텍스트 데이터 분석

CONTENTS

01 서론 **02** 데이터 수집

03 데이터 전처리 **04** 단어 빈도

05 감성 분석 **06** 주제 분석 07 ^{결론}

- 소비자들은 제품을 구매할 때 리뷰를 참고하지만, 인기 있는 제품일수록 리뷰 수가 많아 원하는 정보를 파악하기 어려움
- 이를 해결하기 위해 여러 제품의 리뷰를 요약·분석하여 특성을 정리하고, 소비 자의 용도와 선호도에 맞는 제품을 추천하는 방안을 고려
- 무릎보호대를 대상으로 비교 분석을 진행
- 무릎보호대는 운동 중 부상 방지 및 보조 용도, 혹은 무릎 관절 통증 완화를 위한 일상생활 용도로 사용됨
- 이러한 다양한 사용 목적에 따라 필요한 기능과 특성이 달라지므로, 리뷰 분석을 통해 목적별로 최적의 제품을 추천하는 것이 목표

데이터 수집

02.

- 네이버쇼핑에서 판매되는 무릎보호대 13개를 선정하여 데이터를 수집 Selenium을 활용해 제품 URL을 수집한 후, 리뷰를 검토하여 분석에 적합 한 제품 선택
- 이후, 긍정 리뷰 200개(랭킹순 정렬)와 부정 리뷰 200개(평점 낮은 순 정렬) 를 Selenium으로 크롤링
- 수집된 리뷰와 함께 부여된 평점을 포함한 데이터프레임을 생성하여 분석 준비

사이트 주소 :

https://search.shopping.naver.com/search/all?query=%EB%AC%B4%EB%A6 %8E%EB%B3%B4%ED%98%B8%EB%8C%80&cat id=&frm=NVSHAKW

03. 데이터 전처리

- 리뷰가 중복으로 수집 되었을 수 있기에 중복리뷰를 제거
- 평점이 5점 만점이여서 1~3점까지는 0점으로, 4~5점은 1점으로 변환
- 3점은 긍정보다는 그저 그런 느낌이므로 부정리뷰로 처리
- Stanza와 Kiwi를 비교해 형태소 분석을 시행하였고, Kiwi를 사용하는 것이 더 분석이 잘 된다 생각해서 Kiwi를 사용해 형태소 분석
- 동사를 포함할 경우 "하", "쓰" 같은 단어가 많이 나와서 명사만 추출했다.

04. 단어 빈도









1번 제품

2번 제품

3번 제품

4번 제품







6번 제품



7번 제품

11번 제품 12번 제품

10번 제품



04. 단어 빈도

- 전체적으로 무릎 보호대이므로 '무릎'의 빈도가 높음
- 운동 용품이다 보니 '운동'의 빈도가 높음
- 보호대는 보통 네오프렌(고무)재질이라서 냄새가 날 확률이 높기에 '냄새'의 빈도가 높음
- 착용했을 때 무릎을 잘 잡아주는지, 착용감이 좋고 나쁜지에 대해 리뷰를 작성했으므로 '착용'의 빈도가 높음
- 사이즈가 적당하다 혹은 너무 작아서 불편하다 같은 리뷰를 작성했으므로 '사이즈'의 빈도가 높음

1번 제품

	토큰	가중치		토큰	가중치
189	부분/NNG	-0.128459	437	타사/NNG	0.117732
69	냄새/NNG	-0.121618	169	밤/NNG	0.120556
301	위아래/NNG	-0.120434	108	되ハ٧	0.121021
346	자/NNG	-0.119489	433	친절/NNG	0.121591
236	시/NNG	-0.118832	119	≦ /∨∨	0.125031

- 냄새, 위아래 단어가 나오면 리뷰가 부정적일 가능성이 녹으
- -- 냄새가 많이 나고, 무릎 접는 부분이 불편하고, 위아래로 보호대를 움직이기 불편한 제품이라 볼 수 있음
- 타사, 밤, 친절 단어가 나오면 긍정적일 가능성이 높음 - 타사 제품에 비해 장점이 많고, 보호대를 끼면 밤에 무 릎이 아프지 않고, 친절하다고 볼 수 있음
 - 선형 모형을 이용해 감성 분석을 진행
 - 로지스틱 모형을 이용해 학습 했으며, 과적합을 피하기 위해 Early Stopping을 사용해 학습 중단 시점을 결정했다.

