

# 보건복지 ISSUE & FOCUS

**KIHASA**  
한국보건사회연구원  
Korea Institute for  
Health and Social Affairs  
www.kihasa.re.kr

제324호 (2016-24)  
발행일 2016. 10. 10.  
ISSN 2092-7117

발행인 김상호 발행처 한국보건사회연구원 (30147) 세종시 시청대로 370 세종국책연구단지 사회정책동(1~5층) TEL 044)287-8000 FAX 044)287-8052

## 2016년 소셜 빅데이터를 활용한 통일인식 동향 분석<sup>1)</sup>



송태민  
정보통계연구실 선임연구위원

- 통일에 대한 찬성 비율은 소셜 빅데이터 감성 분석 결과 2011년 55.0%, 2012년 62.0%, 2013년 57.7%, 2014년 58.2%, 2015년 56.2%, 2016년 59.8%로 나타남.
- 핵무기와 통일대박이 동시에 언급된 문서는 통일에 대한 중립과 찬성의 확률이 높아 '통일대박론'이 통일의 필요성에 대한 인식에 영향을 준 것으로 나타남.
- 통일에 대한 찬반, 통일 방법 등과 관련한 국민의 통일인식 조사와 더불어 소셜미디어에서 수집된 빅데이터를 활용, 분석하면 통일인식 예측은 더욱 신뢰성이 있을 것으로 봄.

### 1. 서론

- 현 정부는 '통일'을 핵심 국정과제 중 하나로 설정한 데 이어 통일에 대한 준비와 세부 정책을 다양하게 추진해 오고 있음.
  - 한반도 신뢰 프로세스를 통한 남북 관계 정상화를 주요 국정과제로 설정한 데 이어 2014년 대통령의 '통일은 대박이다' 한마디로 평화통일의 기반 구축에 적극적으로 나설 것임을 선언함.
  - '통일대박론'은 통일된 한반도에 대한 비전을 제시했으며, 이에 대해 북한의 불안정한 정치 상황으로 인한 실질적 흡수통일 준비라는 해석부터 경제적 효과에 중점을 둔 전략적 판단이라는 설까지 다양하게 제기됨.<sup>2)</sup>
- 최근 '한반도 정세'는 북한의 로켓 발사, 개성공단 잠정 폐쇄, 북한에 대한 유엔의 강력한 제재 결정 등으로 국민의 통일에 대한 관심과 열기가 다양하게 변화되고 있는 상황임.

1) 본 연구는 소셜 빅데이터를 분석한 것으로 통일인식에 대한 공식 통계로 인용할 수 없음.

2) 최영준, 황태희, 최우선, 주형민(2016). 박근혜 정부의 통일정책의 정책수용성에 대한 실증연구: 2013년과 2015년 국민인식조사 비교. 국제관계연구 21(1), p.6.

- 통일에 대한 한국인의 전반적인 관심도는 2010년 52.6%에서 2014년 82.6%로 증가하였으나 20대의 무관심은 다른 세대에 비해 높은 상태로 지속됨.<sup>3)</sup>
- 최근의 남북 관계는 모든 세대의 통일인식에 강하게 영향을 미치고 있으나 신자유주의통일세대(19~30세)는 통일 필요성에 대한 인식이 낮은 것으로 나타남.<sup>4)</sup>
- 한반도의 통일은 남북한의 문제이자 동북아 주변국의 미래를 좌우할 국제적 사안이며<sup>5)</sup> 국내 차원에서는 통일에 대한 국민적 공감대 형성과 통일을 맞이할 수 있는 역량 구축이 필요함.<sup>6)</sup>
- 평화로운 통일 한국을 실현하기 위해서는 주변국들의 반응과 함께 우리 국민들의 통일에 대한 인식과 태도를 분석할 필요가 있음.<sup>7)</sup>

