팔티셔츠	커버넌트	39000~59000
	여름에 입기 좋을거같고 가성비 좋...	0	50등 이하	상의	반팔티셔츠	커버넌트	39000~59000
	요. 흰색도 이쁘고 살짝 비취지긴 한...	0	50등 이하	상의	반팔티셔츠	커버넌트	39000~59000
	매일입네요 너무 좋아요 가격도 좋아요	1	50등 이하	상의	반팔티셔츠	커버넌트	39000~59000
	재질이 너무마음에들어요 싼값에 잘건져갑니다	1	50등 이하	상의	반팔티셔츠	커버넌트	39000~59000

상위권 상품

기장	250.554065
티	250.164052
두께	236.155076
가격	234.265301
추천	217.369233
길이	210.805583
맨투맨	204.086142
기모	200.562165
봄	199.808037
색상	181.479935
이너	179.515225

하위권 상품

가격	285.146771
두께	277.630065
기장	272.070678
여름	259.069216
추천	232.668840
티	227.882466
길이	221.629556
셔츠	210.534459
룩	204.017773
디자인	197.329999
기모	192.269096



전체적으로 빈도가 높은 단어 외 단어들 중 비교했을 때 차이가 있는 단어들을 분석

상의에서 순위별, 가격별 단어빈도를 분석해 보았을 때 각각 브랜드 별 대표적인 옷을 판매하더라도 **레이어드 티** 또는 **이너 티** 같은 기본 아이템도 판매하면 좋을 것으로 보인다. 또한 **셔츠**는 다른 상의에 비해 수요가 작기 때문에 셔츠 판매 시 타 브랜드와 차별점을 가져야 할 것이다.

저렴한 상품

레이어드	545.988802
만족	514.245752
재질	486.299704
색감	438.678448
구매	437.459154
색	374.257818
기장	349.515023
티	329.408308
여름	322.645919
마음	317.207212
느낌	309.819876
두께	308.802169
가격	295.425035

고가의 상품

색감	399.012779
만족	358.943345
색	298.775898
재질	290.399671
담	280.540291
구매	277.507803
마음	266.049461
느낌	249.579193
기모	241.578187
가격	224.405576
두께	204.899929
추천	190.368551
봄	187.489487
셔츠	179.002291



순위별 빈도분석: **맨투맨**, **이너** 같은 기본 아이템이 판매율이 높은 것으로 보인다. 비교적 **셔츠**는 손이 많이 가지 않는 제품으로 보인다.



가격별 빈도분석: **여름티**, **레이어드티** 등 얇은 티들이 가격이 저렴한 것으로 보인다. 비교적 **기모티**, **봄옷**, **셔츠** 등 두께가 있는 옷이 비싼 것으로 보인다.

상위권 상품

허벅지	347.089065
길이	346.387076
맘	341.195437
재질	316.173706
느낌	310.130848
마음	302.304640
색	300.671527
와이드	290.001234
옷	283.700512
수선	282.559713

하위권 상품

여름	354.276755
수선	335.824716
허벅지	330.355533
마음	327.656819
재질	325.073815
맘	319.517455
길이	318.827082
와이드	309.971439
색	296.924706
느낌	286.896118
옷	281.609774

전체적으로 빈도가 높은 단어 외 단어들 중 비교했을 때 차이가 있는 단어들을 분석

바지에서 순위별, 가격별 단어빈도를 분석해 보았을 때 바지의 경우에는 **와이드** 하고 **통이 큰** 바지가 요즘 유행하는 것으로 보인다. 그리고 **수선**을 하지 않아도 되는 제품이 수요가 높다. **청바지**의 경우 다른 바지에 비해 고가인 것으로 보인다.

저렴한 상품

재질	347.276114
길이	340.543086
와이드	331.044893
가격	329.991157
옷	301.195476
맘	296.532554
색	285.461224
허벅지	285.028156
마음	278.751124
색감	273.591124
통	271.816623

고가의 상품

허벅지	393.174566
맘	366.616633
수선	362.103094
마음	352.363918
느낌	339.286717
길이	328.269679
색	312.219945
청바지	300.123563
재질	289.980412
모드나인	287.531004
추천	276.929129
와이드	266.898850

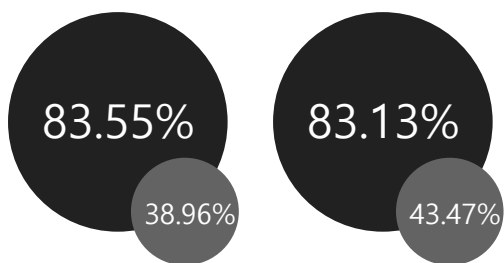
순위별 빈도분석: **수선**의 빈도를 보면 상위권의 상품은 수선을 하지 않아도 잘 맞는 상품이, 하위권에는 그렇지 않은 상품으로 보인다. 그 외에 큰 차이는 없어 보인다.



