

# 빅데이터 분석에 기반한 아동학대의 이해 -머신러닝 알고리즘 개발 기초연구-

배정호<sup>1\*</sup>, 범은애<sup>2</sup>

<sup>1</sup>백석문화대학교 유아교육과 교수, <sup>2</sup>백석문화대학교 간호학과 교수

## Understanding Child Abuse Based on Big Data Analysis -A Basic Study on the Development of Machine Learning Algorithm-

Jungho Bae<sup>1\*</sup>, Eunae Burm<sup>2</sup>

<sup>1</sup>Professor, Child Education, Baekseok Culture University

<sup>2</sup>Professor, Nursing, Baekseok Culture University

**요약** 본 연구의 목적은 아동학대 예방을 위한 방안 마련의 일환으로 빅데이터 분석과 머신러닝 알고리즘을 활용한 정책개발의 기초자료를 제공하는데 있다. 아동학대 예방을 위한 머신러닝 알고리즘 개발을 위한 빅데이터 분석을 위해 학술데이터베이스와 사회관계망서비스 자료를 빅데이터로 정의하고 빈도, 연관어, 감성분석을 시행하였다. 연구결과 예방적 아동학대 알고리즘은 학술빅데이터 분석에 나타난 아동학대 관련 세 주체 피해아동, 가해양육자, 정부당국의 관점에서 아동학대 예방을 위한 데이터 수집 및 공유 네트워크 시스템 마련을 통해 개발이 가능할 것이다. 또한 아동학대 피해아동의 특성에서 자아개념 저하 등으로 우울 및 불안이 나타남을 단서로 영유아 자아존중감 및 우울, 불안 검사를 제도화함으로써 가능할 것이다. 아동학대 예방을 위한 빅데이터 수집 및 분석, 알고리즘 개발 연구의 지속적 진행을 제안하며 아동학대 예방을 위한 실효적 정책 마련이 실현되어 아동학대범죄가 근절되기를 기대한다.

**주제어** : 아동학대, 빅데이터, 머신러닝, 알고리즘

**Abstract** The purpose of this study is to provide basic data on policy development using big data analysis and machine learning algorithms as part of preparing measures to prevent child abuse. In order to analyze big data for developing machine learning algorithms to prevent child abuse, frequency analysis, related word analysis, and emotional analysis were performed after defining academic databases and social network service data as big data. related words, and emotional analysis were conducted. As a result of the study, a preventive child abuse algorithm can be developed by preparing a data collection and sharing network system to prevent child abuse from the perspective of children affected by child abuse, perpetrators, and government authorities. Although it will be possible by institutionalizing infant self-esteem, depression, and anxiety tests with clues that depression and anxiety appear due to a decrease in self-concept in the characteristics of children affected by child abuse. We suggest that continuous progress of big data collection and analysis and algorithm development research to prevent child abuse, and expects that effective policies to prevent child abuse will be realized to eradicate child abuse crimes.

**Key Words** : Child abuse, Big data, Machine learning, Algorithms

## 1. 서론

‘신은 모든 곳에 있을 수 없기에 어머니를 만들었다’는 유대인 격언이 있다. 하지만 그 어머니가 아이에게 폭력과 방임을 하는 학대 가해자라면? 그 부모가 그렇다면 아동의 삶은, 생명은 어떠할까? 상상조차 불가한 고통으로 그 삶은 지옥과 같을 것이다.

아동복지법에 따르면 아동학대란 보호자를 포함한 성인이 아동의 건강, 복지를 해지거나 정상적 발달을 저해할 수 있는 신체적·정신적·성적 폭력 및 가혹행위를 하는 일이다[1]. 신체학대, 정서학대, 성학대는 물론 적극적인 보호의무를 위배하는 방임 또한 아동학대에 포함된다. 즉, 아동의 건강은 물론 더 나아가 복지를 훼손하는 모든 행위를 아동학대로 정의하고 있다. 또한 아동학대처벌법 제2조 제4호 ‘아동학대범죄’란 보호자에 의한 아동학대로서 「형법」 및 「아동복지법」에서 규정하고 있는 죄이다. 상해, 특수상해, 폭행, 특수폭행, 유기, 아동학사, 유기등치사상, 특수협박, 약취, 유인, 매매, 이송, 상해 등이며 위의 몇몇 범죄에 대해서는 미수범까지 아동학대범죄에 포함된다.