## 2. 소셜 빅데이터를 활용한 통일인식 예측 모형

### 가. 분석 대상 및 분석 방법

- 본 연구에서는 88개의 온라인 뉴스 사이트, 9개의 게시판 및 주요 커뮤니티, 1개의 SNS(트위터), 4개의 블로그, 2개의 카페 등 총 104개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 온라인 문서(버즈)를 소셜 빅데이터로 정의함.
  - 통일 토픽(topic)은 모든 관련 문서를 수집하기 위해 '통일'이라는 토픽을 사용하였으며, 유사어로 '남북통일, 한반도통일'이라는 용어를 사용함. '통일신라, 통일교, 통일전망대, 세계통일' 등은 불용어(stop-words)로 처리하여 수집, 분석에서 제외함.
  - 소셜 빅데이터 수집은 2016. 1. 1.~2016. 3. 31.(3개월)의 기간에 해당 채널에서 요일, 주말, 휴일을 고려하지 않고 매 시간 단위로 이루어졌으며<sup>8)</sup> 수집된 총 9만 5507건의 텍스트 문서를 본 연구의 분석에 포함함.
- 국민의 통일인식을 파악하는 데 가장 효율적인 예측 모형을 구축하기 위해 머신러닝의 랜덤포리스트(Random Forest), 의사결정트리, 연관규칙 분석을 사용함.
  - 연관규칙의 분석 알고리즘으로는 선형적 규칙(apriori principle)을 사용하였고, 의사결정트리 형성을 위한 분석 알고리즘으로는 CHAID(Chi-squared Automatic Interaction Detection), 머신러닝 분석에는 R 3.2.1과 IBM SPSS 23.0을 사용함.

### 나. 분석 결과

- 통일 관련 감성 분석(Opinion Mining)과 주제 분석(Text Mining)
  - 통일 감정 키워드는 문서 수집 이후 요인 분석과 주제 분석을 통하여 찬성(13개: 관심, 필요, 중요, 노력, 신뢰, 평화

3) 아산정책연구원(2015). 한국인의 대북·통일 인식 변화, 아산정책연구원. pp.25-26.

4) 김병조(2015). 한국인의 통일인식 2007~2015. 세대별 격차와 세대 내 분화, 서울대학교 통일평화연구원 통일과 평화 7(2), pp.3-41.

5) 김규륜(2013). 한반도 통일의 미래와 주변 4국의 기대. 통일연구원. p.3.

6) 송태민(2015). 소셜 빅데이터를 활용한 국민 통일인식 동향 분석. 보건복지 ISSUE & FOCUS. 제269호.

7) 위의 책.

8) 본 연구를 위한 소셜 빅데이터의 수집 및 토픽 분류는 '(주)SK텔레콤 스마트인사이트'에서 수행함.

(평화적), 든든, 다행, 공감, 행복, 사랑, 위대, 대박), 반대(18개: 쪽박, 천박, 경박, 빨갱이, 환장, 친절, 한심, 문제, 위협, 압박, 포기, 재앙, 혼란, 분열, 갈등, 아픔, 고통, 불신)로 분류하였으며 찬성과 반대의 감정을 동일한 횡수로 표현한 문서는 중립의 감정으로 분류함.

○통일에 대한 주요 안보와 이슈는 주제 분석을 통하여 ‘핵무기, 정상회담, 휴전선, 남북공동선언, 간첩, 천안함, 이산가족상봉, 통일대박, 통일비용’의 9개 요인으로 그룹화함.