가격별 빈도분석: **와이드**의 빈도가 둘 다 높다. **청바지**는 비교적 비싼 것으로 보인다. 예외로 와이드한 바지는 허벅지가 널널하기 때문에 **와이드**와 **허벅지**의 빈도는 음의 상관관계로 보인다

감성분석에서는 앞서 진행했던 단어 빈도분석에서의 가설들에 대한 검증 및 추가 분석을 해보겠다. 별점을 기준으로 라벨링 한 score를 이용하여 상의 · 바지 내에서 순위, 가격 별 감성분석을 진행할 것이다. 감성분석을 통해 각각의 단점, 보완점(**부정의 단어**) 그리고 장점, 유지되고 고려되어야 할 점(**긍정의 단어**)이 무엇인지를 분석해 보겠다. 분석하기 앞서 감정 **감탄사는 제외**하고 비교를 진행하겠다. 모델 학습은 앞서 만들어 준 TDM들을 X, score를 Y로 하여 진행하였고 **경사 하강법**을 이용한 **adam**을 이용하였다. 모든 종류의 학습에서 'val_accuracy'는 변화가 크게 없었지만 'val_loss'는 줄어들었고 그래서 **EarlyStopping** 을 'val_loss' 로 하여 진행하니 결과가 더 좋게 나타났다.

상의 - 순위별



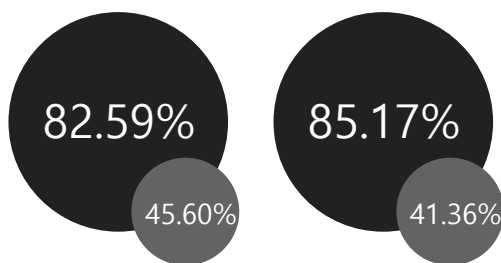
상위권

하위권

● : val_accuracy

● : val_loss

상의 - 가격별



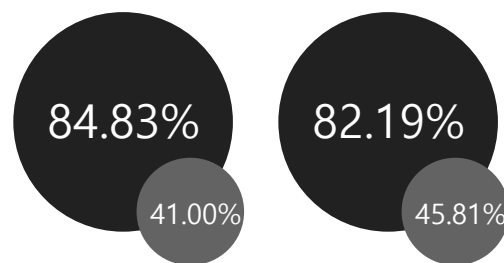
저렴 상품

고가 상품

● : val_accuracy

● : val_loss

바지 - 순위별



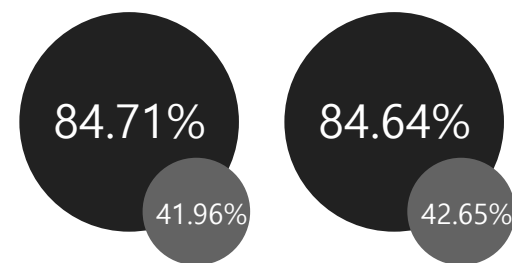
상위권

하위권

● : val_accuracy

● : val_loss

바지 - 가격별



저렴 상품

고가 상품

● : val_accuracy

● : val_loss

상의 상위권

이쁨니	0.998946	실밥	-0.677929
낭낭	1.008764	아쉬움	-0.595177
후드	1.087066	주름	-0.573170
찰떡	1.095474	차이	-0.560136
맘	1.127648	구멍	-0.554898
대박	1.203026	마무리	-0.550889
실물	1.252912	털	-0.539469
애용	1.368184	판	-0.531208
감사	1.459407	상태	-0.526622
최고	1.630585	봉	-0.494951

긍정

부정

상의 하위권

가운데	1.112431	값	-0.982765
주변	1.114545	실밥	-0.868339
여행	1.120723	털	-0.845253
무지티	1.129832	부분	-0.793055
아이템	1.263176	목	-0.785884
절기	1.318794	전반	-0.784253
최고	1.328025	마감	-0.712751
면도	1.333521	빨래	-0.699225
미니멀	1.419314	몸통	-0.683851
완벽	1.523325	초가를	-0.663279

긍정

부정



상위권 긍정의 단어구름



하위권 부정의 단어구름

상의에서 순위별 감성분석을 진행한 결과 **긍정**의 단어에는 긍정의 감탄사 외에 **미니멀**, **여행**, **무지티** 등이 있었다. **미니멀**은 요즘 유행하는 패션으로 **와이드한 바지에 간단한 무지티를 함께 입는** 룩을 말한다. **여행**은 어떤 리뷰에서 나왔는지 찾아보니 “여행가서 입었는데 겁나 이쁘네용 고냥 편하게 입었는데 사이즈도 딱 맞고 좋아요~”와 같이 **편하게 여행 갈 때 입기 좋은 옷**이 수요가 많은 것으로 보인다.

