J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr
Copyright ⓒ 2021 KSII

LDA 토픽 모델링을 이용한 액티브 시니어 콘텐츠 트렌드 분석☆

Active Senior Contents Trend Analysis using LDA Topic Modeling

이 동 우[1] 김 유 신[1] 신 은 정[2*]
Dongwoo Lee Yoosin Kim Eunjung Shin

요 약

베이비 부머 세대가 고령인구로 진입하면서 기존의 시니어와는 달리 활동적인 모습을 보이는 시니어들이 액티브 시니어라는 신조어로 불리며 새로운 소비자 층으로 떠오르고 있다. 많은 국가들과 기업들도 이들을 주목하고 관련 정책이나 서비스를 제공하고자 하지만 액티브 시니어 트렌드에 대한 연구는 매우 부족한 실정이다. 본 논문에서는 이러한 액티브 시니어에 대한 특징과 이들이 생산하고 소비하는 온라인 미디어 콘텐츠 트렌드를 파악하여,액티브 시니어를 적극 포용하고 지원할 수 있는 온라인 미디어에 대한 정책 및 서비스 방향성을 제시하고자한다. 이를 위해 소셜 미디어에서 액티브 시니어를 수집 키워드로 2018년 1월 1일부터 2021년 6월 31일까지 8,740건의 데이터를 수집하여 키워드 빈도 분석, TF-IDF 분석, LDA 토픽 모델링 분석을 하였다. 키워드 빈도 분석 및 TF-IDF 분석을 통해서 액티브시니어에 대한 관심도가 급증하고 있다는 것을 파악하였으며 LDA 토픽 모델링 분석을통해서 온라인 콘텐츠의 주제 영역을 10가지로 분류하고 라이프 스타일, 혜택, 쇼핑, 정부 사업, 정부 교육, 건강, 사회/경제, 케어 산업, 실버 주택, 여가로 명명하였다.

☞주제어: 액티브 시니어, 빅데이터, 텍스트 마이닝, LDA 토픽 모델링, 트렌드 분석

ABSTRACT

The purpose of this study is to understand the characteristics and trends of active senior. As the baby boom generation become the age of the elderly, they are more active than senior. These seniors are called active seniors, a new consumer group. Many countries and companies are also interested in providing relevant policies and services, but there is lack of researches on active senior trends. This study collects the 8,740 posts related to active seniors on social media from January 1st, 2018 to June 31st, 2021, and conducted keyword frequency analysis, TF-IDF analysis and LDA topic modeling. Through LDA topic modeling, topics are classified into 10 categories: lifestyle, benefits, shopping, government business, government education, health, society and economy, care industry, silver housing, leisure. The results of this study can be utilized as fundamental data to help understand the academic and industrial aspects of active senior.

☞ keyword : Active Senior, Big Data, Text Mining, LDA Topic Modeling, Trend Analysis

1. 서 론

통계청이 발표한 '2020 고령자 통계'에 따르면 2020년 65세 이상 고령인구는 우리나라 인구의 15.7%를 차지하고 있으며 2025년에는 20.3%에 이른다고 하였다. 따라서 현재는 고령사회 단계이지만 2025년에는 초고령사회 단계로 진입하고 2060년이되면 43.9%가 고령인구가 될 것으로 전망하였다[1]. 이처럼 우리나라는 고령 인구의 증가가 다른 국가들보다 빠른 편에 속한다[2]. 고령화의 원인은 저출산과 경제성장, 의료기술의 발달로 볼 수 있으며 지속적인 고령인구의 증가는 우리 사회에 많은 영향을 주었다. 노인 개인에게는 신체적 노화로 인한 건강상 문제와 은퇴로 인한 수입의 감소, 사회로부터의 소외 등 삶의 질이 저하되며 가족 구성원과 사회, 경제, 정치 영역에서는 노인 부양에 대한 세대간 갈등과 건강상의 문제로 인한 부담, 생산가능인구의 감소로 인한 경제성장률 저하 그리고 고령화에 비례하여 증가하는 의료비 및 연금의 증가 등이 있다[3, 4].

최근 고령화 현상을 긍정적으로 보고자하는 노력들이 나타나고 있다. 국가마다 액티브 에이징(active aging) 정책을 강조하고 있는데 이는 노년을 건강하고 활기차게 보내며 적극적으로 사회에 참여하여 긍정적으로 보낼 수 있다는 개념이다[5]. 따라서 시니어 세대에 대한 연구가

[1] RTdata Lab, Seoul, 06655, Korea
[2] Yonsei University, Seoul, 03722, Korea
* Corresponding Author (ej_shin@yonsei.ac.kr)
[Received 16 October 2021, Reviewed 19 October 2021(R2 25 October 2021), Accepted 27 October 2021]
☆ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 연구개발 지원사업으로 수행되었음(과제번호: R2020040068)

http://dx.doi.org/10.7472/jksii.2021.22.5.35

국내외활발하게 연구되고 있으며, 노인의 삶의 질과 만족도에 대한 연구 또한 활발히 이루어지고 있다[6]. 그 중 시니어 세대의 콘텐츠 제작에 대한 관심이 급증하고 있다. 시니어들이 디지털 플랫폼을 통해 다른 사람과 상호교류를 형성하면서 능동적으로 다양한 활동 참여를 통해 스스로 만족감을 느끼며 자기존중감을 체험하게 된다[7]. 이렇듯 기존의 시니어보다 활동적으로 노년을 보내는 특징을 가지며 기존 시니어에 비해 다양한 스마트 기기를 활발하게 사용하여 콘텐츠를 이용하고 쇼핑을 하는 시니어를 액티브 시니어라 한다[8,9,10]. 이들은 새로운 소비자 층으로 급부상하고 있다. 그러나 주로 액티브 시니어의 여가에 대한 연구가 주로 이루어졌을 뿐 액티브 시니어 전반에 대해 알아보고, 그 트렌드에 대해 분석한 연구는 부족한 실정이다.