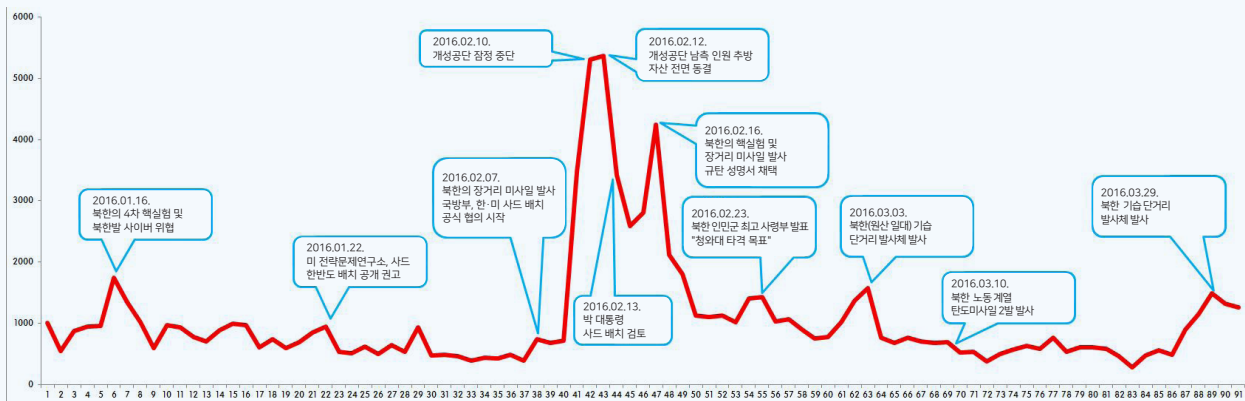
■ 통일 관련 온라인 문서(버즈) 현황

○2016년 1월부터 3월까지 통일 관련 온라인 문서의 추이를 분석한 결과 [그림 1]과 같이 북한과 관련한 이슈가 발생하면 커뮤니케이션이 급증하는 양상을 보임.

- 2016년 1월 6일 북한의 4차 핵실험으로 인해 문서량이 증가하였으며 특히 2월 10일 북한의 개성공단 잠정 폐쇄 발표 후 문서량이 급증한 것으로 나타남.

○통일에 대한 찬성 감정은 대박, 평화, 필요, 노력, 중요, 사랑, 관심, 신뢰 등의 순으로 높게 나타났으며 통일에 대한 반대 감정은 문제, 위협, 쪽박, 포기, 갈등, 고통, 압박, 분열 등의 순으로 높게 나타남[그림 1].

[그림 1] 통일 관련 온라인 문서 일별 추이 및 찬성·반대 감정



2016년 통일 찬성 감정



2016년 통일 반대 감정



■ 통일 관련 국민 인식 변화

○ 통일의 필요성에 대한 국민의 인식은 조사 기관별로 차이를 보이거나 소셜 빅데이터 감성 분석 결과 2011년 55.0%, 2012년 62.0%, 2013년 57.7%, 2014년 58.2%, 2015년 56.2%, 2016년 59.8%로 나타남(표 1).

〈표 1〉 연도별 통일 관련 국민의 통일인식

(단위: %)

연도	찬성			중립			반대		
	소셜 <sup>1)</sup>	서울대 <sup>2)</sup>	KBS <sup>3)</sup>	소셜 <sup>1)</sup>	서울대 <sup>2)</sup>	KBS <sup>3)</sup>	소셜 <sup>1)</sup>	서울대 <sup>2)</sup>	KBS <sup>3)</sup>
2011	55.0	53.7	74.4	18.3	25.0	19.9	26.7	21.3	5.7
2012	62.0	57.0	68.4	10.7	21.6	24.6	27.3	21.4	7.0
2013	57.7	54.8	69.1	8.7	21.5	21.2	33.6	23.7	9.7
2014	58.2	55.8	-	9.0	22.5	-	32.7	21.7	-
2015	56.2	51.0		8.3	24.3		35.5	24.8	
2016	59.8			12.6			27.6		

주: 1) 본 연구의 소셜 빅데이터 감성 분석 결과.  
 2) 서울대학교 통일평화연구원이 매년 7월부터 8월 사이(3주간) 1200명을 대상으로 대면 조사한 결과.  
 3) KBS 방송문화연구소(2013). 2013년 국민 통일인식 조사.

○ 개성공단 중단 및 사드 배치 검토 전(2016.1.1.~2016.2.12.) 통일에 대한 찬성 인식은 57.9%에서 이후 일주일간<sup>9)</sup>(2016.2.13.~2016.2.19.) 62.6%로 높아졌다가 일주일 후(2016.2.20.~2016.3.31.) 61.1%로 낮아짐. 이 기간 통일에 대한 반대 인식은 26.9%→26.4%→29.2%로 변화함(표 2).