부정의 단어에는 실밥과 털, 주름, 마무리 등의 옷의 내구성에 대한 아쉬움을 표현하는 단어가 있었다. 이것 외에 하위권 부정의 단어를 보면 **값**, **빨래**, **목** 등이 있었다. **값**이 비싼 옷에 약간의 **흠**이 있다면 바로 불만을 가지는 것으로 보인다. 그리고 **빨래**를 했을 때 **옷이 변형**되는 것을 조심해야 하고, **목** 부분이 너무 **좁거나** 너무 **넓은** 것을 조심해야 할 것으로 보인다.

상의 저렴한상품

재민	0.740793	실밥	-0.584012
완전	0.743910	통	-0.548118
의향	0.745901	전반	-0.480224
감성	0.776524	교환	-0.459243
무지티	0.779218	어깨	-0.441480
강추	0.797373	목	-0.437836
실물	0.851646	군더더기	-0.403489
최고	0.894701	실망	-0.402668
핏	0.968001	몸통	-0.391192
감사	0.983376	털	-0.374735

긍정

부정

상의 고가상품

패딩	1.055899	아쉬움	-1.008332
별론핏	1.069382	부분	-0.785709
찰떡	1.078075	실밥	-0.745770
대박	1.081359	줄것	-0.717070
매력	1.097962	단면	-0.636909
번재	1.137994	별로	-0.633928
워싱	1.157778	하체	-0.623699
엘	1.184154	차이	-0.609546
아디다스	1.188546	머	-0.597032
최고	1.764600	목	-0.594833

긍정

부정

상의에서 가격 별 감성분석을 진행한 결과 저렴한 상품의 **긍정** 단어에는 긍정의 감탄사 외에 **무지티**가 있었다. **무지티**와 같은 **기본 아이템**이 수요가 높다는 것은 단어 빈도분석에서의 가설이 들어맞는 것이다. 고가 상품의 **긍정** 단어에는 **아디다스**와 같은 고객층이 두터운 유명 브랜드도 있었고 **워싱**과 같이 **디자인이 겸비된 옷**이 가격은 높지만 **수요가 높다는** 것을 볼 수 있었다.

부정의 단어에는 동일하게 **실밥**과 **털**, **주름**, **마감** 등의 옷의 **내구성**에 대한 아쉬움을 표현하는 단어가 있고, 외에 저렴한 상품에서 **교환**에 대한 언급이 많았다. 분석해보니 “선택한 색상은 아니지만 어차피 레이어드용으로 산거라 흰색도 무난해서 그냥 **교환** 귀찮아 입어요”와 같이 **저렴한 기본 아이템은 교환** 귀찮아서 입는다는 내용이 많았다. 기본 아이템이 저렴하다고 해서 **색상**과 **내구성** 부분에서 부족함이 있어서는 안될 것이다.



저렴상품 긍정의 단어구름

저렴상품 부정의 단어구름

바지 상위권

정핏 1.584279	아쉬움 -1.406743
무드 1.617227	부분 -1.051250
맞춤 1.631545	환불 -1.029309
장만 1.643837	값 -0.984458
치트키 1.812973	밀위 -0.872623
제로 1.928102	뒤 -0.835166
이용 1.998587	오염 -0.827846
모드나인 2.034339	깁기 -0.814469
교복 2.230527	뽑기 -0.800557
품질 2.547422	용도 -0.790878

긍정

부정

바지 하위권

생로랑 0.912129	부분 -0.849296
운동화 0.913066	양말 -0.701329
요즘 0.914960	영덩이 -0.657578
룩 0.932645	야외 -0.638989
중청 0.957686	꼬룩튀 -0.575459
크림진 1.028543	느낌 -0.544227
대박 1.089918	커서 -0.544072
예정 1.124021	후기 -0.494821
최고 1.188902	아쉬움 -0.494193
찰떡 1.403963	선택 -0.487512

긍정

부정

바지에서 순위별 감성분석을 진행한 결과 **긍정**의 단어에는 인기있는 **브랜드**와 **요즘 유행**이 어떤 것인지 나타난다. **제로**는 유행하는 **조거팬츠** 브랜드들이고 **모드나인**은 **치트키**라 불리는 **청바지** 브랜드이다. 그리고 **정핏**, **중청**, **크림진**은 **남친룩**이라 불리는 패션이다. 이렇게 바지는 **트렌드를 잘 따라가는 것이 중요**해 보인다.

