소셜빅데이터를 활용한 보건복지정책 동향 분석

Trend Analysis of Health and Welfare Policy Based on Social Big Data

송태민 한국보건사회연구원 연구위원

대상자별, 분야별로 다양한 보건복지 정책이 요구됨에 따라 정부는 보건복지 정책 수요를 사전에 예측하여 보건복지 욕구를 충족시키기 위한 노력이 필요하다. 국민이 요구하는 보건복지 정책 수요를 분석하기 위해서는 다양한 산업의 종사자나 일반인을 대상으로 설문조사를 실시해야 한다. 기존에 실시하던 횡단적 조사나 종단적 조사 등을 대상으로 한 연구는 정해진 변인들에 대한 개인과 집단의 관계를 보는 데에는 유용하나, 사이버 상에서 언급된 문서에서 논의된 관련 변인의 상호간의 연관관계를 밝히고 원인을 파악하는 데는 한계가 있다. 따라서 국민이 필요한 보건복지 욕구를 사전에 파악하여 보건복지 수요를 정확·신속하게 예측하고 계획을 수립하기 위해서는 오프라인 보건복지 수요 조사와 함께 소셜 빅데이터의 지속적인 활용과 분석을 통한 과학적 보건복지 정책수립이 필요할 것으로 본다.

1. 서론

새정부 출범이후 2013년부터 건강보험 보장성 강화를 위한 국민적 요구가 증대하고 고령화로 진료비 지출증가 등이 예상[1]됨에 따라 진료비 지출을 억제하고, 건강보험의 안정적 운영을 위한 정책개발이 요구되고 있다. 특히, 인구고령화가 사회적 문제로 대두됨에 따라 전체 인구의 고령화와 함께 장애인구 역시 고령화되는 현상을 보임에 따라 노인과 장애인을 위한 정책수요는 증가할 것으로 보인다[2]. 그리고 고령화로 인한 노인인구의 양적 증가와 특성의 다양화로 인한 보건복지 서비스에 대한 관심의 증가와 함께 향후 노인의 보건복지서비스에 대한 욕구도 폭발적으로 늘어날 것으로 예측하고 있다[3]. 그동안 정부의 보육정책은 저소득층 위주의 경제적 부담완화에 초점을 맞추어 왔으나 새정부 출범과 함께 전 계층 무상보육과 양육수당 지급 정책이 도입됨에 따라 대상자들의 정책 체감도는 상당히 높아질 것으로 예견되고 있다[4]. 또한

1) 신현웅(2013). 건강보험 진료비 분석 및 정책방향, **보건·복지 Issue & Focus**, 186.
2) 김성희(2013). 장애노인의 실태와 과제, **보건·복지 Issue & Focus**, 208.
3) 이윤경, 염주희, 이선희(2013). **고령화 대응 노인복지서비스 수요전망과 공급체계 개편**, 한국보건사회연구원.
4) 김은정(2013). 소득계층별 출산행태 분석과 시사점, **보건·복지 Issue & Focus**, 191.

2014년 7월부터 상대적으로 형편이 어려운 노인에게 차등 지급하는 기초연금에 관한 정책이 결정되어 많은 국민이 공적연금의 혜택을 받을 수 있게 되었다. 그러나, 영리병원을 비롯한 의료민영화 정책과 원격의료 정책은 아직 정부와 의료계의 합의가 이루어지지 못하고 있다. 이와 같이 대상자별, 분야별로 다양한 보건복지 정책이 요구됨에 따라 정부는 보건복지 정책 동향을 파악하여 보건복지 욕구를 충족시키기 위한 노력이 필요하다.

한편, 스마트기기의 급속한 보급과 소셜미디어의 확산으로 데이터량이 기하급수적으로 증가하고 데이터의 생산, 유통, 소비 체계에 큰 변화를 주면서 데이터가 경제적 자산이 될 수 있는 빅데이터[5] 시대를 맞이하게 되었다[6]. 세계 각국의 정부와 기업들은 빅데이터가 향후 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하고 있으며, McKinsey, The Economist, Gartner 등은 빅데이터를 활용한 시장변동 예측과 신사업 발굴 등 경제적 가치창출 사례 및 효과를 제시하고 있다[7]. 특히, 많은 국가와 기업에서는 SNS를 통하여 생산되는 소셜 빅데이터의 활용과 분석을 통하여 새로운 경제적 효과와 일자리 창출은 물론 사회적 문제의 해결을 위하여 적극적으로 노력하고 있다. 한국은 최근 정부3.0[8]과 창조경제의 추진과 실현을 위하여 모든 분야에 빅데이터의 효율적 활용을 적극적으로 모색하고 있다. 이와 같이 많은 국가에서 빅데이터가 공공과 민간에 미치는 파급효과를 전망함에 따라 빅데이터의 활용은 정부의 정책을 효율적으로 추진하기 위한 새로운 동력이 될 것으로 보여진다. 국민이 요구하는 보건복지 정책 수요를 분석하기 위해서는 다양한 산업의 종사자나 일반인을 대상으로 설문조사를 실시해야 한다. 그러나 기존에 실시하던 횡단적 조사나 종단적 조사 등을 대상으로 한 연구는 정해진 변인들에 대한 개인과 집단의 관계를 보는 데에는 유용하나, 사이버 상에서 언급된 개인별 담론(buzz)에서 논의된 관련 변인의 상호간의 연관관계를 밝히고 원인을 파악하는 데는 한계가 있다. 특히, 트위트, 페이스북 등 소셜미디어에 남긴 정치, 경제, 사회, 문화에 대한 메시지는 그 시대의 감성과 정서를 파악할 수 있는 원천으로 등장함에 따라, 대중매체에 의해 수립된 정책의제는 이제 소셜미디어로부터 파악할 수 있게 되었다[9]. 따라서 본 연구는 2014년 1월과 2월 우리나라 온라인 뉴스사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 바탕으로 보건복지 정책의 동향을 분석하고자 한다.