따라서 본 연구에서는 '액티브 시니어' 키워드로 수집된 소셜 미디어 데이터를 분석하여 새로운 소비자층으로 떠오르는 액티브 시니어를 대상으로 그들의 특징과 액티브 시니어와 관련된 트렌드를 파악하고자 한다. 본 연구의 결과는 액티브 시니어를 대상으로 한 연구와 기업의 기초자료가 될 것으로 사료된다.

2. 이론적 배경

2.1 액티브 시니어

2020년을 기점으로 베이비 부머 세대가 고령인구로 진입하면서 시니어의 대한 정의가 새롭게 정의 되고 있다. 액티브 시니어(active senior)라고 일컫는 신조어가 등장했는데 이들은 기존 시니어와는 구분되는 새로운 시니어층으로 베이비 부머 세대가 포함되어 있는 55~64세를 뜻한다. 은퇴 후에도 적극적으로 사회활동에 참여 하고 경제력 여유를 바탕으로 소비 활동을 하며 이전 시니어층보다 젊고 건강하고 활동적으로 노년을 보내는 특징이 있다[8, 9]. 또한 이들은 기존 시니어에 비해 다양한 스마트 기기를 활발하게 사용하여 콘텐츠를 이용하고 쇼핑을 하면서 모바일 시장에서 새로운 소비자 층으로 급부상하고 있다[10].

이러한 액티브 시니어의 특징을 시니어 유튜버에서 찾아볼 수 있다. '유튜브 크리에이터'란 유튜브와 같은 1인 미디어 플랫폼을 활용하여 스스로 콘텐츠를 기획하고 제작하여 타인과 공유하는 사람을 의미하는데[11] 이들 중 시니어가 직접 콘텐츠를 기획하고 제작 하는 사람을 '시니어 유튜버', '시니어 크리에이터' 라고 부른다. 'Korean Grandma'로 알려진 박막례와 패션유튜버 '밀라논나' 장명숙, '할담비' 지병수가 대표적인 시니어 유튜버이다. 이러한 현상은 디지털 에이징(digital aging)이라는 개념과 관련이 있는데 노화를 디지털 정보통신기술의 융합으로 극복하자는 뜻에서 제안된 개념으로 고영삼(2016)의 연구에서는 디지털 에이징을 "디지털 기술을 활용하여, 고령자가 건강하고, 안전하며, 독립적이고 활동적인 삶을 영위하는 것을 지원함으로써, 고령자의 삶의 질을 향상하고, 고령화 문제 해결에 기여하는 신노인복지 전략"으로 정의하기도 했다[12].

2.2 트렌드 분석

트렌드 및 사용자 인식 분석에 대한 연구는 다양한 분야에서 이루어지고 있다. 특히 소셜 미디어나 커뮤니티의 게시물과 같은 비정형 문서에 대한 분석에 대한 방법으로는 키워드 빈도 분석, LDA(Latent Dirichlet allocation) 토픽 모델링 등의 방법이 있다.

소셜 미디어와 같은 비정형 문서에 대한 언어 분석은 언어통계학적으로 중요하게 다루어지고 있으며 그 중에서도 가장 대표적이며 바탕이 되는 방법이 키워드 빈도 분석이다[13]. 키워드 빈도 분석은 기본적으로 단어가 여러번 등장할 경우 중요도가 높은 단어라는 가정을 근간으로 하며 많이 등장하는 단어를 보고자 할 때 사용한다[14]. TF(Term Frequency) 값은 문서에서 해당 단어가 몇 번 출현했는지 빈도를 나타내는 값이다. 하지만 단순하게 많이 등장하는 단어가 가장 유용하다라고 보기에는 문서에 자주 나타나는 의미 없는 단어들이 존재하기 때문에 그러한 단어의 중요도를 낮추어 볼 필요가 있다. 이러한 상황에서 살펴보는 것이 TF-IDF(Term Frequency-Inverse Document Frequency) 값이다. TF-IDF 값은 단어 중요도의 가중치로 여러 문서로 이루어진 문서군이 주어질 때 특정 단어가 한 문서 내에서 얼마나 중요한 지를 나타내는 통계적 수치이다. TF 값의 한계를 보완하기 위하여 IDF 값을 곱한 값을 의미하며, IDF(Inverse Document Frequency) 값은 전체 문서의 개수 대비 해당 단어가 등장한 문서의 수를 가리키는 DF(Document Frequency) 값의 역수로 다른 문서에서 출현하지 않을수록 높은 값을 가지는 가중치이다.

이진명, 나종연(2015)은 해외 직접 구매에 대한 소비자들의 경향을 알아보기 위하여 온라인 커뮤니티의 글을 대상으로 키워드 빈도 분석을 시행하는 연구를 하였으며 허준석, 이은정(2019)의 연구에서는 특정 브랜드의 트렌

드를 분석하고자 다양한 채널을 통해 해당 브랜드와 관련된 문서들을 수집하고 TF, TF-IDF 값을 도출하였고 이를 통해 트렌드를 분석하였다[15, 16]. 또한 한기향(2019)은 TF, TF-IDF 값을 통해서 하이 서울 패션쇼에 대한 소비자 인식을 조사하는 연구를 하였다[17].