부정의 단어에는 **깁기**는과 **꼬룩튀**, **영덩이**, **밀위** 등의 **바지의 불편함**에 대해 **아쉬움**을 표현하는 단어가 있었다. 외에 **뽑기**, **환불** 등이 상위권 부정의 단어에 있었다. 분석해 보니 “인기있는 상품이 라 당연히 마음에 들 줄 알았는데 **그때그때 옷이 다르다**, **뽑기**운이 있다” 라는 내용이 많이 있었고 그래서 **환불**을 요청하는 리뷰가 많았던 것이었다. 바지는 **사이즈나 불편함에 민감**하기 때문에 상품 제작을 할 때 **일정하게 만드는 것이 중요**해 보인다.



상위권 긍정의 단어구름



상위권 부정의 단어구름

바지 저렴상품

슬랙	0.820687	부분	-0.705197
핏	0.820932	골반	-0.688970
맨투맨	0.826292	보풀	-0.614014
만족	0.850687	밀위	-0.583138
무태	0.855595	재봉	-0.577695
맘	0.891909	불량	-0.570903
예정	0.893785	별점	-0.567073
와이드핏	0.901163	미듐	-0.535402
인생	0.916433	느낌	-0.507135
최고	1.103646	광택	-0.495184

그정



저렴상품 긍정의 단어구름

바지 고가상품

핏	0.783256	아쉬움	-0.531628
최고	0.809435	확정	-0.376687
맘	0.814396	개선	-0.350044
쫄쫄	0.821923	커즈	-0.349307
걱정	0.826162	생활	-0.344480
바지	0.846151	타이	-0.333768
크림진	0.898782	1센치	-0.314572
브랜드즈	0.933620	신죽	-0.311603
감사	0.943319	부분	-0.308493
모드나인	1.053586	먼지	-0.306864

그정



고가상품 긍정의 단어구름

바지에서 가격 별 감성분석을 진행한 결과 저렴
상품의 **긍정** 단어에서 **슬랙(스)**, 그 대표브랜드인
무텐(다드)를 볼 수 있다. 고가 상품의 **긍정** 단어
는 청바지 대표 브랜드인 **모드나인, 브랜드즈**가
를 볼 수 있다. **청바지가 슬랙스에 비해 비교적**
고가인 것을 알 수 있다. 외에 **트렌드**를 살펴보면
와이드 핏이, 청바지는 **크림진**이 **유행**인 것을 알
수 있다. 앞서 순위별 분석과 같이 **유행의 트렌드**
를 잘 따라가는 것이 중요해 보인다.

부정에서는 저렴한 **상품 부정**의 단어가 고가 상품 **부정**의 단어보다 **제품의 질**에 대한 문제(**광택, 재봉, 보풀**)와 **밀위, 골반** 등이 많은 것으로 보인다. 고가상품의 경우 **1센치, 신축** 등 보완했으면 하는 단어가 많았다. 분석하면 저렴한상품에 많은 **슬랙스**의 경우 **광택, 재봉, 허리끈, 보풀** 등 **제품의 질**에 대한 문제를 **신경** 써야 할 것이고, **청바지**의 경우 **1센치, 신축** 등 제품을 만들 때 **디테일적인 부분**을 **신경** 써야 할 것으로 보인다.

주제분석은 상의, 바지 이렇게 두 가지로 나누어 **상,하의 별로 주제를 분석**해 볼 것이다. 분류 별 리뷰들을 **Kiwi**를 이용하여 **불용어에 포함되지 않는** 단어들을 리뷰 별 단어들의 리스트로 묶어주고 **단어사전을** 만들었다. 단어들 중 전체문서의 **80% 이상에서 나온 단어**는 제외하고 **리뷰 20개 이하에서 나온 단어** 또한 **제외**하였다. 그리고 **LDA model** 학습을 진행하였다. 학습한 후 주제 수를 정해야 한다. 주제 수를 정하기 위해 **test_size**를 바꿔가며 그 때 마다 **loss(혼란도)**가 0에 가까운 것과 주제의 **응집도**와 **다양도** 등이 높은 것들을 이용하여 정하려 하였다. **test_size**를 바꾸고 토픽의 수도 여러가지로 시도해봤지만 적정 토픽의 수를 결정하지 못하였다. 그래서 **논문을** 이용하게 되었다.

논문에서는 ¹⁾ **토픽의 수를 정하는 것이** 아직까지 주제 분석에 있어서 가장 **고민**되는 것이라고 하였다. 그래서 여러가지 방법을 고안하는 중이고 그 중 추천하는 방법들을 선택할 것이다.