5) 빅데이터란 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미함(위키백과, 2014. 5. 24).

6) 송태민(2012). 보건복지 빅데이터 효율적 활용방안, **보건복지포럼, 193,** pp.68~76.

7) 송태민(2013). 우리나라 보건복지 빅데이터 동향 및 활용방안, **과학기술정책, 192,** pp.56~73.

8) 정부 3.0이란 공공정보를 적극 개방 · 공유하고, 부처 간 칸막이를 없애 소통 · 협력함으로써 국민 맞춤형 서비스를 제공하고, 일자리 창출과 창조 경제를 지원하는 새로운 정부운영 패러다임을 의미함.

9) 송영조(2012). **빅데이터 시대! SNS의 진화와 공공정책,** 한국정보화진흥원.

2. 연구방법

1) 연구대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅 데이터를 대상으로 하였다. 본 분석에서는 116개의 온라인 뉴스사이트, 4개의 블로그(네이버, 네이트, 다음, 티스토리), 2개의 카페(네이버, 다음), 2개의 SNS(트위터, 미투데이), 4개의 게시판(네이버지식인, 네이트지식, 네이트톡, 네이트판) 등 총 128개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서(버즈)를 소셜 빅데이터로 정의하였다. 보건복지 관련 토픽의 수집은 2014. 1. 1~2014. 2. 28까지 해당 채널에서 요일별, 주말, 휴일을 고려하지 않고 매 시간단위로 수집하였으며, 수집된 총 111,596건(1월: 57,830건, 2월: 53,766건)[10]의 텍스트(Text) 문서를 본 연구의 분석에 포함시켰다. 본 연구를 위한 소셜 빅데이터의 수집[11]은 크롤러(Crawler)를 사용하였고, 토픽의 분류는 주제분석(text mining) 기법을 사용하였다. 보건복지 토픽은 모든 관련 문서를 수집하기 위해 '보건', '복지', 그리고 '보건복지'를 사용하였다.

2) 연구도구

보건복지와 관련하여 수집된 버즈는 주제분석(text mining)과 요인분석(factor analysis)의 과정을 거쳐 다음과 같이 정형화 데이터로 코드화하여 사용하였다.

(1) 보건복지 관련 수요

본 연구의 종속변수인 보건복지 수요(찬성, 반대)의 정의는 주제분석과 요인분석의 과정을 거쳐 '운영, 지원, 계획, 예정, 강화, 실시, 확대, 진행, 이용, 사용, 도입, 추진, 참여'은 찬성의 감정으로, '문제, 지적, 반대, 거짓말, 논란, 비판, 걱정, 억울, 외면'은 반대의 감정으로 정의하였다.

(2) 보건복지 관련 정책

보건복지 관련 정책의 정의는 주제분석과 요인분석의 과정을 거쳐 '연금(기초연금, 국민연금), 기초생활(최저생계비, 기초생활보장, 기초생활수급권), 양육수당(양육수당, 보조금, 맞춤형), 무상정책(무상보육, 무상의료, 무상급식), 의료민영화(포괄수가제, 의료상업화, 민영화, 의료민영화), 건강보험(국민건강보험, 민간보험, 건강보험), 원격의료(법인약국, 의료영리화, 원격진료, 원격의료), 전세대책(전세대책, 서민정책), 반값등록금(반값등록금, 4대강), 퇴직연금(희망누리, 퇴직연금, 산재보험), 증세(증세, 세법개정안, 부자증세, 부자감세), 과다복지(방만경영, 과다복지, 복리후생비, 복지카드)[12], 행복온도, 육아휴직(출산휴가, 보건휴가, 육아휴

10) 본 연구에서 수집된 문서는 SNS 77.7%(트위트: 75.9%(84,721건), 미투데이: 1.8%(2,010건)), 온라인 뉴스사이트 7.4%(8,345건), 카페 6.8%(7,508건), 블로그 5.9%(6,704건), 게시판 2.2%(2,317건) 순으로 나타남.

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

직), 중독법(중독법, 게임중독법)'의 15개 정책으로 해당 정책이 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

(3) 보건복지 관련 공공기관

보건복지 관련 공공기관의 정의는 주제분석과 요인분석의 과정을 거쳐 '국회(보건복지위원회, 국회), 보건복지부(보건복지부, 질병관리본부), 청와대, 고용여성부처(고용노동부, 여성가족부), 공기업(공기업, 공공기관, 근로복지공단, 근로복지넷), 교육농림부처(교육부, 농림축산식품부), 지자체(보건소, 주민센터)'의 7개 기관으로 정의하였다. 정의된 모든 공공기관은 해당 공공기관이 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