보건복지부 ‘2020 아동학대 연차보고서’에 따르면 2020년 아동학대 신고는 38,929건이다[2]. 하지만 아동학대 발견율은 4.02%로 아동학대 실제 아동학대 사건은 매해 약 100만 건(973,225건) 발생한다고 예측할 수 있다. 아동학대는 암수범죄가 많다. 특히 신생아기와 영유아 아동학대범죄는 대부분이 암수범죄이다. 이를 뒷받침하는 구체적 수치는 아동학대 사망자 수로 총 43명의 아동이 아동학대로 생명을 잃었으며, 그중 38명이 신생아와 영유아이다. 결국, 신생아와 영유아 아동학대는 죽어야 끝나는, 아이가 사망한 뒤에야 드러나는 암수범죄기에 선제 대응과 예방이 중요하다.

그동안 우리나라의 아동학대 예방을 위한 정책은 많은 변화와 발전을 이루었다. 아동학대 예방을 위한 중요한 변곡점에는 아동학대 사건이 발생하였으며 이에 대한 국민적 관심과 공분, 공감의 아동학대 예방정책의 변화를 견인하였다. 2020년 온 국민의 공분을 산 ‘정인이 사건’ 이후 「아동학대범죄처벌특례법」이 개정되었으며, 아동학대살해죄가 신설됨은 물론 양형의 기준도 강화되었다. 예방을 위한 아동학대 전담공무원 제도도 도입되었다. 2013년 ‘울산 아동학대 사망 사건’과 ‘칠곡 아동학대 사망 사건’ 이후 2014년 「아동학대범죄의 처벌 등에 관한 특례법」이 시행되어 이후의 아동학대 예방을 위한 정책에 이정표가 되었다. 어린이집 CCTV 의무화(2015), 장

기결석 전수조사(2016년), e아동행복지원서비스(2018년), 아동보호서비스 전달체계 공공성 강화(2019년)의 시행으로[3], 아동학대 예방 정책에서 공적 시스템강화와 선제적 대응에 대한 초석을 다질 수 있었다. 처벌 수위도 한층 강화되어 2014년 이후 아동학대 형사처벌 사례가 급격히 증가했으며, 2021년부터 아동학대 가해자에 대한 양형 기준이 강화되어 이전과 달리 유사 형사범죄의 권고형량보다 높게 적용되었다. 아동학대 예방을 위한 접근의 한 방안으로 처벌의 강화가 있으며 어느 정도 효과를 거두고 있다. 하지만 처벌강화만으로 아동학대 예방에는 한계가 있다. 우리나라의 경우 아동학대 예방을 위한 교육 및 국가 감시 시스템을 적극적으로 활용하고 있으나 아동학대 위험 그룹, 가족과 아동을 대상으로 한 선제적 중재·지원이 더 필요한 실정으로 이를 위한 국가의 재원과 인력이 요구된다. 하지만 현실적으로 아동학대 예방을 위해 막대한 재원과 인력을 투입하는 것은 요원한 일이다. 또한, 재원과 인력 투입만으로 아동학대를 근절할 수도 없을 것이다.