토픽 모델링은 다수의 문서 집합에서 구별되는 주제를 추출하는 기법이다. 문서 집합을 연구자가 정해준 주제 수로 구분하고 각 주제별로 어떠한 단어가 사용되었는지 분석한다[18]. 토픽 모델링은 주제와 단어 간에 어떠한 관계가 존재한다는 생각에서 출발한다. 즉, 단어 자체에서 의미를 찾는 것이 아니라 그 단어가 어떤 단어들과 주로 같이 사용되는지에 따라 의미를 정의하는 것이다. 그러므로 문서 내의 단어의 위치에 관계없이 동시에 등장하는 단어들을 분석하고 빈번하게 반복 등장할수록 하나의 주제로 인식하는 기법이다. 그 중 단어의 집합으로 이루어진 문서들의 주제에 대한 단어 생성 확률 모델을 적용한 기법이 가장 많이 사용되는 LDA 토픽 모델링이다.

LDA 토픽 모델링은 비정형 텍스트 분석에서 자주 쓰이는 기법으로 문서 집합에서 연구자가 원하는 주제 수만큼 주제를 도출할 수 있으며 도출된 주제의 형태는 단어로 구성된 벡터들의 확률 분포가 도출된다. 각 주제별 단어의 확률 분포를 통해 단어가 해당 주제에서 가지는 영향력도 알 수 있다.

안효선, 박민정(2017)의 연구에서는 패션 디자인에 대한 사용자 인식을 알아보기 위해 소셜 미디어 게시물을 수집, LDA 토픽 모델링을 적용하고 분석하였으며 박상언(2016)은 드라마 관련한 트렌드를 알아보고자 소셜 미디어 게시물을 대상으로 LDA 토픽 모델링을 하였다[19, 20]. 또한 박건철, 이치형(2019)의 연구는 스마트시티와 관련된 학술 논문들의 제목과 초록을 수집하여 LDA 토픽 모델링을 통해 관련 주제들을 알아보았다[21].

2.3 시니어 트렌드 분석

'액티브 시니어'라는 용어가 등장하기 이전에는 '고령'이라는 키워드로 연구가 많이 이루어 졌다. 오민정, 정진철, 이미현(2012)의 연구는 1991년부터 2011년까지 총 64편의 연구를 살펴보고, 주제별로 분류하여 시기별로 실버세대와 관련하여 어떤 부분에 관심이 많은지 알아보았으며 1991년부터 1999년에는 소비자 문제, 소비자 사회화, 불평행동, 구매행동, 정보탐색, 마케팅 전략, 인지연령처럼 학문적인 영역에서 그쳤으며 개인적인 영역에서의 연구는 거의 이루어 지지 않았으나, 2000년도에는 초기에는 저축 행동, 여가, 스포츠, 실버타운, 연령과 호칭 등 개인적인 영역의 주제에 관심을 가지기 시작하였으며 2010년대에 들어서는 라이프 스타일, 실버화장품, 관광지, 의료비, 노후의 삶 등 문화적인 영역의 주제까지 관심분야가 늘어났다는 점을 알아내었다[22].

또한 오민정(2019)은 1996년부터 2018년까지 '고령'과 관련있는 1,291개의 논문의 초록을 대상으로 토픽 모델링을 통해 핵심 키워드를 도출하였다. 1996년부터 2009년까지는 정부의 고령자 부양 제도 등에 관심이 많았으며 2010년부터 2014년까지는 고령자 처우, 은퇴, 고용에 대한 실질적 지원 정책 2015년부터 2018년까지는 지원 정책에서 벗어나 스스로 실질적인 소득 및 재정의 마련에 대한 관심과 건강 등에 관심이 생겼다는 점을 알아냈다[23].

3. 연구 방법

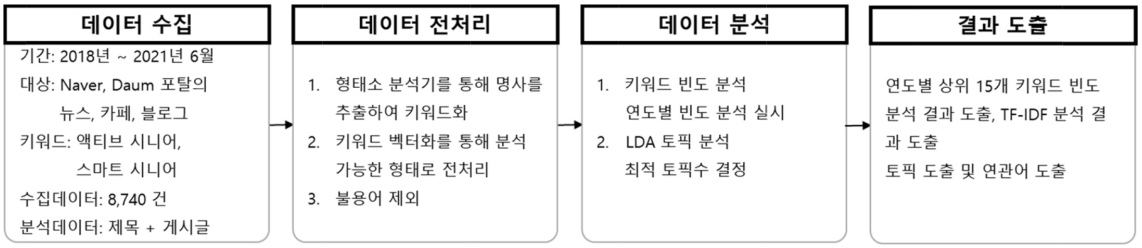
본 연구에서는 빅데이터 분석을 활용하여 액티브 시니어 트렌드에 대해 실증적으로 분석하고자 한다.

3.1 데이터 수집

액티브 시니어의 관심사를 조사하기 위해 대한민국 대표 포털사이트인 네이버와 다음에서 이들과 관련된 사회적 이슈와 개인의 생각이나 경험, 정보 등을 솔직하게 공유할 수 있는 채널인 뉴스, 블로그, 카페 게시글을 분석 대상으로 선정하였다. 자료 수집은 해당 포털사이트에서 제공하는 API를 사용하였고, 제공하지 않는 메타 데이터가 있는 채널에서는 직접 웹 크롤러(web crawler) 프로그램을 개발하여 베이비 붐 세대가 고령인구로 진입하기 시작한 2018년 1월 1일부터 2021년 6월 31일까지 소셜미디어에서 대상 키워드를 검색하여 글의 제목, 본문, 게시일, url 등을 수집하였다. 데이터 수집 키워드는 '액티브 시니어', '액티브 시니어'와 의미가 유사한 '스마트 시니어'[10]를 활용하였고, 이를 통해 총 8,740건의 데이터를 수집하여 최종 분석 데이터로 사용하였다.