(4) 보건복지 관련 민간기관

보건복지 관련 민간기관의 정의는 주제분석과 요인분석의 과정을 거쳐 '요양병원(인증요양병원, 요양병원, 의료기관), 관련협회(약사회, 의사협회, 보건의료노조, 보건의료단체, 대한의사협회), 기업(대기업, 중소기업, 회사, 기업, 삼성, 복지시설), 대학(대학, 학교), 시민단체(시민단체, 병원), 사회복지시설(장애인복지시설, 아동복지시설, 사회복지시설, 장애인복지단체)'의 6개 기관으로 정의하였다. 정의된 모든 민간기관은 해당 민간기관이 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

(5) 보건복지 관련 대상

보건복지 관련 대상의 정의는 주제분석과 요인분석의 과정을 거쳐 '가족(모녀, 부자, 가족, 할머니, 자식들, 아이), 청소년(아동, 청소년), 환자, 의사, 저소득층(저소득층, 취약계층, 기초생활수급자, 소외계층, 빈곤층), 장애인(장애인, 노약자), 여성, 노동자(노동자, 노동자들, 근로자), 서민(서민, 서민들), 중산층, 노인(노인들, 노인, 어르신들, 어르신), 피해자, 공무원(공무원, 군인), 비정규직, 학생(대학생, 학생, 청년), 노숙자(노숙자, 노숙자들), 외국인'의 17개 대상으로 정의하였다. 정의된 모든 대상은 해당 대상이 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

(6) 보건복지 관련 분야

보건복지 관련 분야의 정의는 주제분석의 과정을 거쳐 '주거(주거복지, 주거안정), 교육, 사회복지, 보건의료, 경제, 문화, 환경, 통일, 가정, 노동, 보육, 범죄, 안보, 다문화'의 14개로 정의하였다. 정의된 모든 분야는 해당 분야가 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

12) '과다복지'는 '비정상의 정상화'와 관련하여 공기업 · 공공기관의 '방만경영'의 실태를 나타내는 개념으로 보건복지 정책과는 무관하여 본 분석에서는 제외함.

(7) 보건복지 관련 주요이슈

보건복지 관련 주요이슈의 정의는 주제분석과 요인분석의 과정을 거쳐 '중증질환, 의료비, 담배, 건강, 일자리(일자리, 취업), 의료수가, 자살, 등록금(등록금, 학자금), 세금, 행복지수(행복지수, 국민행복지수), 개인정보, 부동산, 정규직, 위안부, 결혼, 출산율, 양극화, 성폭행, 반려동물'의 19개 이슈로 정의하였다. 정의된 모든 이슈는 해당 대상이 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

3) 분석 방법

본 연구에서는 한국의 보건복지 정책 수요를 설명하는 가장 효율적인 예측모형을 구축하기 위해 특별한 통계적 가정이 필요하지 않은 데이터마이닝의 의사결정나무 분석방법을 사용하였다. 데이터마이닝의 의사결정나무 분석은 방대한 자료 속에서 종속변인을 가장 잘 설명하는 예측모형을 자동적으로 산출해 줌으로써 각기 다른 속성은 가진 보건복지 정책 수요에 대한 요인을 쉽게 파악할 수 있다. 본 연구의 의사결정나무 형성을 위한 분석 알고리즘은 CHAID (CHi-squared Automatic Interaction Detection)[13]를 사용하였다. CHAID는 이산형인 종속변수의 분리기준으로 카이제곱(χ^2)-검정을 사용하며, 모든 가능한 조합을 탐색하여 최적분리를 찾는다. 정지규칙(stopping rule)으로 관찰치가 충분하여 상위노드(부모마디)의 최소케이스 수는 100으로 하위노드(자식마디)의 최소 케이스 수는 50으로 설정하였고, 나무깊이는 3수준으로 정하였다. 본 연구의 기술분석, 다중응답분석, 로지스틱 회귀분석, 의사결정나무분석은 SPSS v. 21.0을 사용하였다.

3. 연구 결과

1) 보건복지 관련 분야 버즈 현황

보건복지 관련 수요는 찬성의 감정을 가진 버즈가 62.6%로 나타났다. 보건복지 관련 주요정책으로는 의료민영화의 버즈가 40.3%로 가장 높게 나타났으며, 증세(8.6%), 양육수당(7.8%), 원격의료(7.5%), 연금(6.8%), 건강보험(6.3%)의 순으로 나타났다. 보건복지 관련 주요이슈로는 일자리의 버즈가 20.8%로 가장 높게 나타났으며, 건강(16.4%), 세금(15.7), 결혼(6.4%), 자살(5.9), 의료비(5.8%), 출산율(4.6%) 등의 순으로 나타났다. 보건복지 관련 분야의 버즈로는 교육(19.2%), 사회복지(18.1%), 보건의료(15.7%), 경제(13.8%), 문화(7.0%), 환경(6.2%), 가정(4.9%) 등의 순으로 나타났다. 보건복지 관련 대상의 버즈로는 노인(15.5%), 가족(15.2%), 공무원(9.4%), 저소득층(8.9%), 장애인(7.7%), 노동자(6.5%), 학생(5.6%) 등의 순으로 나타났다. 보건복지 관련 공공기관의 버즈로는 청와대(34.4%),

13) Kass, G.(1980), An exploratory technique for investigating large quantities of categorical data, *Applied Statistics, 292*, pp.119~127.

보건복지부(32.1%), 공기업(17.5%), 국회(7.0%), 지자체(6.0%) 등의 순으로 나타났다. 보건복지 관련 민간기관의 버즈로는 기업(32.8%), 대학(16.6%), 사회복지시설(16.5%), 관련협회(15.6%) 등의 순으로 나타났다.