아동학대 예방을 위한 방안 마련을 위해 빅데이터 분석과 머신러닝 알고리즘을 활용한 정책이 전 세계적으로 주목받고 있다[4]. 이는 기존의 아동학대 시스템의 한계를 극복하기 위한 노력으로 미국과 네덜란드, 뉴질랜드에서는 이미 아동학대 예방을 위한 머신러닝 알고리즘을 개발하여 실제 아동보호 현장에 도입함으로써 행정적 성과를 거두고 있다[5-6]. 아동학대의 징후 예측에 있어 구조화된 데이터를 활용하는 사례도 늘고 있다. 하지만 구조화된 데이터는 대부분 의료데이터이기에 일반 연구자들이 접근하기에는 어려움이 있다. 대체적 방법으로 비정형 텍스트 수집과 분석 일반인의 비정형 텍스트는 아동학대 공통어해 증진은 물론 머신러닝을 통해 아동학대를 예측하는 중요한 데이터로 활용될 수 있다[4-6]. 아동학대의 예방적 증재를 위한 머신러닝 적용 기법은 크게 두 가지로 구분된다. 빅데이터 분석에 기반한 머신러닝 개발과 수학과 논리 알고리즘에 기반한 머신러닝이다. 이에 본 연구는 아동학대 관련 빅데이터 분석으로 아동학대 예측 요인을 조사하고자 한다. 연구문제로 첫째, 빅데이터 분석을 통해 아동학대에 대한 사회적 인식과 전문가들의 견해를 분석하고 둘째, 아동학대 예방을 위한 위험요인을 제안하여 예측조합 머신러닝 알고리즘 개발을 위한 기초자료로 활용되고자 한다.

## 2. 아동학대 예방을 위한 빅데이터 분석

아동학대에 대한 일반인과 전문가의 다양한 관점을 담지한 방대한 자료를 수집하여 분석하고자, 아동학대와 관련된 학술연구자료와 데이터베이스(Web) 에서 수집되는 자료를 빅데이터로 정의하였다.

### 2.1 학술연구 자료수집

아동학대에 대한 전문가들의 이해를 알아보기 위해, 아동학대 관련 선행연구 문헌분석을 실시하였다. 한국학술정보서비스(RISS)에서 ‘아동학대’라는 키워드로 검색된 연구물(2022.03.01. output) 4,751건에서 관련이 없는 자료와 중복자료, 그리고 2010~2021년 까지의 자료를 선별하여 총 3,079건을 분석하였다.

### 2.2 사회관계망서비스(Social Networking Service: SNS) 기반 자료수집

아동학대에 대한 일반인들의 이해를 알아보기 위해, 대표적인 사회관계망서비스인 영상공유 플랫폼 유\*브와 \*이버 지식in에서 2010~2021년 까지의 아동학대 관련 영상 및 질문 제목과 댓글을 수집하였다. 유\*브 영상 및 댓글자료는 총 5,737건(영상 488건, 댓글5,249건)이며, 관련 없는 자료 및 중복자료를 제거한 총 5,043개의 자료를 분석하였다. \*이버 지식in 질문 제목 40,213건 중 중복자료와 관련없는 자료를 제외한 총 39,679건의 자료를 분석하였다.

### 2.3 데이터베이스 기반 자료수집(포털 검색엔진 OPEN API)

아동학대에 대한 일반인의 이해를 알아보기 위해 과 뉴스빅데이터 분석 플랫폼 빅카인즈(<https://www.bigkinds.or.kr/>)에서 ‘아동학대’를 주제로 2010~2021년 까지의 자료를 수집하였다. 빅데이터 자료 수집현황 및 자료 정제 절차는 다음 <Table 1>과 같다.

<Table 1> Bigdata collection

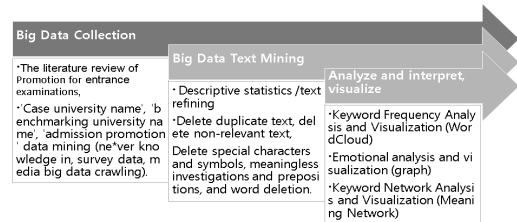
Content	Classification	Social Networking Service	Articles
Collection period	Data collection Target period	2010.01.01.~2021.12.31.	
	Collection period	2022.01.15.~01.31.	
Collection data type	Na*er Jisik-n	Question 40,213-> 39,679(Filter)	

	*outube	Comment5,737-> 5,271(Filter)	
	Dissertation		Dissertation:1,828-> 1,006(Filter)
	Research papers		Research papers ->2,073(Filter)

## 3. 자료의 분석

선행연구 자료들은 문헌분석을 기반으로 기술통계(빈도와 누적통계 등)와 문헌분석, 텍스트분석(워드클라우드 네트워크분석 등)을 실시하였다.