3.2 데이터 전처리

수집된 데이터를 분석 목적에 따라서 조사나 감탄사, 부사, 대명사, 영문, 특수문자 등의 분석에 방해가 되는 요소를 제거하고 키워드 형태로 추출하고자 다음과 같은 데이터 전처리 과정을 거친다.



(그림 1) 연구 프레임 워크
(Figure 1) Research framework of the study

먼저, 형태소 분석을 하였다. 형태소 분석은 텍스트를 의미를 가지는 가장 최소의 단위인 형태소로 나누는 것을 말한다. 자연어 처리를 하기 위해서는 의미상 최소 단위로 나누어 주어야 분석이 용이하다. 한국어 형태소 분석기는 다양한 종류가 있지만 가장 최근까지 업데이트 되고 분석단계에서 정확도와 속도 등을 고려하여 kiwipiepy라는 Python 라이브러리를 사용하였다.

둘째, '일반명사, '고유명사'만을 추출하였다. 또한 명사 추출 과정에서 코로나 백신을 뜻하는 새로운 단어 '아스트라제네카'나 1인경제를 뜻하는 합성어 '1코노미' 등의 형태소 분석기가 인식하지 못하는 명사들이 등장하였고 이러한 총 25개의 명사의 경우는 사용자 사전에 추가하였다. 그 다음, 모델이 인식할 수 있는 벡터 형태로 변환하는 과정을 거쳤다.

셋째, '액티브 시니어'와 관련있는 주제 이지만 문장내에서 분석에 방해가 되는 '불용어'를 결정하였다. '이번', '최근', '최초', '올해', '내년', '지난해', '오후', '이날', '오늘', '처음', '이후'와 같이 시간과 관련있는 명사들을 제외하였고 '이상', '대상', '적극', '과정', '진행', '증가', '관련', '기준', '가운데', '전체', '다양', '결과', '경우', '이하', '비율', '과정', '상황', '기반'과 같이 전치 명사의 범위를 한정하는 명사들도 제외하였다. 이외에 '뉴시스', '이데일리' 신문사나 매체 이름과 '사진', '제공'과 같이 주석에 해당하는 단어도 불용어에 포함시켰다.

3.3 데이터 분석

데이터 분석은 Python을 활용하여 tomotopy라는 라이브러리를 사용하였으며 키워드 빈도 분석, TF-IDF 분석, LDA 토픽 모델링 기법을 활용하였다. 전처리 과정을 모두 거치면 각 문서들은 형태소 단위 변환되는데 이 형태의 데이터를 모델에 학습시키면 키워드 빈도 분석과 TF-IDF 분석이 가능하다. 추가로 일관성 점수를 도출하여 최적 주제 수까지 결정하여 LDA 토픽 모델링 분석까지 하였다.

4. 연구 결과

4.1 키워드 빈도 분석과 TF- IDF

소셜 미디어 데이터를 수집하여 빈도 분석과 TF- IDF 분석을 한 결과는 다음과 같다. 액티브 시니어와 관련된 핵심 키워드 중 연도별 상위 15개 키워드는 표 1-4와 같다.

2018년 액티브 시니어와 관련하여 소셜 미디어 상에서

(표 1) 2018년 TF 및 TF-IDF 분석 결과
(Table 1) TF and TF-IDF results in 2018

순위	키워드	TF	키워드	TF-IDF
1	건강	69	건강	117.8
2	활동	64	소비	106.2
3	소비	63	활동	104.2
4	세대	50	여행	97.8
5	생활	47	비타민	95.4
6	사회	46	생활	92.2
7	은퇴	42	세대	88.5
8	여가	39	사회	87.8
9	여행	39	고령	86.1
10	고령	38	문화	82.6
11	문화	37	은퇴	80.2
12	비타민	33	여가	76.5
13	자신	31	페어	73.1
14	페어	27	자신	70.3
15	능동	26	멀티	68.1

(표 2) 2019년 TF 및 TF-IDF 분석 결과
(Table 2) TF and TF-IDF results in 2019

순위	키워드	TF	키워드	TF-IDF
1	건강	255	건강	419.0
2	사회	216	타운	395.1
3	활동	194	서비스	332.4
4	서비스	194	사회	323.9
5	타운	170	협력	319.7
6	은퇴	152	활동	310.6
7	세대	152	세대	303.2
8	협력	144	여행	288.3
9	고령	142	고령	277.1
10	소비	126	라이프	273.0
11	여행	126	교육	272.1
12	라이프	115	은퇴	266.6
13	교육	114	문화	264.4
14	문화	112	클래식	264.4
15	클래식	112	소비	262.9

(표 3) 2020년 TF 및 TF-IDF 분석 결과
(Table 3) TF and TF-IDF results in 2020

순위	키워드	TF	키워드	TF-IDF
1	활동	284	문화	422.5
2	문화	227	건강	399.3
3	건강	201	활동	391.4
4	사회	181	콘텐츠	381.7
5	콘텐츠	164	여행	363.1
6	여행	156	서비스	340.6
7	세대	154	사회	324.3
8	소비	152	사업	322.1
9	서비스	147	세대	319.0
10	시대	138	소비	300.8
11	사업	129	시대	279.3
12	맞춤	110	노인	273.3
13	지원	108	생활	259.5
14	생활	108	지원	252.5
15	고령	101	라이프	247.1

가장 많이 언급된 TOP 5 키워드는 '건강(69)', '활동(64)', '소비(63)', '세대(50)', '생활(47)'이다. 문서 내에서 특정 단어의 중요도를 나타내는 지표인 TF-IDF 분석 결과로는