첫번째는 **혼란도**에 대한 것이다. 혼란도는 낮은 것(0에 가까운 것)이 좋지만 주제의 수가 작아지면 혼란도도 작아지기 때문에 **적절한 토픽 수를 정하기 어렵다**. 그래서 논문의 저자는 **혼란도가 변하는 값 중 완전한 곳이 적합한 토픽의 수**라고 말하였다. 두번째 방법은 **응집도**에 관한 것이다. 토픽의 수가 많아지면 응집도는 증가하게 된다. 응집도 또한 **증가하다가 완만해지는 구간이 있고 그 구간이 적절한 토픽의 수**라고 말하였다. 그래서 이 두가지를 비교하여 적절한 토픽의 수를 정하였다. **상의에 대한 분석**은 **test_size**가 0.3이고 **토픽의 수는 9개**일 때 가장 적합하다고 나왔다. **바지에 대한 분석**은 마찬가지로 **test_size**가 0.3이고 **토픽의 수는 8개**일 때 적합하다고 나왔다. 이제 상의 별, 바지 별 주제분석을 통해 **감정분석에서 나왔던 단어들과 어떤 단어들이 연관되어** 나오고 이를 분석하여 어떤 점을 **보완**하고 어떤 점을 **고려**하여야 할지를 분석해 보겠다.

1) 이대영(Dae Young Lee), 이현숙(Hyun Sook Yi). (2021). LDA 토픽 모델링의 적정 토픽 수 결정 방법 탐색. 교육평가연구, 34(1): 1-30

토픽	토픽 1번	토픽 2번	토픽 3번	토픽 4번	토픽 5번
단어	만족, 가격, 대비, 제품, 퀄리티, 세일, 원단, 데일리	옷, 색, 색깔, 청, 블랙, 매치, 검정, 크림, 아이보리	재질, 여름, 셔츠, 촉감, 심플, 편안, 옷감	맨투맨, 기본, 티셔츠, 흰색, 무탠다드, 기본템, 가성비, 기본티	레이어드, 티, 후드, 반팔, 넥라인, 이너, 품, 보들보들
주제	만족한 구매의 요소	상품의 색감과 매칭에 대한 요소	상품과 피부가 닿는 느낌	매칭이 쉬운 기본 아이템과 관련 브랜드	레이어드 제품, 레이어드 하기 좋은 제품
토픽	토픽 6번	토픽 7번	토픽 8번	토픽 9번	
단어	구매, 색감, 고민, 핏, 어깨, 색상, 할인, 두께감	사이즈, 오버핏, 스타일, 평소, 치수, 커서, 박시, 다운	두께, 기모, 봄, 가을, 겨울, 안감	느낌, 목, 시보리, 짱짱, 자수, 프린팅, 쏘쏘, 밑단	
주제	구매할 때 고려하는 요소	사이즈 관련 요소	계절 별 옷의 두께중요성	디테일적 상품의 고려요소	

먼저 토픽 4번을 보면 상의 단어빈도에서 가중치가 높았던 맨투맨 또는 기본 티셔츠 같이 매칭이 쉽고 코디하고 편한 기본 아이템이 인기가 많다는 것을 볼 수 있다. 토픽 5번 역시 단어빈도에서 언급했던 레이어드티에 대한 토픽이다. 추가적으로 레이어드를 할 때 잘 어울리는 상의로는 후드티라는 것을 확인할 수 있었다. 더불어 만족한 구매에 대한 토픽 1번을 보면 가격 대비 제품의 퀄리티나 원단이 좋으면 만족하고 정가에서 세일을 하는 옷이 만족도가 높은 것으로 보인다. 구매할 때 고려하는 요소가 있는 토픽으로 6번이, 디테일한 고려요소는 9번이 있었다. 색감이나 핏, 할인, 두께감을 일반적으로 고려하고 목이나 시보리가 짱짱한지, 자수나 프린팅이 관찮은지 등 디테일하게 고려하는 것으로 보인다. 8번 토픽을 보면 두께나 기모는 계절에 맞춰 제품을 출시해야하는 것으로 보인다.

1번 토픽: 가격 대비 제품의 **퀄리티**와 **원단**이 좋고 **세일**을 하는 제품, 데일리로 입을 수 있는 **가성비** 아이템이 만족스러운 것을 알 수 있다.

2번 토픽: 바지와 상의가 어떤 **색 조합**이 좋은지에 대한 토픽으로 청바지와 잘 어울리는 상의 색깔에는 블랙(검정)과 크림, 아이보리 같은 흰색계열의 상의가 매치하기 편한 것으로 보인다.