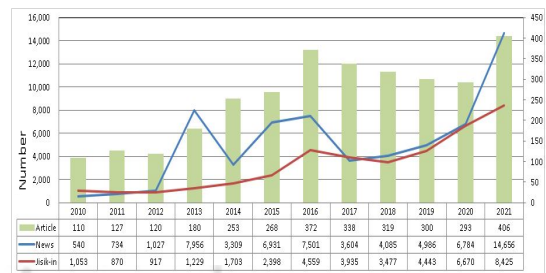
검색엔진 데이터베이스에서 수집된 자료들은 텍스트마이닝(Textmining)을 통해 주요 키워드(Key word)의 기술통계와와 관계성 분석을 실시하고 워드클라우드(Wordcloud)와 네트워크분석((Social Network Analysis)으로 시각화 하였다. 이상의 절차를 빅데이터수집, 텍스트마이닝, 분석 및 시각화(제시)의 세 단계로 제시하면 다음 [Fig. 1]과 같다.



[Fig. 1] Big data analysis and interpretation, visualization procedures[6]

## 4. 아동학대 빅데이터 분석결과

아동학대 빅데이터 자료 빈도 양상은 다음[Fig. 2]과 같다. 학술연구와 뉴스, 지식in 자료 빈도가 모두 함께 상승한 연도는 2021년 ‘정인이사건’ 으로 온 국민의 아



[Fig. 2] Frequency of Bigdata materials by year

동학대에 대한 높은 관심을 보여주고 있다. 이와 더불어 2013년, 2015년, 2016년 사회적으로 이슈가 되었던 아동학대 사건이 뉴스 자료로 드러난 것으로 보이며 이를 반영한 국민적 관심의 증가 시점은 지식in 자료가 증가한 2016년으로 유추해 볼 수 있다.

#### 4.1 학술빅데이터 분석

아동학대에 학술빅데이터 워드클라우드 관계망 분석 결과는 다음 [Fig. 3]와 같다. 아동학대 학술빅데이터는 크게 피해아동, 가해부모, 정부당국 세 주제로 나누어 볼 수 있다. 또한 피해아동의 트라우마 및 후유장애 양상과 극복을 위한 방안 및 지원, 가해부모의 위험요인, 마지막으로 아동학대 예방을 위한 정부당국의 법제정 및 행정 지원 등이 주요 키워드로 분석되었다.

#### 4.2 사회관계망서비스(Social Networking Service: SNS) 빅데이터 분석

아동학대 관련 사회관계망서비스 빅데이터 분석 결과는 [Fig. 4-5], <Table 2>와 같다. 사회관계망 플랫폼의 특징별로 살펴보면, Jisik-in 은 집단지성 기반 지식 공유 플랫폼으로 일반인들(수요자)의 다양한 관심사(질문)을 이해할 수 있다면 Youtube 는 제작자가 영상으로 일반인들(수요자)의 관심 주제나 내용등을 제공한다. 이에 Jisik-in 아동학대 내용은 주로 개별적으로 경험하고 있는 아동학대 개념 및 대처, 신고 등에 대한 내용이 주로 많았다면, Youtube는 조회수를 이끌어 낼 만한 아동학대 사건 및 처벌 등의 정보를 제공하는 자료들이 많은 것을 알 수 있다.

아동학대 관련 개별적 경험에 대한 현상적 이해를 위해 Jisik-in을 팔꿈치기법(주요유목을 정하는 기법)으로 8주제로 유목화한 결과는 [Fig. 5]와 같다. 첫째, 아동학대예방시스템 문제, 가정폭력이 아동학대로 연속, 아동학대 및 성범죄자 취업제한 관련, 학대부모신고 및 처벌 관련, 주양육자의 아동학대 행위 관련, 영유아교육기관 아동학대 관련, 가정폭력 및 아동학대와 이혼 관련, 마지막으로 아동학대로 개념 및 피해관련으로 나누어 볼 수 있다. 이와 함께 주목할 점은 다음 <Table 2> 아동학대 Jisik-in 감성분석 결과에서 부정적 감성지수(Negative sentimentlog\_odds\_ratio)로 가장 높은 단어는 '내가'라는 점이다. 주체이자 존재 당사자가 스스로의 존재에 대해 가장 높은 부정적 감정을 나타내고 있다.