(표 4) 2021년 TF 및 TF-IDF 분석 결과
(Table 4) TF and TF-IDF results in 2021

순위	키워드	TF	키워드	TF-IDF
1	세대	332	건강	487.3
2	건강	256	세대	452.5
3	활동	238	시장	423.4
4	은퇴	196	노인	404.3
5	노인	183	활동	373.9
6	시장	176	은퇴	331.5
7	콘텐츠	147	놀이터	331.2
8	대표	139	콘텐츠	313.4
9	사회	131	여행	310.9
10	소비	129	대표	288.5
11	서비스	126	사회	283.7
12	참여	122	소비	279.3
13	도전	118	서비스	277.2
14	문화	113	문화	276.9
15	놀이터	112	도전	270.6

'건강(117.8)', '소비(106.2)', '활동(104.2)', '여행(97.8)', '비타민(95.4)' 순으로 가장 중요도가 높은 단어로 나타났다. 2019년 액티브 시니어와 관련하여 소셜 미디어 상에서 가장 많이 언급된 TOP 5 키워드는 '건강(255)', '사회(216)', '활동(194)', '서비스(194)', '타운(170)'이다. TF-IDF 분석 결과로는 '건강(117.8)', '타운(106.2)', '서비스(332.4)', '사회(323.9)', '협력(319.7)' 순으로 가장 중요도가 높은 단어로 나타났다. 2018년 빈도 분석 결과와 비교하여 2019년 상위 15개 결과에서 새롭게 도출된 키워드로는 시니어 하우스 서비스를 제공하는 기업이름인 '더 클래식 500'의 '클래식'이라는 키워드와 '서비스', '타운', '협력', '라이프', '교육'과 같이 기존 실버 세대을 타깃으로 한 지원 사업과 관련된 키워드들이 나타났다. 2018년과 비교하여 빈도 분석 결과 상위 15개로 도출되는 키워드의 빈도수가 약 4배 증가한 것을 통해 소셜 미디어 상에서 액티브 시니어에 대한 언급이 급증함을 알 수 있었고, 이를 통해 액티브 시니어에 대한 관심이 급증함을 확인할 수 있었다. 2020년 액티브 시니어와 관련하여 소셜 미디어 상에서 가장 많이 언급된 TOP 5 키워드는 '활동(284)', '문화 (227)', '건강 (201)', '사회 (181)', '콘텐츠(164)'이다. TF-IDF 분석 결과로는 '문화(422.5)', '건강(399.3)', '활동(391.4)', '콘텐츠(381.7)', '여행(363.1)' 순으로 가장 중요도가 높은 단어로 나타났다. 2019년 빈도 분

(표 5) LDA 토픽 모델링 결과
(Table 5) LDA topic modeling results

번호	주제	1 순위	2 순위	3 순위	4 순위	5 순위
주제 1	라이프 스타일	공개	스타일	콘텐츠	활약	개성
주제 2	혜택	프로모션	강좌	구매	자동차	기아차
주제 3	쇼핑	클럽	롯데	회원	교육	유료
주제 4	정부 사업	시장	김상돈	윤화섭	놀이터	노인
주제 5	정부 교육	인생	준비	스마트	일자리	기대
주제 6	건강	건강	라이프	기능	출시	식품
주제 7	사회/경제	소비	세대	문화	사회	고령
주제 8	케어 산업	콘텐츠	지도	복약	말벗	탑재
주제 9	실버 주택	타운	클래식	협력	양성	서울
주제 10	여가	디자인	놀이터	노인	공공	뉴딜

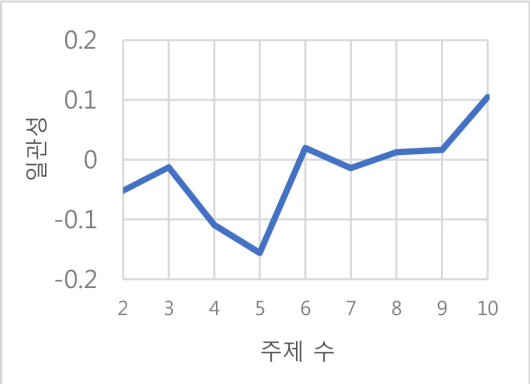
석 결과와 비교하여 상위 15개 결과로 새롭게 도출된 키워드로는 ‘콘텐츠’, ‘시대’, ‘사업’, ‘맞춤’, ‘지원’, ‘생활’, ‘노인’ 이 있다. 2021년 액티브 시니어와 관련하여 소셜 미디어 상에서 가장 많이 언급된 TOP 5 키워드는 ‘세대(332)’, ‘건강(256)’, ‘활동(238)’, ‘은퇴(196)’, ‘노인(183)’이다. TF-IDF 분석 결과로는 ‘건강(487.3)’, ‘세대(452.5)’, ‘시장(423.4)’, ‘노인(404.3)’, ‘활동(373.9)’ 순으로 가장 높은 단어로 나타났다. 2020년 빈도 분석 결과와 비교하여 2021년 상위 15개 결과에서 새롭게 도출된 키워드로는 ‘은퇴’, ‘노인’, ‘시장’, ‘대표’, ‘참여’, ‘도전’, ‘놀이터’가 있다.

4.2 LDA 토픽 모델링

전체 수집 기간인 2018년 1월 1일부터 2021년 6월 31일까지의 텍스트를 LDA 알고리즘에 적용하여 10개의 토픽을 도출하였다. 본 연구의 구체적인 토픽 모델링 과정은 다음과 같다. 그림 2와 같이 npmi 방식으로 각 토픽 수별로 일관성 점수를 도출하여 가장 높은 점수일 때 최적으로 판단한다. 따라서 토픽이 10개일 때 가장 일관성이 높게 도출되어 최적 토픽 수는 10개로 결정하였다. 표 5와 같이 각 토픽을 구성하는 연관어들은 토픽 모델링의 수행 결과로 출력된 것으로 토픽에서 해당 단어가 차지하는 비중이 높은 순으로 5개의 연관어가 추출되었다.