<표 2>과 같이 청와대와 관련한 정책 버즈는

표 1. 보건복지 관련 버즈 현황

구분	항목	N(%)	구분	항목	N(%)	구분	항목	N(%)
이슈	중증질환	266(1.9)	정책 분야	연금	1,419(6.8)	수요	반대	9,912(37.4)
	의료비	823(5.8)		기초생활	693(3.3)		찬성	16,594(62.6)
	담배	283(2.0)		양육수당	1,644(7.8)		계	26,506
	건강	2,345(16.4)		무상정책	968(4.6)	민간기관	요양병원	902(5.8)
	일자리	2,970(20.8)		의료민영화	8,456(40.3)		관련협회	2,430(15.6)
	의료수가	154(1.1)		건강보험	1,322(6.3)		기업	5,104(32.8)
	자살	844(5.9)		원격의료	1,573(7.5)		대학	2,592(16.6)
	등록금	502(3.5)		전세대책	117(0.6)		시민단체	1,982(12.7)
	세금	2,245(15.7)		반값등록금	705(3.4)		사회복지시설	2,563(16.5)
	행복지수	499(3.5)		퇴직연금	855(4.1)		계	15,573
	개인정보	307(2.1)		증세	1,856(8.8)	대상	가족	4,300(15.2)
	부동산	356(2.5)		행복온도	733(3.5)		청소년	1,516(5.4)
	정규직	262(1.8)		육아휴직	366(1.7)		환자	1,310(4.6)
	위안부	239(1.7)		중독법	267(1.3)		의사	1,063(3.8)
	결혼	909(6.4)		계	20,974		저소득층	2,511(8.9)
	출산율	662(4.6)		주거	429(2.0)		장애인	2,166(7.7)
	양극화	205(1.4)		교육	4,019(19.2)		여성	1,514(5.4)
	성폭행	240(1.7)		사회복지	3,790(18.1)		노동자	1,838(6.5)
	반려동물	182(1.3)		보건의료	3,299(15.7)		서민	1,473(5.2)
	계	14,293		경제	2,884(13.8)		중산층	278(1.0)
공공기관	국회	2,183(7.0)		문화	1,458(7.0)		노인	4,400(15.5)
	보건복지부	10,041(32.1)		환경	1,297(6.2)		피해자	250(0.9)
	청와대	10,778(34.4)		통일	870(4.1)		공무원	2,652(9.4)
	고용여성부처	488(1.6)		가정	1,019(4.9)		비정규직	417(1.5)
	공기업	5,469(17.5)		노동	802(3.8)		학생	1,591(5.6)
	교육농림부처	468(1.5)		보육	427(2.0)		외국인	501(1.8)
	지자체	1,890(6.0)		범죄	241(1.1)		노숙자요인	516(1.8)
	계	31,317		안보	233(1.1)		계	28,296
				다문화	197(0.9)			
				계	20,965			

의료민영화, 증세, 연금, 원격의료, 양육수당, 등의 순으로 나타났다. 국회와 관련한 정책 버즈는 의료민영화, 연금, 원격의료, 중독법, 양육수당, 건강보험 등의 순으로 나타났다. 보건복지부와 관련한 정책 버즈는 의료민영화, 건강보험, 행복온도, 연금, 원격의료, 양육수당 등의 순

표 2. 공공기관별 보건복지 정책 및 이슈 버즈 현황

N(%)