			_			
	토큰	가중치			토큰	가중치
208	불편/NNG	-0.156220		410	지지력/NNP	0.121794
261	아쉽/VA	-0.140233		473	III/NNG	0.121810
172	못하/VX	-0.130174		5	가족/NNG	0.122022
262	0101/NNG	-0.123516		197	보이	0.127028
207	불량/NNG	-0.120856		237	성능/NNG	0.129870

- 불편, 못하, 아이 단어가 나오면 리뷰가 부정적일 가능성 이 높음
- 불량이 많고, 불편하며 아이가 사용 못할 단점이 있는 제품이라 볼 수 있음
- 지지력, 피, 가족 단어가 나오면 긍정적일 가능성이 높음
- 피가 안 통할 절도로 지지력이 좋고, 가족한테 선물해줬 는데 만족 하고, 성능이 좋은 제품이라 볼 수 있음

3번 제	푿
------	---

	도근	가중지		도근	가장시
209	아쉽/VA	-0.122802	423	통증/NNG	0.119375
19	걱정/NNG	-0.119588	189	스쿼트/NNP	0.120498
170	살///	-0.117154	414	친절/NNG	0.120765
43	⊒/VA	-0.116777	198	신랑/NNG	0.120815
135	방/NNG	-0.116304	464	함/NNG	0.122515

- 아쉽, 걱정, 살 단어가 나오면 리뷰가 부정적일 가능성이 높음

- 제품이 아쉽고, 무릎을 잘 잡아줄지 걱정되며 착용시 살이 집히는 제품이라 볼 수 있음
- 통증, 신랑, 스쿼트 단어가 나오면 긍정적일 가능성이 높음
- 통증을 잘 잡아주고, 스쿼트 운동 시 도움이 되며 신랑한테 사줬을 때 만족하는 제품이라 볼 수 있음

4번 제품

	토큰	가중치		토큰	가중치
165	버리//	-0.127748	286	옷/NNG	0.117880
446	페이지/NNG	-0.121636	31	고민/NNG	0.117897
307	일본/NNP	-0.121634	160	받///	0.119606
484	환불/NNG	-0.119706	79	내려가///	0.120762
84	₩NA	-0.119696	310	일상/NNG	0.121322

- 페이지, 일본, 버리 단어가 나오면 리뷰가 부정적일 가능성이 높음
- 일본 제품이어서 거부감이 있고, 보호대가 1쪽만 와서 페이지를 잘 보고 사야 하는 제품이라 볼 수 있음
- 옷, 고민, 내려가 단어가 나오면 긍정적일 가능성이 높음 - 사이즈 고민하지 않아도 되고, 옷 위에 착용하고, 보호대가 잘 내려가지 않는 제품이라 볼 수
- 5번 제품

있음

ni.	토큰	가중치		토큰	가중치
266	아쉽/VA	-0.127918	36	교환/NNG	0.113955
182	밴드/NNG	-0.124840	465	∓∧v	0.117677
329	임/NNP	-0.122029	323	일/NNG	0.119942
109	대하//	-0.120247	38	구매/NNG	0.122445
470	필요/NNG	-0.117294	146	많/VA	0.122593

- 아쉽, 밴드, 필요 단어가 나오면 리뷰가 부정적일 가능성이 높음
- 제품이 아쉽고, 밴드가 약해서 고정이 잘 안되고, 살 필요가 없는 제품이라 볼 수 있음
- 교환, 구매 단어가 나오면 긍정적일 가능성이 높음사이즈 교환을 잘해주고, 제품에 만족해 같은 브랜드의 다른 제품을 구매할 생각을 가지고 있다고 볼 수 있음