3번 토픽: **재질과 촉감** 등 옷감에 대한 토픽으로 **상품과 피부가 닿는 느낌**을 고객들이 중요시하게 여긴다는 것을 알 수 있고 특히 여름에는 날이 덥기 때문에 여름 옷에 더욱 신경 써야 한다는 것, 셔츠와 같은 뽀뽀한 재질의 상품에 더욱 신경 써야 한다는 것을 알 수 있다.

4번 토픽: 감성분석에서 봤듯이 맨투맨, 기본 흰 티셔츠 등 **기본 아이템, 기본 티가 유행**을 타지 않고 인기를 유지하는 것으로 보이고 기본 제품이 디자인이 첨부되지 않아 가격이 저렴하기 때문에 **가성비** 아이템인 것을 알 수 있다. 그에 대표적인 예로 무탠다드가 있다.

5번 토픽: **레이어드 티, 이너 티** 또한 4번 토픽과 같이 유행을 타지 않는 것으로 보인다. 그리고 반팔 티가 레이어드 아이템으로 좋고 보들보들한 느낌을 선호하는 것으로 보인다. 그리고 레이어드를 하면 잘 어울리는 옷은 후드 티인 것을 확인할 수 있다.

6번 토픽: 상의를 구매할 때 색감에 대한 고민이 많은 것으로 보인다. 그리고 핏, 어깨, 두께감 등을 걱정하고 할 인하는 제품을 찾는 것으로 보인다.

7번 토픽: 상의 사이즈가 **오버핏 스타일**이 유행하는 것으로 보이고 그렇기 때문에 평소 치수보다 크게 입어야 하는지 다운해서 입어야 하는지에 대한 부분을 고려하는 것 같다.

8번 토픽: 계절 별로 옷의 두께가 중요하다는 것을 알 수 있고 상품을 출시할 때 안감과 두께감이 어떠한지, 어떤 계절에 적절한 상품인지를 따로 표시해주면 좋을 것으로 보인다.

9번 토픽: **디테일한 고려요소**들로 목 부분, 시보리가 짙은 것을 선호하고 자수와 프린팅이 괜찮은지 고려하는 것으로 보인다.

토픽	토픽 1번	토픽 2번	토픽 3번	토픽 4번
단어	핏, 마음, 원단, 구매, 느낌, 만족, 색상, 재질, 디테일	기장, 수선, 청바지, 모드나인, 허리, 색감, 워싱, 브랜드, 밑위	와이드, 색감, 핏, 추천, 퀄리티, 가성비, 배송, 데님, 슬랙스	사이즈, 허벅지, 허리, 여유, 참고, 종아리, 치수, 엉덩이, 롤업, 체형, 교환
주제	만족한 구매의 요소	청바지 관련 키워드	바지 구매 시 추천하는 요소	사이즈 관련 고려 요소
토픽	토픽 5번	토픽 6번	토픽 7번	토픽 8번
단어	길이, 통, 키, 기장, 무릎, 신발, 포인트, 주머니, 발사, 밑단	색, 여름, 두께, 겨울, 블랙, 봄, 가을, 검정, 진, 그레이, 회색, 크림	가격, 재질, 슬랙스, 무탠다드, 가성비, 기본, 대비, 세일	허리, 밴딩, 운동, 발목, 편안, 재질, 불편, 타이트, 소재, 스판
주제	디테일적 고려 요소	계절과 색상의 관계	슬랙스 관련 키워드	운동과 바지의 관련 요소

가장 먼저 눈에 띈 토픽은 3번 토픽이다. 바지 구매 시 추천하는 요소로 단어 빈도분석에서 요즘 유행하는 것으로 추측했던 와이드 바지를 추천하고 이는 데님과 슬랙스 둘 다 있다는 것을 알 수 있다. 그리고 토픽 4번, 5번은 바지 구매 시 고려해야 하는 요소들이 나와있다. 사이즈를 고를 때 허벅지, 허리, 종아리, 엉덩이가 여유 있는 것을 선택해야 하는 것을 확인할 수 있다. 디테일적으로는 통, 무릎발사 주머니 밑단 등을 고려하고 신발과의 코디를 잘 생각해야하는 것으로 보인다. 2번은 청바지, 7번은 슬랙스 관련 키워드로 보인다. 6번은 계절 별로 색상이 관련성에 대해 말하는 것 같다. 그리고 구매 시 만족한 요소는 1번에 나와있었다. 핏, 원단, 느낌, 색상, 재질, 디테일이 마음에 들면 만족하는 것 같다. 바지는 사이즈와 관련성이 높아 운동과 바지 구매에 대한 고려사항이 토픽8번에 나오는 것으로 보인다. 불편하고 타이트하지 않는 재질, 밴딩이 있고 스판이 있는 소재를 선호하는 것 같다.