〈표 2〉 개성공단 중단 및 사드 배치 검토 전후의 국민 통일인식

N(%)

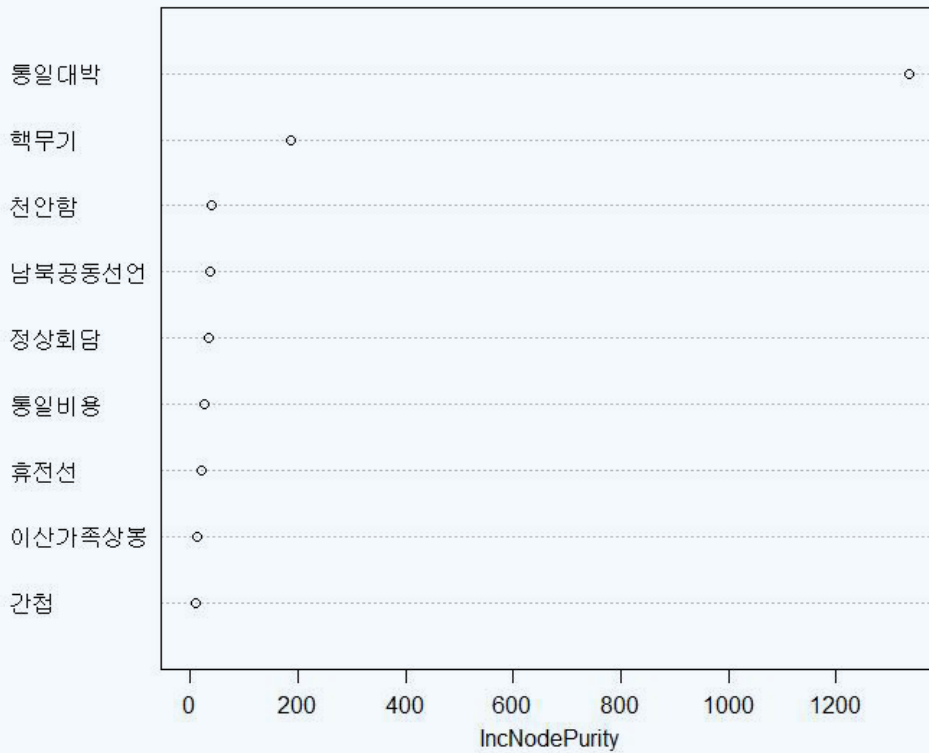
	찬성	중립	반대	계
2016.1.1. - 2016.2.12.	7,039(57.9)	1,849(15.2)	3,279(26.9)	12,167
2016.2.13. - 2016.2.19.	2,613(62.6)	460(11.0)	1,101(26.4)	4,174
2016.2.20. - 2016.3.31.	5,039(61.1)	796(9.7)	2,408(29.2)	8,243
계	14,691(59.8)	3,105(12.6)	6,788(27.6)	24,584

■ 통일에 대한 국민 인식 예측 모형

○ 랜덤포리스트 분석을 활용하여 통일인식에 영향을 주는 주요 안보와 이슈 요인을 살펴보면 [그림 2]와 같이 통일 인식(찬성, 중립, 반대)에 가장 큰 영향을 미치는(연관성이 높은) 주요 안보와 이슈 요인은 ‘통일대박’이었으며 이어 핵무기, 천안함, 남북공동선언, 정상회담, 통일비용, 휴전선, 이산가족상봉, 간첩 순으로 나타남.

9) 'URL을 통해 확산되는 온라인 문서(자살 등)는 약 3주 정도의 생명 주기를 가지며 발생 후 첫 주에 급속히 전파되는 경향을 보인다[National Information Society Agency(2012). Implications for Suicide Prevention Policy of Youth Described in the Social Analysis. Seoul, Korea: Author.]'에 따라 일주일 전후에 대한 인식을 분석함.

[그림 2] 랜덤포리스트 모델의 통일인식의 중요도



○ 랜덤포리스트의 중요도에 나타난 주요 안보와 이슈 요인들이 통일인식에 미치는 영향에 대한 로지스틱 회귀분석 결과는 (표 3)과 같이 천안함, 핵무기, 통일비용 순으로 찬성보다 반대의 확률이 높으며 그 외 통일대박, 남북공동선언, 이산가족상봉(p<.05), 휴전선 순으로 반대보다 찬성의 확률이 높은 것으로 나타남.