		6번	제품		
31	토큰	가중치	게占	토큰	가중치
27	교체/NNG	-0.120299	7	Zt/NNG	0.151422
182	뭐/NP	-0.109623	374	雲/∨A	0.152889
348	저렇/VA	-0.109016	482	†⊦/√∨	0.155376
91	LHAV	-0.108627	372	종아리/NNG	0.158962
28	교환/NNG	-0.104106	274	알///	0.160498
		7번	제품		
	토큰	_		토큰	가중치
37		-0.135475	232	수술/NNG	0.126437
118	딸띠/NNG	-0.131018	290	어렵/VA	0.129141
92	CHHI/NNG	-0.129419	45	근육/NNG	0.130971
300	여름/NNG	-0.127141	392	주문/NNG	0.133467
450	E-OIE/NNG	-0.124807	359	잡///	0.136560
		ρН	제품		
	토큰	가중치	^ 🗖	토큰	가중치
193	생각/NNG	120 0	159	보호/NNG	
	참고/NNG		63		0.119237
	일어나///			클래식/NNG	
166	부분/NNG	-0.126001	401		0.120312
116		-0.124520	136		0.120459

감성 분석 05.

305

333

주/VX 0.205531

	토큰	가중치	" -		가중치					
249	안쪽/NNG -	0.137636	412	짱짱하게땡겨/NNG	0.128376	이쯔 무취 소프로 다이기 나이면 기보기 보저지의 기노성이 노오				
391	AI/NNB -	0.137401	301	일/NNG	0.128918	- 안쪽, 못하, 스포츠 단어가 나오면 리뷰가 부정적일 가능성이 높음 - 무릎 안쪽을 잘 지지해주지 못하고, 스포츠 할 때 사용하기 별로인 제품이라 볼 수				
162	못하/VX -	0.132332	142	III/NNG	0.131063	구표 단기를 본 MM에 1 M				
70	남/NNG -	0.130864	248	안정/NNG	0.132153	-안정, 좋아서 단어가 나오면 긍정적일 가능성이 높음				
221	스포츠/NNG -	0.130452	369	좋아서/NNP	0.135119	- 안정적으로 잘 잡아줘서 좋은 제품이라 볼 수 있음				
	토큰	10번 가중치	l 제·		가중치					
153	핑크/NNP	-0.121164	143	무릎보호대/NNP	0.124968	- 핑크, 이틀, 심하 단어가 나오면 리뷰가 부정적일 가능성이 높음				
299	0 ∰/NNG	-0.118761	456	하나/NR	0.127406	- 핑크색 보호대가 이상하고, 이틀이 지나도 냄새가 심한 제품이라고 볼 수 있음				
170	해/NNG	-0.115585	23	거/NNB	0.130430	- 하나, 만족 단어가 나오면 긍정적일 가능성이 높음				
327	잠스트보호/NNG	-0.108498	485	卒/NNG	0.132873	- 하나 더 사고 싶을 정도로 만족하는 제품이라고 볼 수 있음				
231	심하/VA	-0.107692	126	만족/NNG	0.136248					
	E ē	11번 가중치	제골	<u>.</u> 토큰	가중치					
52	₹/NNG	-0.142299	194			- 터지, 벨크로 단어가 나오면 리뷰가 부정적일 가능성이 높음				
27	터지//V	-0.139699	193	사이즈/NNG	0.169345	- 벨크로가 약하고 잘 터지는 제품이라고 볼 수 있음				
82	꼼꼼했음더좋/NNG	-0.139601	387	착용/NNG	0.171892	- 사이즈, 착용 단어가 나오면 긍정적일 가능성이 높음				
61	벨크로우/NNG	-0.136074	283	있사	0.172447	- 사이스, 역용 단어가 다오면 등정역을 가능성이 높음 ²⁴⁴⁷ - 사이즈가 잘 맞고 착용감이 좋은 제품이라고 볼 수 있음				

토큰

투크

12번 제품

가중치

가주치

160	//NNB	-0.143953	10	<u></u> == /∨X	0.101200	- (
112	뭐/NP	-0.142641	409	친구/NNG	0.162300	- 0
203	여성/NNG	-0.140800	270	윘VX	0.167844	
34	L-I/NP	-0.134808	403	추천/NNG	0.177796	- <u>-</u>
292	. 저한태/NNP	-0.131709	450	편하/VA	0.186171	0]

토큰

토큰

가중치

가증치

	1	1001		100 TO	\$100 M
487	확실/NNG	-0.128974	285	0 ##WA	0.143392
341	조임/NNP	-0.115255	189	비싸/va	0.147365
446	특성/NNG	-0.113251	211	☆/NNB	0.149019
334	제/NP	-0.109890	412	추천/NNG	0.151893
230	안쪽/NNG	-0.107911	172	배송/NNG	0.163774