- 1번 토픽:** 구매를 할 때 만족하는 요소는 핏이 좋고 색상과 느낌, 재질, 디테일 등을 보는 것을 알 수 있다.
- 2번 토픽:** 청바지를 살 때 고려해야 할 요소가 기장과 허리, 워싱 색감, 밑위 인 것을 알 수 있다. 그리고 청바지는 수선을 많이 하는 것으로 보여 이를 유의하면 좋을 것 같다. 모드나인과 브랜드드가 청바지 대표 브랜드이다.
- 3번 토픽:** 바지를 구매 한 후 만족스러워서 추천을 하는 토픽이다. 와이드 바지를 많은 고객이 만족하여 추천하는 것을 알 수 있고, 와이드 데님과 와이드 슬랙스가 퀄리티 좋게 가성비 아이템으로 판매하는 것을 볼 수 있다.
- 4번 토픽:** 사이즈 관련 고려요소로 사이즈를 선택할 때 허벅지와 허리, 종아리, 엉덩이가 여유 있는지를 고려하고 자신의 치수와 체형을 잘 참고하여 선택해야 한다는 것을 알 수 있다. 바지 판매 시 정확한 사이즈를 기재하면 좋을 것으로 보인다.
- 5번 토픽:** 디테일한 고려요소들이 나온 토픽으로 길이가 적당한지, 통이 넓은지 등을 고려하는 것 같고 무릎발사가 있는 바지인지 확인하는 것을 알 수 있다. 이를 유의하여 상품을 만들면 좋을 것 같다. 그리고 주머니와 밑단을 어떻게 디자인 하여 포인트를 줄 것인지도 고려하면 좋을 듯 하다. 또한 어떤 신발과 잘 어울리는 바지인지도 함께 기재해주면 좋을 것 같다.
- 6번 토픽:** 계절과 바지의 관계로 두께와 색상을 중요하게 여기는 것으로 알 수 있다. 그래서 상품을 출시할 때 어느 계절에 적합한 두께이고 어떤 계절에 잘 어울리는 색인지를 알려주면 좋을 듯 하다. 모두에게 인기있는 색상은 검정, 그레이(회색), 크림 색의 바지로 보인다.
- 7번 토픽:** 슬랙스는 기본으로 하나씩 가지고 있는 아이템이기 때문에 가격면에서 가성비인 아이템, 세일하는 아이템을 선호하는 것으로 보인다. 가성비 슬랙스의 대표 브랜드에는 무텐다드가 있다.
- 8번 토픽:** 요즘 운동하는 사람이 많아져 운동하는 고객들이 불편함을 느끼는 요소들에 대한 토픽으로 불편하고 타이트한 재질을 입지 못하고 허리 밴딩이 있고 소재가 스판인 바지를 선호하는 것으로 보인다. 운동하는 고객을 주요 고객층으로 하는 바지를 출시하면 좋을 것으로 보인다.

코로나 기간 온라인을 이용한 의류 쇼핑이 원활히 일어나는 가운데 무신사가 국내 온라인 패션몰 1등을 하고 있다. 무신사에는 **랭킹이 중요한데 몇몇 브랜드가 독점**하는 현상이 발생했고 이를 해결하기 위해 **상품별로 장단점을 찾고 어떤 점을 고려하고 보완하면 좋을지를 분석**해보았다.

먼저 상·하의 모두 **가격이 저렴하고 가성비지만 퀄리티와 원단까지 좋은 아이템, 세일·할인하는 아이템을 선호**한다. 상·하의 모두 수요가 많은 색감은 검정, 흰색, 그레이색이다. **상의는 매칭이 쉬운 기본 아이템(맨투맨, 기본 티, 레이어드·이너 티)의 구매 양이 많은 것으로 보인다.** 상의에서는 색감에 대한 매칭이 중요하여 **색감적으로 어떤 바지와 매칭을 하면 좋을지 기재**해주면 좋겠다. 또한 상품과 피부에 닿는 느낌을 중요하게 여긴다. 그래서 **재질, 소재에 대해 신경을 써야 할 것이다.** 리뷰를 통해서도 어떤 소재가 좋은지는 알 수가 없었다. 사이즈는 **오버핏이 유행**이다. 때문에 **평소 치수보다 크게 입어야 하는지 또는 작게 입어야 하는지 헷갈려 하기 때문에 따로 표시를 해주면 좋을 것이다.** 상의는 계절과 두께가 연관성이 크다는 것을 볼 수 있다. **두께감이 어느 계절에 적합한지를 기재**해주면 좋을 것 같다. 마지막으로 **상의의 목 부분과 시보리가 짝짝한, 잘 늘어나지 않게 제작**하여야 할 것이다.