각 토픽별 키워드를 대표하는 토픽에 대한 명명을 수행한 결과는 다음과 같다. 토픽 1은 공개(0.01), 스타일(0.01), 콘텐츠(0.01), 활약(0.01), 개성(0.01)가 연관어로 도출이 되어 ‘라이프 스타일’로 명명하였으며, 토픽 2는 프로모션(0.02), 강좌(0.01), 구매(0.01), 자동차(0.01), 기아차(0.01) 가 연관어로 도출이 되어 ‘혜택’으로 명명하였다. 또한 토픽 3은 클럽(0.02), 롯데(0.01), 회원(0.01), 교육(0.01), 유료(0.01) 가 연관어로 도출이 되어 ‘쇼핑’으로 명명하였으며, 토픽 4는 시장(0.03), 김상돈(0.02), 윤화(0.02), 놀이터(0.02), 노인(0.02) 가 연관어로 도출이 되어 ‘정부 사업’으로 명명하였다. 토픽 5는 인생(0.01), 준비(0.01), 스마트(0.01), 일자리(0.01), 기대(0.01) 가 연관어로 도출이 되어 ‘정부 교육’으로 명명하였고, 토픽 6은 건강(0.03), 라이프(0.02), 기능(0.01), 출시(0.01), 식품(0.01) 가 연관어로 도출이 되어 ‘건강’이라 명명하였다. 토픽 7은 소비(0.01), 세대(0.01), 문화(0.01), 사회(0.01), 고령(0.01) 가 연관어로 도출이 되어 ‘사회 경제’로 명명하였으며 토픽 8은 콘텐츠(0.02), 지도(0.01), 복약(0.01), 말벗(0.01), 탑재(0.01) 가 연관어로 도출이 되어 ‘케어 산업’으로 명명하였다. 토픽 9는 타운(0.03), 클래식(0.02), 협력(0.02), 양성(0.02), 서울(0.02) 가 연관어로 도출이 되어 ‘실버 주택’으로 명명하였고 토픽 10은 디자인(0.02), 놀이터(0.02), 노인(0.01), 공공(0.01), 뉴딜(0.01)가 연관어로 도출이 되어 ‘여가’로 명명하였다.

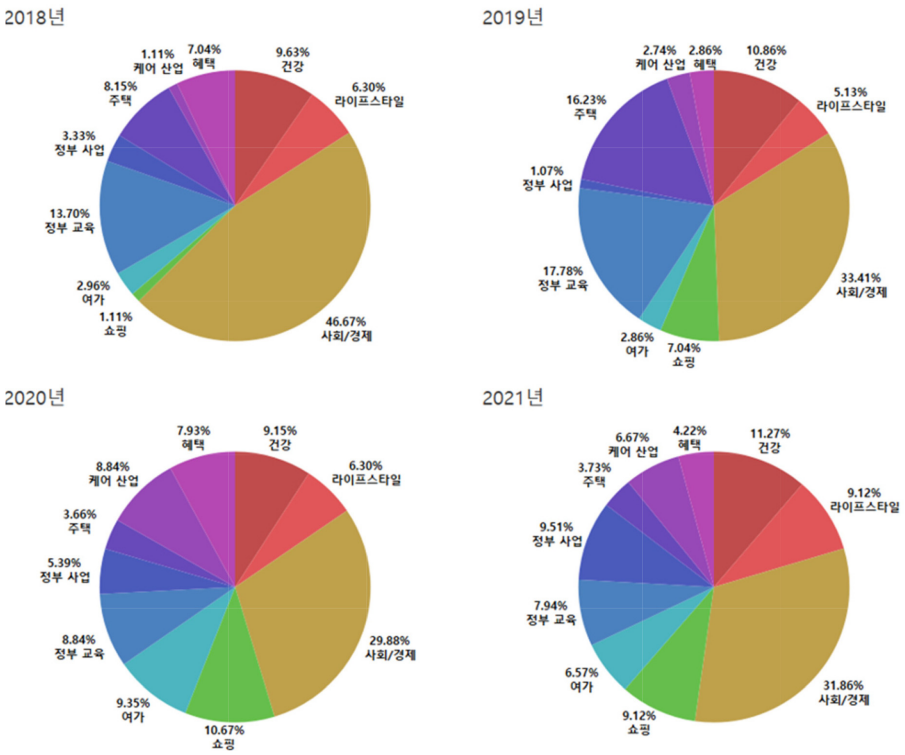
각 토픽별로 등장할 확률은 표 6과 같다. 토픽 7: 사회/경제가 32.94%로 가장 많이 등장한다는 것을 알 수 있었고, 그 다음은 토픽 5: 정부 교육(11.38%), 토픽 6: 건강(10.35%), 토픽 3: 쇼핑(8.35%), 토픽 9: 실버 주택(7.46 %), 토픽 1: 라이프 스타일(6.91 %), 토픽 10: 여가(6.14 %), 토픽 4: 정부 사업(5.40 %), 토픽 2: 혜택(5.27 %) 순서로 등장한다는 것을 알 수 있었다.