미디어빅데이터 분석결과는 이전 결과들과 다른 양상

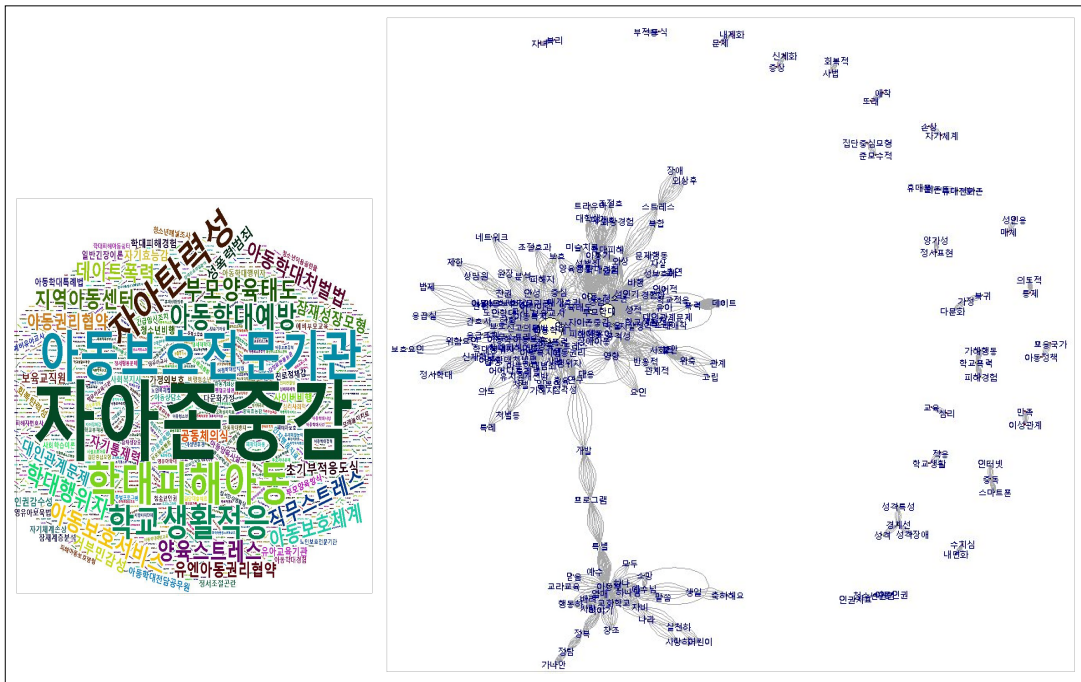
으로 아동학대 관련 내용이 가정 내 보다 영유아 교육기관, 즉 어린이집 및 유치원 등에서 주로 드러났다. 보육시설, 유치원, 종사자, 보육교사와 더불어 인천시 보육시설 아동학대 사건 키워드(인천시, 남구, 가해자 등)이 주요 키워드로 나타나고 있다. 관계망 분석의 기관 및 장소 등에서도 이와 같은 특징이 나타나는 것을 확인할 수 있었다.