분야	속성	국회	보건복지부	청와대	고용여성부처	공기업	교육농림부처	지자체	합계
정책	연금	245(16.6)	615(10.0)	479(10.2)	25(11.8)	89(4.5)	14(18.9)	57(10.8)	1,524
	기초생활	61(4.1)	115(1.9)	57(1.2)	17(8.0)	45(2.3)	10(13.5)	75(14.3)	380
	양육수당	122(8.2)	268(4.3)	362(7.7)	88(41.5)	214(10.7)	17(23.0)	190(36.1)	1,261
	무상정책	59(4.0)	48(.8)	79(1.7)	4(1.9)	92(4.6)	9(12.2)	6(1.1)	297
	의료민영화	358(24.2)	2,763(44.8)	2,484(52.7)	6(2.8)	495(24.8)	10(13.5)	58(11.0)	6,174
	건강보험	112(7.6)	893(14.5)	127(2.7)	31(14.6)	74(3.7)	7(9.5)	76(14.4)	1,320
	원격의료	218(14.7)	430(7.0)	369(7.8)	6(2.8)	19(1.0)	4(5.4)	46(8.7)	1,092
	전세대책	2(.1)	3(.0)	4(.1)	1(.5)	2(.1)	–	–	12
	반값등록금	29(2.0)	145(2.4)	65(1.4)	2(.9)	32(1.6)	2(2.7)	5(1.0)	280
	퇴직연금	3(.2)	26(.4)	9(.2)	13(6.1)	708(35.5)	–	4(.8)	763
	증세	92(6.2)	29(.5)	624(13.2)	2(.9)	184(9.2)	1(1.4)	2(.4)	934
	행복온도	–	732(11.9)	–	–	–	–	–	732
	육아휴직	13(.9)	21(.3)	54(1.1)	15(7.1)	38(1.9)	–	7(1.3)	148
	중독법	166(11.2)	82(1.3)	2(.0)	2(.9)	–	–	–	252
이슈	계	1480	6170	4715	212	1992	74	526	15,169
	중증질환	63(10.0)	151(9.4)	72(3.5)	22(7.3)	50(5.6)	–	51(8.0)	409
	의료비	76(12.0)	142(8.8)	326(16.1)	12(4.0)	131(14.7)	2(1.8)	83(13.0)	772
	담배	43(6.8)	96(6.0)	25(1.2)	1(.3)	27(3.0)	–	30(4.7)	222
	건강	103(16.3)	376(23.4)	121(6.0)	54(18.0)	83(9.3)	29(26.4)	236(36.9)	1,002
	일자리	123(19.4)	339(21.1)	423(20.8)	154(51.3)	127(14.2)	37(33.6)	117(18.3)	1,320
	의료수가	16(2.5)	34(2.1)	16(.8)	–	3(.3)	2(1.8)	1(.2)	72
	자살	14(2.2)	32(2.0)	70(3.4)	5(1.7)	7(.8)	2(1.8)	9(1.4)	139
	등록금	17(2.7)	28(1.7)	76(3.7)	4(1.3)	154(17.3)	14(12.7)	14(2.2)	307
	세금	52(8.2)	126(7.9)	548(27.0)	7(2.3)	192(21.5)	5(4.5)	16(2.5)	946
	행복지수	1(.2)	2(.1)	4(.2)	–	1(.1)	–	–	8
	개인정보	31(4.9)	57(3.6)	99(4.9)	6(2.0)	27(3.0)	4(3.6)	20(3.1)	244
	부동산	14(2.2)	28(1.7)	41(2.0)	4(1.3)	17(1.9)	1(.9)	7(1.1)	112
	정규직	15(2.4)	18(1.1)	33(1.6)	4(1.3)	21(2.4)	2(1.8)	7(1.1)	100
	위안부	8(1.3)	20(1.2)	9(.4)	5(1.7)	4(.4)	2(1.8)	8(1.3)	56
	결혼	24(3.8)	76(4.7)	70(3.4)	13(4.3)	27(3.0)	4(3.6)	18(2.8)	232
	출산율	5(.8)	21(1.3)	16(.8)	3(1.0)	3(.3)	2(1.8)	10(1.6)	60
	양극화	17(2.7)	16(1.0)	61(3.0)	2(.7)	12(1.3)	3(2.7)	–	111
	성폭행	11(1.7)	43(2.7)	20(1.0)	3(1.0)	5(.6)	1(.9)	10(1.6)	93
	반려동물	–	–	–	1(.3)	1(.1)	–	2(.3)	4
	계	633	1,605	2,030	300	892	110	639	6,209

으로 나타났다. 청와대와 관련한 주요이슈 버즈는 세금, 일자리, 의료비, 건강, 개인정보, 등록금 등의 순으로 나타났다. 국회와 관련한 주요이슈 버즈는 일자리, 건강, 의료비, 중증질환, 세금, 담배 등의 순으로 나타났다. 보건복지부와 관련한 주요이슈 버즈는 건강, 일자리, 중증질환, 의료비, 세금, 담배 등의 순으로 나타났다.

<표 3>과 같이 기업과 관련한 정책 버즈는 양육수당, 의료민영화, 연금, 건강보험, 무상정책, 원격의료 등의 순으로 나타났다. 시민단체와 관련한 정책 버즈는 의료민영화, 원격의료, 건강보험, 연금, 양육수당, 기초생활 등의 순으로 나타났다. 기업과 관련한 주요이슈 버즈는 일자리, 건강, 세금, 결혼, 의료비, 등록금 등의 순으로 나타났다. 시민단체와 관련한 주요이슈 버즈는 건강, 일자리, 의료비, 의료수가, 중증질환, 등록금 등의 순으로 나타났다.

표 3. 민간기관별 보건복지 정책 및 이슈 버즈 현황

N(%)

분야	속성	요양병원	관련협회	기업	대학	시민단체	사회복지시설	합계
정책	연금	57(10.4)	18(1.1)	136(12.7)	81(16.4)	97(8.4)	20(9.8)	409
	기초생활	44(8.1)	–	51(4.8)	47(9.5)	44(3.8)	19(9.3)	205
	양육수당	87(15.9)	14(.9)	213(19.9)	103(20.9)	90(7.8)	118(57.6)	625
	무상정책	8(1.5)	10(.6)	90(8.4)	39(7.9)	31(2.7)	13(6.3)	191
	의료민영화	76(13.9)	1,071(65.0)	183(17.1)	29(5.9)	325(28.3)	8(3.9)	1,692
	건강보험	131(24.0)	93(5.6)	102(9.6)	76(15.4)	199(17.3)	13(6.3)	614
	원격의료	111(20.3)	436(26.5)	83(7.8)	11(2.2)	284(24.7)	–	925
	전세대책	–	–	4(.4)	1(.2)	3(.3)	–	8
	반값등록금	15(2.7)	1(.1)	38(3.6)	31(6.3)	17(1.5)	–	102
	퇴직연금	9(1.6)	–	50(4.7)	14(2.8)	15(1.3)	7(3.4)	95
	증세	2(.4)	4(.2)	70(6.6)	41(8.3)	22(1.9)	7(3.4)	146
	육아휴직	6(1.1)	–	46(4.3)	21(4.3)	11(1.0)	–	84
	중독법	–	–	2(.2)	–	12(1.0)	–	14
	계	546	1647	1068	494	1150	205	5,110
이슈	중증질환	63(10.8)	17(6.4)	36(2.1)	43(3.6)	53(4.6)	8(2.9)	220
	의료비	104(17.9)	64(24.2)	111(6.6)	71(6.0)	205(17.8)	18(6.6)	573
	담배	27(4.6)	11(4.2)	38(2.3)	42(3.5)	27(2.3)	2(.7)	147
	건강	158(27.1)	60(22.7)	265(15.8)	198(16.7)	267(23.2)	62(22.8)	1,010
	일자리	101(17.4)	35(13.3)	417(24.9)	376(31.7)	261(22.7)	80(29.4)	1,270
	의료수가	19(3.3)	28(10.6)	15(.9)	3(.3)	82(7.1)	–	147
	자살	4(.7)	6(2.3)	18(1.1)	26(2.2)	25(2.2)	9(3.3)	88
	등록금	16(2.7)	–	90(5.4)	167(14.1)	44(3.8)	6(2.2)	323
	세금	12(2.1)	10(3.8)	233(13.9)	61(5.1)	43(3.7)	18(6.6)	377