〈표 3〉 통일인식에 영향을 미치는 안보·이슈 요인<sup>1)</sup>

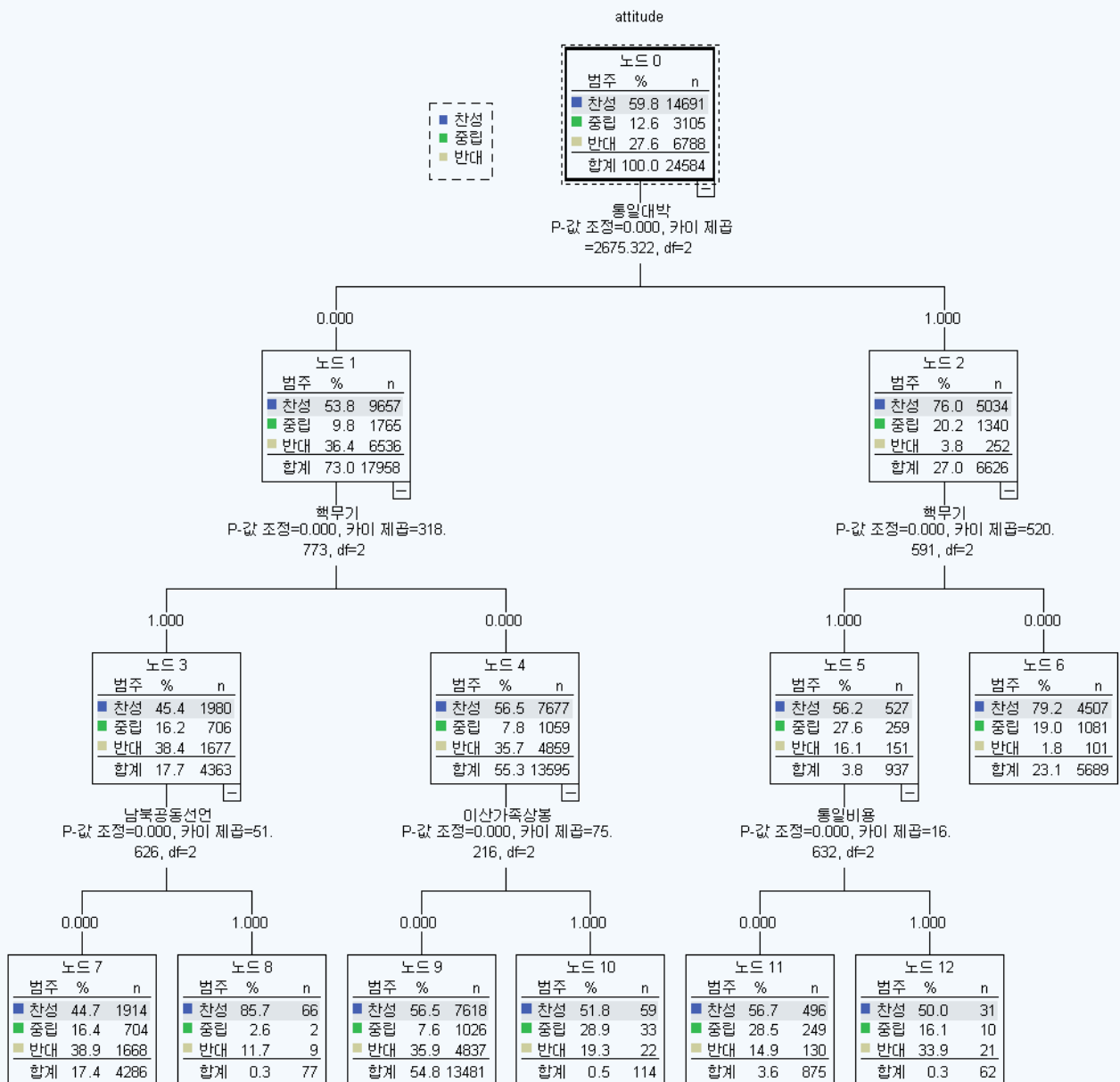
변수	찬성				중립			
	b <sup>†</sup>	S.E. <sup>‡</sup>	OR <sup>§</sup>	P	b <sup>†</sup>	S.E. <sup>‡</sup>	OR <sup>§</sup>	P
핵무기	-.412	.039	.663	.000	.330	.054	1.390	.000
정상회담	.000	.069	1.000	.996	.451	.089	1.570	.000
휴전선	.271	.099	1.311	.006	.623	.125	1.864	.000
남북공동선언	1.193	.251	3.298	.000	-.465	.469	.628	.321
간첩	.073	.198	1.076	.710	.760	.234	2.138	.001
천안함	-.846	.146	.429	.000	-.607	.188	.545	.001
이산가족상봉	.375	.148	1.454	.012	.693	.186	1.999	.000
통일대박	2.577	.067	13.156	0.000	3.080	.074	21.764	0.000
통일비용	-.363	.089	.696	.000	-.347	.133	.707	.009