#### 4.3 미디어 빅데이터 분석결과

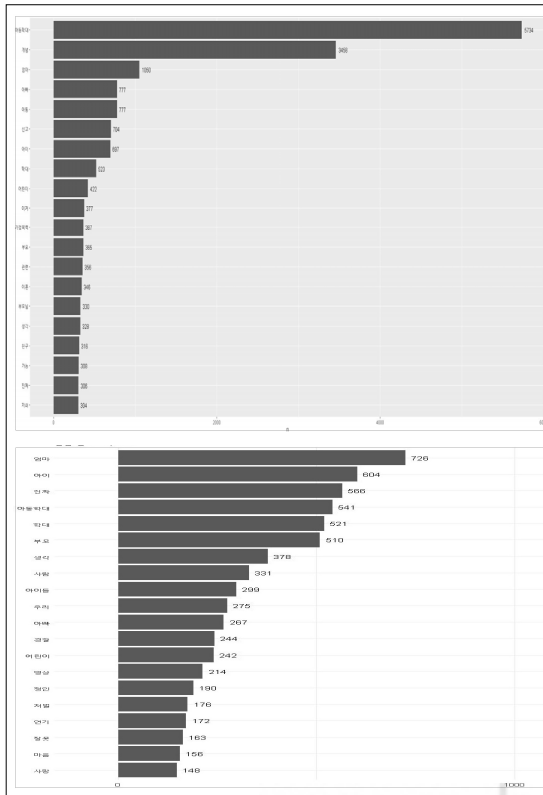
미디어 빅데이터 분석결과는 다음과 같다[Fig. 6]. 보육시설, 유치원, 종사자, 보육교사와 더불어 인천시 보육시설 아동학대 사건 키워드(인천시, 남구, 가해자 등)이 주요 키워드로 나타나고 있다. 이전 결과들과 다른 양상으로 가정 내 아동학대 사건보다 영·유아 교육기관 즉 어린이집 및 유치원 등의 아동학대 관련 내용이 많았다. 인물, 장소, 기관, 키워드를 구분하는 색깔별로 살펴보면 기관은 주로 아동학대 대처를 담당하는 행정기관이, 장소는 공론화된 아동학대 발생 사건장소가, 키워드는 관련 당사자들로 나타났다. 관계망 분석의 기관 및 장소 등에서도 이와 같은 특징이 나타나는 것을 확인할 수 있었다.

<Table 2> Text Analysis of Jisik-in analysis on the sentimentlog\_oddsratio and rank

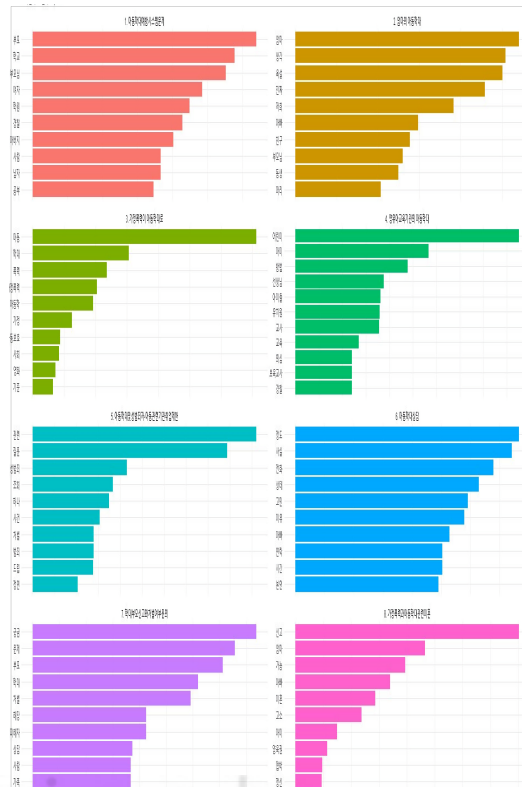
sentiment rank	word	neg word	pos word	sentimentlog_odds_ratio
	<chr>	<int>	<int>	<dbl>
neg-1	내가	104	0	-3.86
neg-2	아빠는	77	0	-3.56
neg-3	남편은	73	0	-3.51
neg-4	남편이	52	0	-3.18
neg-5	폭행	40	0	-2.92
neg-6	의심	79	1	-2.89
neg-7	우울증	35	0	-2.79
neg-8	처벌을	33	0	-2.73
neg-9	범죄	30	0	-2.64
neg-10	때려서	29	0	-2.61
pos-1	예방	0	10	3.19
pos-2	취업	0	9	3.10
pos-3	자격증	0	6	2.74
pos-4	아동	0	6	2.74
pos-5	필수	0	6	2.74
pos-6	공무원	0	5	2.59
pos-7	부모님에	0	5	2.59
pos-8	살면	0	5	2.59
pos-9	여자들이	0	5	2.59
pos-10	이수증	0	5	2.59



[Fig. 3] Text Analysis of Academic Bigdata Word Cloud & Network Analysis



[Fig. 4] Jisik-in & Youtube Word Frequency



[Fig. 5] Jisik-in Topic modeling



REFERENCES

[1] Korea Law Information Center, "CHILD WELFARE ACT," 2022.05.30. <https://law.go.kr/LSW/lsInfoP.do?lsiSeq=198257&viewCls=engLsInfoR&urlMode>

[2] Ministry of Health and Welfare, "2020 Annual Report on Child Abuse," <http://www.mohw.go.kr/react/al/sal0301vw>.