〈표 3〉 계속

분야	속성	요양병원	관련협회	기업	대학	시민단체	사회복지시설	합계
이슈	행복지수	–	–	5(.3)	4(.3)	2(.2)	1(.4)	12
	개인정보	24(4.1)	7(2.7)	26(1.6)	17(1.4)	25(2.2)	16(5.9)	115
	부동산	1(.2)	–	61(3.6)	29(2.4)	13(1.1)	5(1.8)	109
	정규직	10(1.7)	2(.8)	53(3.2)	24(2.0)	20(1.7)	2(.7)	111
	위안부	5(.9)	6(2.3)	32(1.9)	23(1.9)	7(.6)	2(.7)	75
	결혼	13(2.2)	5(1.9)	162(9.7)	59(5.0)	33(2.9)	8(2.9)	280
	출산율	9(1.5)	–	21(1.3)	8(.7)	13(1.1)	2(.7)	53
	양극화	14(2.4)	9(3.4)	36(2.1)	21(1.8)	19(1.6)	3(1.1)	102
	성폭행	2(.3)	4(1.5)	51(3.0)	13(1.1)	11(1.0)	29(10.7)	110
	반려동물	–	–	5(.3)	2(.2)	2(.2)	1(.4)	10
	계	582	264	1,675	1,187	1,152	272	5,132

2) 보건복지 정책 수요에 미치는 영향 요인

보건복지 정책에서 전세대책, 기초생활, 양육수당, 건강보험, 퇴직연금, 중독법, 원격의료 요인은 정적(+)인 영향을 미치는 것으로 나타나, 동 요인들에 대한 찬성의 확률이 높은 것으로 나타났다. 반값등록금, 증세, 무상정책, 의료민영화, 연금 요인은 부적(–)인 영향을 미치는 것으로 나타나, 동 요인들에 대한 반대의 확률이 높은 것으로 나타났다(표 4).

3) 보건복지 정책 수요 예측모형

본 연구에서는 주요정책에 대한 수요를 예측하기 위하여 연금, 무상정책, 의료민영화, 원격의료, 증세 요인에 대한 데이터마이닝 분석을 실시하였다. 보건복지 정책 수요 예측모형에 미치는 영향은 [그림 1]과 같다. 나무구조의 최상위에 있는 네모는 루트노드로서, 예측변수(독립변수)가 투입되지 않은 종속변수(찬성, 반대)의 빈도를 나타낸다. 루트노드에서 보건복지 전체 정책의 찬성은 62.6%(16,594건), 반대는 37.4%(9,912건)으로 나타났다. 루트노드의 하단의 가장 상위에 위치하는 요인이 보건복지 정책 수요 예측에 가장 영향력이 높은(관련성이 깊은)요인으로 '의료민영화요인'의 영향력이 가장 큰 것으로 나타났다. '의료민영화요인'이 높을 경우 보건복지 정책 찬성이 이전의 62.6%에서 48.0%로 감소한 반면, 보건복지 정책의 반대는 이전의 37.4%에서 52.0%로 증가하였다. '의료민영화요인'이 높고 '원격의료요인'이 높은 경우 보건복지 정책의 찬성이 이전의 48.0%에서 69.4%로 증가한 반면, 보건복지 정책의 반대는 이전의 52.0%에서 30.6%로 감소하였다. <표 5>의 보건복지 정책 수요 예측에 대한 이익도표와 같이 보건복지 정책의 찬성에 가장 영향력이 높은 경우는 '의료민영화요인'이 높고 '원격의료요인'이 높은 조합으로 나타났다. 즉, 6

표 4. 보건복지 정책 수요에 미치는 영향요인*

변수	b[a]	S.E.[b]	OR(95%CI)[c]	P
연금	−.725	.097	.484(.400−.586)	.000
기초생활	2.696	.312	14.813(8.031−27.324)	.000
양육수당	2.637	.163	13.969(10.157−19.213)	.000
무상정책	−.860	.123	.423(.333−.538)	.000
의료민영화	−.858	.038	.424(.393−.457)	.000
건강보험	2.285	.144	9.828(7.405−13.046)	.000
원격의료	.330	.082	1.391(1.185−1.634)	.000
전세대책	3.309	.460	27.356(11.109−67.363)	.000
반값등록금	−1.957	.170	.141(.101−.197)	.000
퇴직연금	2.270	.464	9.680(3.901−24.017)	.000
증세	−1.488	.094	.226(.188−.271)	.000
행복온도	.804	1.118	2.235(.250−20.000)	.472
육아휴직	.155	.262	1.167(.699−1.951)	.555
중독법	.939	.254	2.556(1.554−4.207)	.000
상수	.582	.015	1.790	.000