주: 1) 기본 범주는 반대, †Standardized coefficients, ‡Standard error, § odds ratio.

○ 통일인식 예측 모형에 대한 의사결정나무에서 [그림 3]과 같이 뿌리 마디의 통일에 대한 감정 비율은 찬성 59.8%, 중립 12.6%, 반대 27.6%임.

- 뿌리 마디 하단의 가장 상위에 위치하는 요인이 종속변수에 대한 영향력이 가장 높은(관련성이 깊은) 요인이므로 ‘통일대박’이 가장 영향력이 큰 것으로 나타남(온라인 문서에 ‘통일대박’ 요인이 있으면 찬성은 이전의 59.8%에서 76.0%로 증가함).
- ‘통일대박’ 요인이 없고 ‘핵무기’ 요인과 ‘남북공동선언’ 요인이 있는 경우 찬성은 이전의 45.4%에서 85.7%로 크게 증가함.

[그림 3] 통일인식 예측의 의사결정나무 모형



○(표 4)와 같이 주요 안보와 이슈에 대한 통일인식 연관성 예측에서 찬성 확률이 가장 높은 연관규칙은 {통일대박}⇒{찬성}이며 세 변인의 연관성은 지지도 0.05, 신뢰도는 0.70, 향상도는 4.57로, 온라인 문서에서 통일대박이 언급되면 언급되지 않는 문서보다 통일에 대한 찬성 확률이 4.57배 높아지는 것으로 나타남.

- 통일에 대한 반대 확률이 가장 높은 연관규칙은 {핵무기, 천안함}⇒{반대}이며 온라인 문서에서 ‘핵무기, 천안함’이 언급되면 언급되지 않는 문서보다 통일에 대한 반대 확률이 5.92배 높아지는 것으로 나타남.

**〈표 4〉 안보와 이슈 요인에 대한 통일인식의 연관성 예측**

규칙	지지도	신뢰도	향상도
{핵무기, 통일대박}⇒{중립}	0.002700533	0.26811594	8.2815445
{핵무기, 정상회담}⇒{중립}	0.002158341	0.19202226	5.9311688
{핵무기, 천안함}⇒{반대}	0.001053104	0.41908714	5.9212419
{핵무기, 통일비용}⇒{반대}	0.001459747	0.41297935	5.8349456
{통일대박}⇒{중립}	0.013971869	0.18644775	5.7589838
{휴전선}⇒{중립}	0.001334626	0.17273954	5.3355656
{천안함}⇒{반대}	0.001334626	0.37426901	5.2880108
{정상회담}⇒{중립}	0.002888215	0.16838906	5.2011882
{핵무기, 정상회담}⇒{반대}	0.003784917	0.33673469	4.7576921
{통일대박}⇒{찬성}	0.052488348	0.70043133	4.5726137
{정상회담}⇒{반대}	0.004879727	0.28449848	4.0196517
{핵무기, 휴전선}⇒{반대}	0.001219932	0.27725118	3.9172554
{통일비용}⇒{반대}	0.002721386	0.26605505	3.7590662
{핵무기, 통일대박}⇒{찬성}	0.005494907	0.54554865	3.5614958
{핵무기}⇒{중립}	0.010061831	0.11222235	3.4663153
{휴전선}⇒{반대}	0.001845538	0.23886640	3.3749204
{핵무기}⇒{반대}	0.019060131	0.21258286	3.0035628
{이산가족상봉}⇒{찬성}	0.001522308	0.44108761	2.8795446
{핵무기, 휴전선}⇒{찬성}	0.001939379	0.44075829	2.8773947
{핵무기, 정상회담}⇒{찬성}	0.004817167	0.42857143	2.7978354
{휴전선}⇒{찬성}	0.003242725	0.41970310	2.7399405
{정상회담}⇒{찬성}	0.006839960	0.39878419	2.6033759
{핵무기, 통일비용}⇒{찬성}	0.001397187	0.39528024	2.5805011
{핵무기, 통일대박}⇒{반대}	0.001574442	0.15631470	2.2085554
{통일비용}⇒{찬성}	0.003284432	0.32110092	2.0962375