[3] S. H. OH & H. A. Kang, "Analysis of News on Child Abuse based on the Major Policy Changes: Using Text Mining," Journal of the Korean society of child welfare, Vol.70, No.3, pp.1-31, 2021.

[4] Amrit, C., Paauw, T., Aly, R., & Lavric, M., "Identifying child abuse through text mining and machine learning," Expert systems with applications, Vol.88, pp.402-418, 2017.

[5] E.S.Gokten and C.Uyulan, "Prediction of the development of depression and post-traumatic stress disorder in sexually abused children using a random forest classifier," Journal of Affective Disorders, Vol.279, pp.256-265, 2021.

[6] J.I.Sorensen, R.M.Nikam and A.K.Choudhary, "Artificial intelligence in child abuse imaging," Pediatric radiology, Vol.51, No.6, pp.1061-1064, 2021.

[7] J.H.Bae and E.A.Burm, "Strategies for College Entrance Based on Big Data Analysis," Journal of Internet of Things and Convergence, Vol.8, No.2, pp.25-33, 2022.

[8] J.H.Bae and E.A.Burm, "Big data analysis for child well-being on academic research," Medico Legal Update, Vol.20, No.1, pp.2058-2062, 2020.

[9] A. Amado, P. Cortez, P. Rita, and S. Moro, "Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis," European Research on Management and Business Economics, Vol.24, No.1, pp.1-7, 2018.

[10] J.H.Bae and E.A.Burm, "Big data analysis: Medical accident," Medico-Legal Update, Vol.19, No.1, pp.646-652, 2019.

[11] J.H.Bae and E.A.Burm, "Big data analysis for child well-being on academic research," Medico Legal Update, Vol.20, No.1, pp.2058-2062, 2020.

[12] A.V.Annapragada, M.D.Donaruma-Kwuh, A.V. Anna-pragada and Z.A.Starosolski, "A natural language processing and deep learning approach to identify child abuse from pediatric electronic medical records," PLoS ONE Vol.16, No.2, e0247404.

[13] C.Armit, T.Paauw, R.Aly & M.Lavric, "Identifying child abuse through text mining and machine learning," Expert Systems With Applications, 88(1), pp402-418, 2017.

[14] M. Fattori, G. Pedrazzi & Turra, R. "Text mining applied to patent mapping," A practical business case.

World Patent Information, 25(4), pp335-342, 2003.

[15] C.Henry, S.Carnochan, & M.J.Austin, "Using qualitative data-mining for practice research in child welfare," Child Welfare: Arlington, 93(6), pp7-26, 2014.

배 정 호(Jungho Bae)

[정회원]



- 2005년 2월 : 중앙대학교 교육대학원 교육학과 유아교육전공 (교육학석사)
- 2013년 2월 : 중앙대학교 일반대학원 유아교육학과 (문학박사)
- 2012년 3월 ~ 현재 : 서울특별시 유치원 컨설팅 장학위원
- 2016년 3월 ~ 현재 : 백석문화대학교 유아교육과 교수

<관심분야>

유아교육학, 교육학, 사회학, 빅데이터, 머신러닝

범 은 애(Eunae Burm)

[정회원]



- 2013년 2월 : 단국대학교 일반대학원 예방의학과 (의학석사)
- 2016년 8월 : 단국대학교 일반대학원 보건학과 (보건학박사)
- 2006년 5월 ~ 2015년 2월 : 단국대학교 예방의학교실 연구원
- 2018년 3월 ~ 현재 : 백석문화대학교 간호학과 교수

<관심분야>

간호학, 보건학, 사회학, 빅데이터, 머신러닝