- 여성 단어가 나오면 리뷰가 부정적일 가능성이 높음
 - 여성이 사용하기 별로인 제품이라 볼 수 있음
- 드리, 친구, 추천 단어가 나오면 긍정적일 가능성이 높음
- - 부모님에게 드리거나, 친구에게 선물해주거나 추천할 만한 제품이라 볼 수 있음

- 확실, 조임, 안쪽 단어가 나오면 리뷰가 부정적일 가능성이 높음무릎 안쪽을 잘 지지해주지 못하고, 조임이 확실히 약한 제품이라 볼 수 있음
- 이쁘, 비싸, 배송 단어가 나오면 긍정적일 가능성이 높음
- 디자인이 예쁘고 비싸지만 좋고, 배송이 빠른 제품이라 볼 수 있음

	무릎	loading		착용	, 운동 Ioading		무: word	르 교 Ioading		착용 word	loading		운 l word	동 loading		
270	OF SE	0.098281	529	아프	0.255687	65	구매	0.208547	294	바지	0.097952	587	윘	0.165338		
868	편하	0.106213	298	무류	0.265338	0	21	0.225082	799	활동	0.100945	374	사	0.244912		
158		0.132591	474	스쿼트	0.328638	221	CCH	0.481628	388	살	0.111128	330	보	0.256952		
370	早車	0.201843	609	운동	0.368489	337	旦車	0.570953	412	성능	0.130282	510	엄마	0.260736		
298	무릎	0.803010	789	착용	0.554830	214	등산	0.588398	12	감	0.796785	379	사용	0.261323		
675	잡	0.920245	243	CCH	0.639121	192	되	0.647371	355	불편	0.807574	476	아프	0.28066		
735	-	0.953604	892	하	0.794661	270	무릎	0.788045	687	착용	1.120962	778	10	1.490003		

- 무릎은 1번 주제와 관련도가 강하다.
- 착용,운동은 7번 주제와 관련도가 강하다.
- 편하, 느낌, 보호 같은 단어도 1번 주제와 관련
- 아프, 스쿼트 같은 단어는 7번 주제와 관련

- 무릎은 6번 주제와 관련도가 강하다.
- 착용은 3번 주제, 운동은 5번 주제와 관련

2번제품

- 구매, 등산 같은 단어는 6번 주제와 관련
- 불편, 바지 같은 단어는 3번 주제와 관련
- 엄마, 아프 같은 단어는 5번 주제와 관련

LSA, NMF, LDA 세 가지 방법 모두로 분석을 했을 때 해석하기가 편하다고 느낀 NMF방 법으로 주제분석을 시행하였다.

무릎 word loading 약하 0.108207 하 0.195786 286 무리 0.130790 753 크로스핏 0.112886 저 0.196068 오 0.161438 통증 0.121160 433 스퀴트 0.264778 저 0.162904 井 0.136918 착용 0.372555 489 안전 0.176322 무리 0.174814 사 0.373695 9 가능 0.176760 아프 0.593780 운동 1.681335 田 1.777823 무류 1.812282 563 운동 1.907238

- 무릎은 6번 주제와 관련도가 강하다.
- 착용은 11번 주제와 관련도가 강하다. - 운동은 17번 주제와 관련도가 강하다.
- 크로스핏, 통증 단어는 6번 주제와 관
- 려 - 스쿼트 단어는 11번 주제와 관련
- 안전, 무리 단어는 17번 주제와 관련

4번 제품

	-	무릎		착용			운동			
	word	loading		word	loading		word	loading		
35	겉	0.121433	138	남성	0.104463	35	겉	0.245427		
535	않	0.173842	435	성인	0.113035	752	지	0.261413		
242	런닝	0.180709	56	고정력	0.184613	667	자전거	0.281369		
866	통증	0.240663	933	해보	0.227000	844	E	0.292131		
81	구입	0.327010	15	감	0.743658	162	느끼	0.301467		
284	무릎	1.049339	890	편하	1.328278	232	Щ	1.302648		
524	아프	1.335865	796	착용	1.482664	608	운동	1.369743		