### 3. 요약 및 결론

- 통일 관련 온라인 문서는 2016년 1월 6일 북한의 4차 핵실험 이후 꾸준히 증가하여 2월 10일 북한의 개성공단 잠정 폐쇄 발표 이후 급속히 증가함. 이는 2016년 상반기 북한의 핵실험과 개성공단 폐쇄로 인하여 남북 간 긴장이 고조되면서 국민들의 관심이 급증한 것으로 볼 수 있음.
- 통일에 대한 찬성 비율은 소셜 빅데이터 감성 분석 결과 2011년 55.0%, 2012년 62.0%, 2013년 57.7%, 2014년 58.2%, 2015년 56.2%, 2016년 59.8%로 서울대 통일평화연구원의 여론조사<sup>10)</sup> 결과와 비슷한 추이를 보임. 개성공단 중단 및 사드 배치 검토 전후 통일에 대한 인식 변화에서는 찬성 인식(3.95% 상승)이 반대 인식(0.9% 상승)보다 더 많이 상승한 것으로 나타남.
- 통일대박, 남북공동선언, 이산가족상봉, 휴전선 요인이 있으면 통일에 대한 찬성 확률이 높은 반면 천안함, 핵무기, 통일비용 요인이 있으면 통일에 대한 반대의 확률이 높은 것으로 나타남. 이는 통일대박, 남북 공동선언 등은 긍정적 담론을 확산시키는 반면 천안함, 핵무기 등은 부정적 담론을 확산시키는 것으로 볼 수 있음.
- 핵무기와 통일대박이 동시에 언급된 문서는 통일에 대한 중립과 찬성의 확률이 높은 것으로 나타났으며 핵무기, 천안함, 휴전선이 동시에 언급된 문서는 통일에 대한 반대의 확률이 높은 것으로 나타나 ‘통일 대박론’이 통일의 필요성에 대한 인식에 영향을 준 것으로 드러남.
- 최근 미디어 환경은 다양한 플랫폼과 이를 통해 제공되는 콘텐츠들로 확장됨에 따라<sup>11)</sup> 통일에 대한 공감대 형성에서 소셜미디어가 중요한 역할을 수행할 수 있는 것으로 나타남. 특히 소셜미디어는 통일과 관련한 폭넓은 정보를 쉽게 얻을 수 있는 곳으로 담론이 형성, 공유되는 통로 역할을 함.<sup>12)</sup>
- 따라서 통일에 대한 찬반, 통일 방법 등과 관련한 국민의 통일인식 조사와 더불어 소셜미디어에서 수집된 빅데이터를 활용, 분석하면 통일인식의 예측은 더욱 신뢰성이 있을 것으로 봄.<sup>13)</sup>

10) 서울대학교 통일평화연구원이 매년 7월부터 8월 사이(3주간) 1200명을 대상으로 대면 조사한 결과임.

11) 홍종윤(2016). 미디어의 평화통일 지향성 지수 개발 연구. 언론정보연구 53(1), pp.5-56.

12) 연세대학교 산학협력단(2014). 평화통일에 대한 국민 공감대 형성 연구.

13) 송태민·진달래(2015). 2015년 소셜 빅데이터 기반 보건복지 이슈 동향 분석. 한국보건사회연구원, p.123.