바지는 청바지와 슬랙스가 구매가 많고 그 중 **와이드한 바지가 유행**이기 때문에 와이드한 바지 출시를 고려하면 좋겠다. 청바지는 기장과 허리를 잘 맞춰야 하고 워싱의 색감이 이쁜 것이 인기가 많다. 그래서 청바지를 출시할 경우 허리와 밑위가 여유 있고 특히 워싱에 신경을 많이 써야 할 것이다. 대표 브랜드가 모드나인과 브랜드이기 때문에 이에 적대할 수 있는 상품을 출시해야 한다. **슬랙스는 저렴하고 가성비가 좋은 무탠다드가 대표적**이고 이에 적대할 수 있는 상품을 출시해야 할 것이다. 전체적으로 바지는 사이즈를 맞추기 어렵다. **허벅지, 허리, 종아리, 엉덩이 등을 여유 있게 만드는 것이 좋고 만든 바지의 정확한 사이즈를 따로 기재**하면 좋겠다. 또 바지는 무릎발사의 유무, 주머니·밑단 같은 포인트를 중요하게 본다. **무릎발사에 신경 쓰고 주머니·밑단 같은 경우 디자인적으로 잘 고려**해야 할 것이다. 계절과 바지는 두께·색상과 연관성이 있다. 그래서 **두께와 색상이 어떤 계절에 적합한 상품인지**를 기재해주면 좋겠다. 마지막으로 요즘 헬스를 하는 사람이 많아지고 있기 때문에 **기존 바지가 타이트한 고객층이 많다.** 그래서 같은 옷이지만 **밴딩이나 스판을 추가하여 운동하는 고객에게 적합한 바지를 새롭게 출시**해보는 것도 좋을 것 같다.

항목	점수	평가 근거
서론	2/2	"온라인 쇼핑물 상의 하의 별 고려, 보완할 점 찾기 " 를 주제로 하여 문제점과 함께 작성하였고 온라인 쇼핑물 이용이 많다는 현황을 보여줄 자료를 통계청의 연구 및 뉴스 기사를 인용하여 작성하였다.
데이터 수집	3/3	수업 내용을 반영하여 웹 스크랩을 수행하였고 추가적으로 request 와 셀레늄을 같이 사용, 별점을 위해 class 자체의 텍스트 추출, 중간중간 몇 개의 리뷰 스크랩 하였는지 확인하는 문장, 셀레늄 구동 중간중간 쉬는 시간을 주는 등 효과적인 방법을 추가적으로 조사하여 사용하였다.
전처리	3/3	수업 내용을 반영하여 전처리 하였고 추가적으로 가격, 등수를 4분위수를 이용하여 라벨링, 한국어 불용어 데이터로 불용어 처리를 한 형태소 분석, '생지데님'을 '생지데' 와 '님'으로 분석하는 것과 같이 한 단어를 나누어 분석하는 것을 따로 kiwi dictionary에 추가하는 등 추가적인 방법을 사용하였다.
단어 빈도	3/3	수업에 설명한 방식처럼 단어 빈도를 제시하고 주제와 관련된 해석 및 설명을 적절하게 제시하였다. 추가적으로 단어 빈도를 할 때 전체 데이터 빈도가 높은 단어들 중 상 하의로 데이터를 분할하여 분석을 해도 많은 단어는 빈도분석에서 제외하여 진행하니 결과가 좋게 나왔다.
감성 분석	3/3	수업에서 설명한대로 감성분석을 하고 결과를 해석하였다. 추가적으로 단어빈도에서도 사용했듯이 데이터를 상의 하의로 분류하고 거기서 가격 별, 등수 별로 나누어 총 4번의 감성분석을 진행하였다. 가격과 등수별로 어떤 차이가 있는지를 확인하기에 용이하였다.
주제 분석	3/3	수업에서 설명한대로 주제분석을 하고 결과 해석을 했다. 추가적으로 토픽의 수를 정하는 것에서 논문을 인용하여 혼란도와 응집도를 이용하여 토픽 수를 정하는 방법을 사용하였다. 앞서 진행한 단어 빈도와 감성분석의 결과와의 연관성을 이용하여 주제분석을 진행하였다.
결론	2/2	서론에서 제기한 문제에 부합하는 해결책들을 분석을 통해 도출하게 되었고 결론에 이를 정리하여 제시하였다.
합계	19/19	

THE END