5번 제품

무	٠릎,	운동		777	5	
	word	loading		word	loading	
372	보호대	0.161729	924	해보	0.094799	
768	지	0.161766	147	내려가	0.102515	
186	달리기	0.192994	228	드내	0.106081	
606	운동	0.458914	745	좋아보여	0.112205	
523	아프	0.538122	821	첫날	0.113932	
243	때	0.795626	13	감	1.362241	
301	무릎	1.501127	812	착용	1.691305	

차요

- 무릎은 17번 주제와 관련도가 강하다.
- 착용은 8번 주제와 관련도가 강하다.
- 운동은 29번 주제와 관련도가 강하다.
- 무릎은 5번 주제와 관련도가 강하다.
- 착용은 6번 주제와 관련도가 강하다.
- 운동은 5번 주제와 관련도가 강하다.

주제 분석 06.

6번 제품							7번	제품			8번 제품									
	무	-릎		착용	3	5	농동		두	-릎	, 운동		ᄎ	용	무릎	를, 윤	온동		착성	용
	word	loading		word	loading		word	loading		word	loading		word	loading		word	loading		word	loading
477		0.074781	579	올리	0.130878	859	편하	0.227480	41	겉	0.394556	357	밀착	0.078774	210	도움	0.297052	219		
466			16	-	0.168865	715	8	0.239051	982	享	0.478080	183	다리	0.086513	741	지	0.318863	208	_	
			432		0.190754	300	_		763	X	0.654740	990	호르	0.097802	81	구매	0.351140	922	*	
160	느낌	0.115173					20000		340	무릎	1.131756	151	내리	0.099942	215	되	0.443061	362		
530	안정	0.120669	297	무릎	0.196386	880	하체	0.292788					m.1 mm	0.440040	609	0.5	0.540005	471	시	0.677812
297	0.0	0.686464	372	불편	0.286533	405	사용	0.670008	888	86	1.876647	628	0 ==	0.142313	609	운동	0.548935	19	감	0.822468
			909	힘들	0.436017	244	EER	0.792110	267	щ	2.019616	13	감	1.240148	246	CCS	0.703553	780	착용	2.171846
706	÷	0.980592	770	찬용	1.072001	507	-	1.010071	611	운동	2.246700	813	착용	1.336595	301	무릎	0.815289			

- 무릎은 3번 주제와 관련도가 강하다.
- 착용은 6번 주제와 관련도가 강하다.
- 운동은 10번 주제와 관련도가 강하다.
- 무릎은 0번 주제와 관련도가 강하다.
- 착용은27번 주제와 관련도가 강하다.
- 운동은 0번 주제와 관련도가 강하다.

- 무릎은 13번 주제와 관련도가 강하다. - 착용은 0번 주제와 관련도가 강하다.
- 운동은 13번 주제와 관련도가 강하다.

주제 분석 06.

9번 제품								10번 제품										11	1번	제품						
	무릎	<u>.</u>		착	용		운	동			무릎		Ž	착용		운분	동		무릎	-		착용	3		운	동
	word	loading		word	loading		word	loading		word	loading		word	loading		word	loading		word	loading		word	loading	_	word	loading
687	짱퐝	0.164231	601	있	0.143264	204	등산	0.116261	80	구인	0.149114	317	보	0.173664	439	신랑	0.080424	394	아프	0.077625	410	앞	0.110708	669	하체	0.108430
764	통증	0.193719	258	무름	0.143814	395	소재	0.122097	738	축구	0.152176	373	상담	0.175344	72	괜찮	0.088007	579	찍찍이	0.079528	509	재질	0.116667	667	하제	0.111861
0	가	0.224974	848	힘들	0.150631	592	일	0.134587	216		0.177420	121	남기	0.180397	329	부담	0.118049	623	타성	0.085675	109	버	0.133919	640	팔	0.141902
465	아프	0.233992	206	디자인	0.168108	184	되	0.248675		2000	0.206533	848		0.192570	307	배드민턴	0.129174	120			17	개종	0.141334	598	최고	0.193436
336	불편	0.234682	313	보	0.177682	177	도움	0.248946	209		0.227899	82		0.200450	75		0.227085	219		0.449857	265	보	0.180713	394	야프	0.293787
321	보호대	0.340995	16	감	0.558909	211	OH.	1.136947								0.0			-		7	71	0.976261	171	Щ	0.941840
261	무릎	1.464971	709	착용	1.168992	550	운동	1.332100	274	무릎	1.278796	11	감	1.136574	216		0.935979	539	30	1.146182				449	0.5	1.429786
									479	아프	1.295653	713	착용	1.683791	552	운동	1.051453	498	잡	1.226898	587	착용	1.206475		To	1.423100