(그림 2) 토픽수 별 일관성 점수
(Figure 2) Coherence score by number of topics

(표 6) 토픽별 출현 비중
(Table 6) Proportion of appearance by topic

토픽	비중(%)
토픽 1	6.91
토픽 2	5.27
토픽 3	8.35
토픽 4	5.40
토픽 5	11.38
토픽 6	10.35
토픽 7	32.94
토픽 8	5.82
토픽 9	7.46
토픽 10	6.14

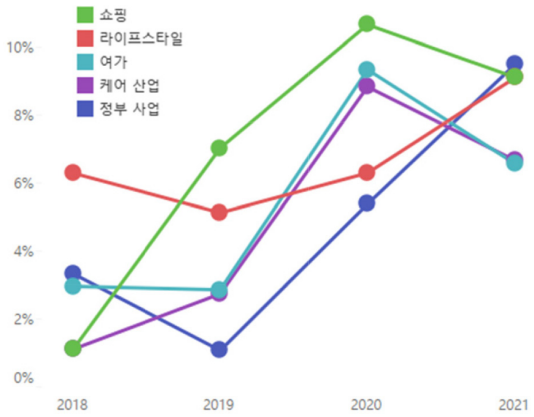


(그림 3) 액티브 시니어 토픽 트렌드
(Figure 3) Active senior topic trend

4.3 연도별 토픽 트렌드

액티브 시니어를 키워드로 최근 3년 6개월 기간의 데이터를 수집하여 LDA 토픽 모델링 결과, 10개의 토픽을 도출하였다. 10개의 토픽의 시계열적 변화를 통해 토픽 트렌드를 분석하기 위해 연도별로 각 토픽에 해당되는 문서 수를 도출하였다. 토픽의 시계열적 변화를 살펴본 결과는 그림 3과 같다. 2018년 상위 5개 토픽은 사회경제(46.47%), 정부교육(13.7%), 건강(9.63%), 주택(8.15%), 혜택(7.04%)이었다. 2019년 상위 5개 토픽은 사회경제(33.41%), 정부교육(17.78%), 주택(16.23%), 건강(10.86%), 쇼핑(7.04%) 이었다.2020년 상위 5개 토픽은 사회경제(29.88%), 쇼핑(10.67%), 여가(9.35%), 건강(9.15%), 정부교육(8.84%)이었다. 2021년 상위 5개 토픽은 사회경제(31.86%), 정부교육(11.27%), 건강(9.51%), 주택(9.12%), 혜택(9.12%)이었다. 사회/경제 토픽이 매해 가장 큰 비중을 차지하였으나 2018년 이후 비중이 10% 이상 감소함을 확인할 수 있었다. 또한 2019년 이후 정부교육, 주택 토픽에 대한 비중이 2018년 대비 약 50% 감소함을 확인할 수 있었다.

시간의 흐름에 따른 토픽이 차지하는 비중이 급증하는 토픽을 선택하여 그래프화한 결과는 그림 4와 같다. 여가, 케어 산업 토픽의 비중이 시간의 흐름에 따라 급격하게 늘어남을 확인할 수 있었고, 라이프 스타일 토픽은 2018년, 2019년, 2020년까지 비슷한 비중으로 나타났으나 2018년 대비 2021년에는 1.5배 증가한 비중으로 나타남을 확인할 수 있었다. 가장 급격한 비중 변화를 나타내는 토픽은 쇼핑이었는데, 2018년 이후 비중이 급격하게 증가하여



(그림 4) 주요 증가 토픽
(Figure 4) Major increased topics

2021년에 이르러 약 9배 증가한 것으로 나타났다. 또한 정부 사업 토픽 또한 2018년 이후 비중이 줄었다가 2018년 대비 2021년에 이르러 약 3배의 비중 증가를 확인할 수 있었다.

5. 결론 및 시사점

본 연구는 빅데이터 분석을 활용하여 액티브 시니어와 관련된 핵심 키워드와 트렌드를 탐색하기 위해 네이버와 다음의 뉴스, 블로그, 카페 게시글을 분석 대상으로 2018년 1월 1일부터 2021년 6월 31일까지 3년 6개월간 '액티브 시니어'와 관련있는 소셜 미디어 데이터를 수집하여 총 8,740건의 데이터를 수집하였다. 수집된 소셜 미디어 데이터를 대상으로 빈도 분석과 TF-IDF 분석, 토픽 모델링하였고, 주요 결과는 다음과 같다.

첫째, 빈도 분석과 TF-IDF 분석을 통해 알아본 연도별 액티브 시니어와 관련하여 가장 높은 빈도로 언급된 키워드와 중요도가 높은 키워드는 다음과 같았다. 2018년 가장 많이 언급된 상위 3개 키워드는 '건강(69)', '활동(64)', '소비(63)', 이며 중요도가 높은 단어는 '건강(117.8)', '소비(106.2)', '활동(104.2)', '여행(97.8)', '비타민(95.4)' 이다. 2019년 가장 많이 언급된 상위 3개 키워드는 '건강(255)', '사회(216)', '활동(194)'이며 중요도가 높은 단어는 '건강(117.8)', '타운(106.2)', '서비스(332.4)'이다. 2020년 가장 많이 언급된 상위 3개 키워드 '활동(284)', '문화 (227)', '건강(201)''이며, TF-IDF 분석 결과로는 '문화(422.5)', '건강(399.3)', '활동(391.4)'순으로 가장 중요도가 높은 단어로 나타났다. 새롭게 도출된 키워드로는 '콘텐츠', '시대', '사업', '맞춤', '지원', '생활', '노인' 이 있었는데, 이러한 결과를 통해 기업들이 액티브 시니어 맞춤 서비스에 관심을 가지고 있다는 것을 알 수 있었다. 또한 2020년에는 2018년, 2019년까지 빈도수가 가장 높고 가장 문서상 단어 중요도가 높게 나타난 건강이 아닌 활동이 가장 높은 빈도 수로 나타났고, 문화가 가장 단어 중요도가 높은 단어로 나타난 것이 특징적이다. 가장 특징적인 점은 콘텐츠 키워드가 많이 언급된 키워드로 도출되었는데 이는 액티브 시니어 층이 콘텐츠와 관련한 관심이 높아졌다는 것을 확인할 수 있었다. 2021년 높은 빈도수로 언급된 키워드는 '세대(332)', '건강(256)', '활동(238)'이며 TF-IDF 분석 결과로는 '건강(487.3)', '세대(452.5)', '시장(423.4)'이 가장 중요도가 높은 키워드로 나타났다. 은퇴', '노인', '시장', '대표', '참여', '도전', '놀이