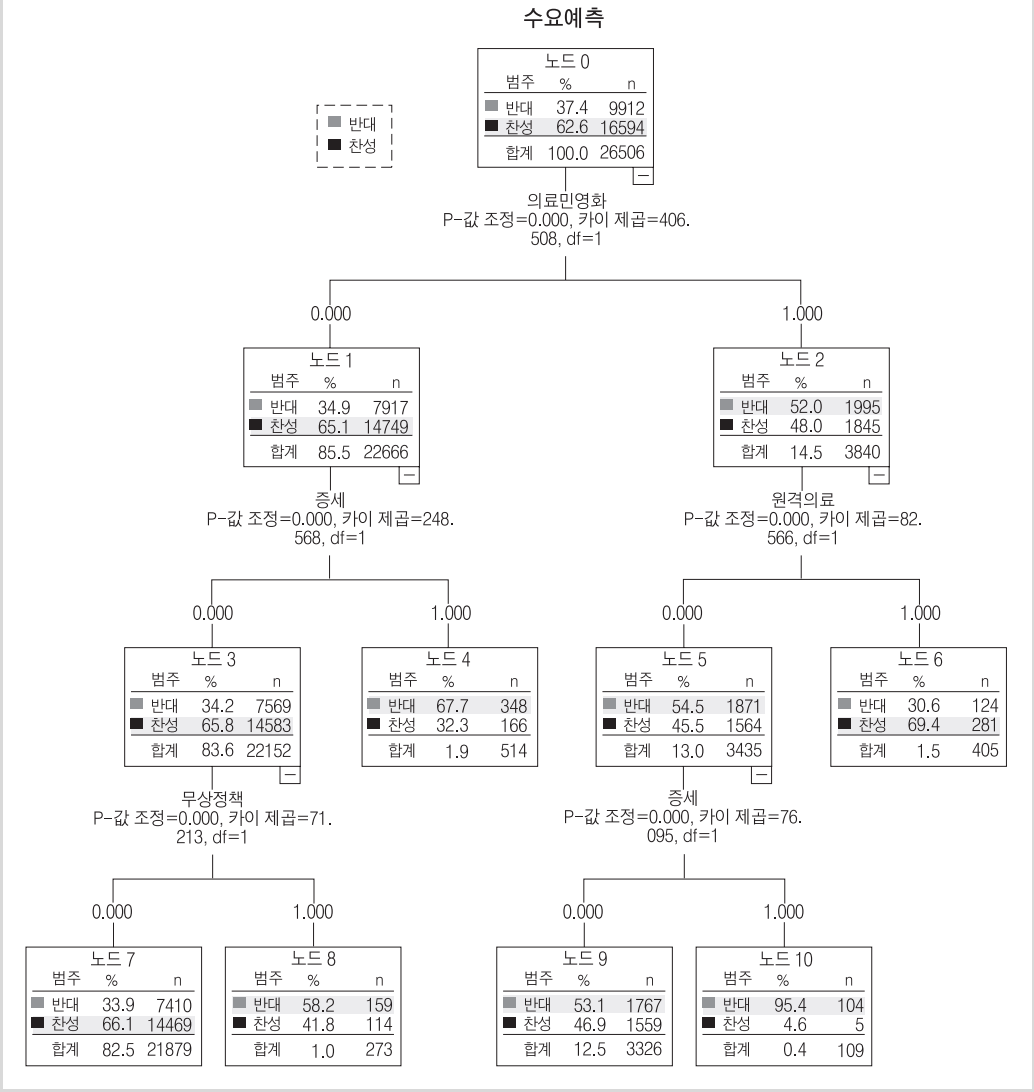
주: *기준범주: 반대, a Standardized coefficients, b Standard error, c Adjusted odds ratio(95% Confidence interval)

표 5. 보건복지 정책 수요 예측모형에 대한 이익도표

구분	노드	이익지수				누적지수			
		노드(n)	노드(%)	이익(%)	지수(%)	노드(n)	노드(%)	이익(%)	지수(%)
찬성	6	405	1.5	1.7	110.8	405	1.5	1.7	110.8
	7	21879	82.5	87.2	105.6	22284	84.1	88.9	105.7
	9	3326	12.5	9.4	74.9	25610	96.6	98.3	101.7
	8	273	1.0	.7	66.7	25883	97.6	99.0	101.4
	4	514	1.9	1.0	51.6	26397	99.6	100.0	100.4
	10	109	.4	.0	7.3	26506	100.0	100.0	100.0
반대	10	109	.4	1.0	255.1	109	.4	1.0	255.1
	4	514	1.9	3.5	181.1	623	2.4	4.6	194.0
	8	273	1.0	1.6	155.7	896	3.4	6.2	182.4
	9	3326	12.5	17.8	142.1	4222	15.9	24.0	150.6
	7	21879	82.5	74.8	90.6	26101	98.5	98.7	100.3
	6	405	1.5	1.3	81.9	26506	100.0	100.0	100.0

번 노드의 지수(index)가 110.8%로 뿌리마디와 비교했을 때 6번 노드의 조건을 가진 집단이 보건복지 정책을 찬성할 확률이 약 1.11배로 나타났다. 보건복지 정책의 반대에 가장 영향력이 높은 경우는 '의료민영화요인'이 높고 '원격의료요인'이 낮고, '증세요인'이 높은 조합으로 나타났다. 즉, 10번 노드의 지수(index)가 255.1%로 뿌리마디와 비교했을 때 10번 노드의 조건을 가진 집단이 보건복지 정책을 반대할 확률이 약 2.55배로 나타났다.