- 무릎은 19번 주제와 관련도가 강하다.
- 착용은 21번 주제와 관련도가 강하다.
- 운동은 9번 주제와 관련도가 강하다.
- 무릎은 19번 주제와 관련도가 강하다.
- 운동은 3번 주제와 관련도가 강하다.
- 무릎은 1번 주제와 관련도가 강하다. - 착용은 7번 주제와 관련도가 강하다. - 착용은 13번 주제와 관련도가 강하다.

 - 운동은 8번 주제와 관련도가 강하다.

		무	릎		착	용
		word	loading		word	loading
	637	착용	0.359722	668	키로	0.125232
	427	아프	0.373089	171	두께	0.162526
	375	스쿼트	0.424073	12	감	0.254563
	490	운동	0.607801	637	착용	0.334075
	246	무릎	0.735879	168	되	0.376725
	194	ICH	0.936797	22	같	1.068198
	713	*1	1 205138	31	것	1.554885

- 무릎은 7번 주제와 관련도가 강하다. 착용은 4번 주제와 관련도가 강하다.
- 운동은 7번 주제와 관련도가 강하다.

	무	릎		ž	착용		2	운동
	word	loading		word	loading		word	loading
523	주	0.110058	159	드리	0.114933	144	도움	0.118658
271	부담	0.139910	394	어머니	0.124612	627	필수품	0.125828
392	양호	0.158469	192	리뷰	0.128259	27	걱정	0.138404
376	아프	0.210027	383	않	0.298587	602	통증	0.171550
149	되	0.378441	64	굿	0.438471	632	하체	0.195914
267	보호	1.022945	11	감	0.829895	175	CCH	0.734138
223	무릎	1.281671	565	착용	0.942276	433	운동	1.258672

- 무릎은 8번 주제와 관련도가 강하다.
- 착용은 17번 주제와 관련도가 강하다.
- 운동은 5번 주제와 관련도가 강하다.

결론

07.

1번 제품: 냄새가 나지만 성능이 나쁘지 않아 스쿼트 같은 운동을 할 때 좋다.

2번제품: 착용감이 불편하지만 지지력이 좋고, 등산에 유용하며 부모님이 사용하기 좋다.

3번제품: 착용감은 별로지만 통증을 잘 잡아주고 크로스핏, 스쿼트 같은 운동에 좋다.

4번제품: 일본제품이지만 착용감이 좋고 런닝, 자전거 같은 운동에 좋다. 5번제품: 밴드가 약하지만 달리기 같은 운동에 좋다.

6번제품: 가격도 비싸고 잘 잡아주기에 스쿼트, 하체운도에 좋다.

7번제품: 여름에 끼기는 덥지만 수술한 사람이 끼면 좋고, 통증 완화에 좋다.

8번제품: 보풀이 일어나지만 보호력이 좋다.

9번제품: 안정적이지만 약해서 등산에 좋다.

10번제품: 냄새가 나지만 잘 지지해서 축구, 등산, 배드민턴에 좋다.

10면세품: 냄새가 나시면 잘 시시해서 국구, 등산, 배드민단에 좋다. 11번제품: 내구성이 약하지만 착용감이 좋다.

12번제품: 여성에게는 너무 단단하지만 남성이 끼기에 좋아 하체 운동에 좋다.

13번제품: 지지력이 약하지만 디자인이 좋고 부모님 선물에 좋다.

- 남성이고 헬스를 좋아하는 소비자의 경우에는 1번(냄새 상관하지 않는 경우), 3번, 6번(가격 생각 안하는 경우), 12번 제품을 추천 할 수 있다

- 가벼운 운동을 한다면 2번, 4번, 5번, 9번, 10번 제품을 추천 할 수 있다.
- 무릎이 좋지 않아 보호하려 한다면 7번(두꺼워도 상관 없는 경우), 8번(보풀 신경 안쓰는 경우), 11번, 13번 제품을 추천 할 수 있다.