그림 1. 보건복지 정책 수요 예측모형



4. 결론

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 데이터마이닝의 의사결정나무 분석기법을 적용하여 분석함으로써 한국의 보건복지 정책 수요에 대한 예측모형을 개발하고자 하였다. 본 연구의 결과를 요약하면 다음과 같다. 첫째, 보건복지 정책과 관련한 버즈는 의료민영화, 과다복지, 증세, 양육수당, 원격의료, 연금, 건강보험 등의 순으로 많이 언급되는 것으로 나타났다. 둘째, 청와대와 관련한 보건복지 정책 관련 버즈는 의료민영화, 증세, 연금, 양육수당, 과다복지 등의 순이며, 보건복지부는 의료민영화, 건강보험, 행복온도, 연금, 원격의료, 양육수당 등의 순으로 나타났다. 그리고 시민단체와 관련한 정책 버즈는 의료민영화, 원격의료, 건강보험, 연금, 양육수당, 기초생활 등의 순으로 나타났다. 셋째, 보건복지 관련 정책은 전세대책, 양육수당, 기초생활, 건강보험, 퇴직연금, 중독법, 원격의료 정책에 대한 찬성의 확률이 높은 것으로 나타났다. 그리고 과다복지, 증세, 반값등록금, 무상정책, 의료민영화, 연금 정책은 반대의 확률이 높은 것으로 나타났다. 넷째, 보건복지 정책 수요 예측에 가장 영향력이 높은 정책은 '의료민영화요인'으로 나타났다. 따라서 보건복지 정책의 찬성에 가장 영향력이 높은 경우는 '의료민영화요인'이 높고 '원격의료요인'이 높은 조합으로 나타났다. 반면, '의료민영화요인'이 높고 '원격의료요인'이 낮고, '증세요인'이 높으면 보건복지 정책의 반대할 확률이 높은 것으로 나타났다. 본 연구를 근거로 한국의 보건복지 정책 수요에 대한 예측과 관련하여 다음과 같은 정책적 함의를 도출할 수 있다.

첫째, 의료민영화(포괄수가제, 의료상업화, 민영화, 의료민영화), 연금(기초연금, 국민연금), 원격의료(법인약국, 의료영리화, 원격의료, 원격의료), 건강보험(국민건강보험, 민간보험, 건강보험)에 대해서 보건복지 정책의 관심이 많은 것으로 나타났다. 이는 최근 정부의 원격의료 추진과 의료법인의 영리 자법인 허용 계획, 그리고 기초연금의 논의와 무관하지 않은 것으로 보여진다. 둘째, 원격의료 정책은 찬성의 확률이 높은 것으로 나타났지만 의료민영화 요인은 반대의 확률이 높은 것으로 나타났다. 따라서 정부는 원격의료 정책이 의료민영화의 시작이라는 인식을 불식시키기 위한 의료계와의 충분한 논의와 합의가 있어야 할 것으로 본다. 셋째, '의료민영화요인'이 높고 '원격의료요인'이 낮고, '증세요인'이 높은 집단의 보건복지 정책에 대한 반대가 높은 것으로 나타났다. 따라서 원격의료와 관련한 보건복지 정책의 실행을 위해서는 의료민영화 논의와는 분리하여 원격의료 시범사업을 실시해야 할 것으로 본다.

본 연구는 개개인의 특성을 가지고 분석한 것이 아니고 그 구성원이 속한 전체 집단의 자료를 대상으로 분석하였기 때문에 이를 개인에게 적용하였을 경우 생태학적 오류(ecological fallacy)가 발생할 수 있다[14]. 또한, 본 연구에서

14) Song TM, Song J, An JY, Hayman LL, Woo JM(2014), Psychological and Social Factors Affecting Internet

정의된 보건복지 관련 요인(용어)은 버즈 내에서 발생된 단어의 빈도로 정의되었기 때문에 기존의 조사 등을 통한 이론적 모형에서의 의미와 다를 수 있으며, 2개월간의 제한된 소셜 빅데이터를 분석함으로써 보건복지 정책의 동향 분석에 한계가 있을 수 있다. 본 연구는 소셜 빅데이터에서 수집된 빅데이터 주제분석과 데이터마이닝 분석을 통하여 한국의 보건복지 정책에 대한 수요 예측모형을 제시한 점에서 정책적 · 분석방법론적으로 의의가 있다. 또한, 실제적인 내용을 빠르게 효과적으로 파악하여 사회통계가 지닌 한계를 보완할 수 있는 새로운 조사방법으로서의 빅데이터의 가치를 확인하였다는 점에서 조사방법론적 의의를 가진다고 할 수 있다. 끝으로 국민이 필요한 보건복지 욕구를 사전에 파악하여 보건복지 수요를 정확 · 신속하게 예측하고 계획을 수립하기 위해서는 오프라인 보건복지 수요 조사와 함께 소셜 빅데이터의 지속적인 활용과 분석을 통한 과학적 보건복지 정책수립이 필요할 것으로 본다. 보건복지

Searches on Suicide in Korea: A Big Data Analysis of Google Search Trends, *Yonsei Med Journal, 55*(1), pp.254~263.