터'가 새롭게 높은 빈도수를 나타내는 키워드로 도출되었는데, 이는 2021년부터는 시니어들을 위한 노인 놀이터의 키워드 '노인', '놀이터'의 키워드를 포함하며 '참여', '도전'과 같이 시니어 개인의 활동에 관심이 많아졌으며 '은퇴' 후에 삶에 대한 관심도 많이 생긴 것을 알 수 있다. 또한 빈도수를 살펴보면 2018년에서 2019년에는 약 4배가 증가하였고, 매해 빈도수가 증가하여 2020년에서2021년에는 같은 기간으로 비교해보았을 때 약 2배 이상 증가함을 확인할 수 있었다. 이를 통해 액티브 시니어에 대한 관심도가 급격하게 증가하였음을 확인할 수 있었다.

둘째, 3년 6개월간 액티브 시니어와 관련하여 수집된 데이터를 기반으로 토픽 모델링을 한 결과 10개의 토픽을 도출할 수 있었다. 다양한 스마트 기기를 활발하게 사용하여 콘텐츠를 이용하고 쇼핑을 하면서 모바일 시장에서 새로운 소비자 층인 액티브 시니어의 특징을 보여주는 토픽인 라이프 스타일, 쇼핑, 여가와 연령에 따른 특성인 기존의 시니어 세대와 공통적으로 도출되는 토픽인 건강이 있었고, 시니어를 보호하기 위한 정부차원의 혜택, 정책,사업 등과 관련된 토픽으로는 혜택, 정부 사업, 정부 교육, 케어 산업이 도출되었다. 문서 건수에 따른 토픽 출현율은 토픽 7: 사회/경제, 토픽 5: 정부 교육, 토픽 6: 건강, 토픽 3: 쇼핑, 토픽 9: 실버 주택, 토픽 1: 라이프 스타일, 토픽 10: 여가, 토픽 4: 정부 사업, 토픽 2: 혜택 순서임을 알 수 있었다. 10개 토픽의 시계열 변화를 살펴본 결과, 토픽의 트렌드 변화를 살펴볼 수 있었다. 사회/경제가 매해 모두 가장 비중이 높은 토픽으로 나타났고, 시니어 연령 특성상 건강 토픽은 약 10%의 비중으로 매해 상위 5개의 토픽으로 나타남을 확인할 수 있었다. 액티브 시니어가 기존의 시니어와의 차이점으로 자신의 여가와 라이프 스타일을 즐기며 쇼핑 활동에 주목한다는 특징을 가지는데, 최근 액티브 시니어 계층의 증가와 그들의 활동이 활발해짐과 더불어 여가, 케어 산업, 라이프 스타일, 쇼핑 토픽의 비중이 증가함을 시계열적 변화를 통해 확인할 수 있었다. 이와 관련하여 액티브 시니어와 관련한 정부 사업 토픽의 비중 또한 증가함을 확인할 수 있었다.

본 연구의 결과가 제시하는 학술적 시사점은 다음과 같다. 새로운 소비자 계층으로 나타난 액티브 시니어와 관련된 연도별 키워드 확인과 시계열적 변화 확인을 통해 액티브 시니어 전반에 대한 이해를 높일 수 있었고, 이와 관련한 연구의 기초자료로 활용될 수 있을 것이다. 또한 본 연구의 실무적 시사점은 다음과 같다. 액티브 시니어의 콘텐츠에 대한 관심과 관심도의 증가를 확인할 수 있었고, 이를 통해 액티브 시니어 콘텐츠 시장의 가능성을 제시해주며 액티브 시니어의 콘텐츠 개발에 있어 10개의 토픽이 가이드 라인을 제시해 줄 수 있을 것이다.

하지만 본 연구는 이러한 시사점과 더불어 다음과 같은 한계점을 가지고 있다. 불용어 사전 구축을 통해 정제가 잘 이루어졌다고 하나, 동일 의미 단어를 하나의 단어로 변환하는 과정이 향후 보완되어져야 할 것으로 생각된다.

References

[1] Statistics Korea,"2020 Senior Statistics" 2020. https://kostat.go.kr/portal/korea/kor_nw/1/1/index.board?bmode=read&aSeq=385322

[2] J. Yun, "Korea's Aging and Medical Industry", The World of Electricity, Vol.67, No.1, pp.34-38, 2018. https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07293119

[3] I. Kim, "Necessity for Changing the Aging Perceptions as Enter the Aged Society - Focusing on the concept of Gerotranscendence in the Korean Sundo", Sundo Culture, Vol.30, No.0, pp.373, 2021. https://dx.doi.org/10.35573/JKSC.30.10

[4] M. Kim, J. Won, M. Seo, B. Gang, G. Kim, and Y. Im, "The Socioeconomic problems of an aging society and policy responses", Korea Institute For Health And Social Affairs, 2003. http://repository.kihasa.re.kr/handle/201002/366

[5] D. Han, "A Study on Active Aging Policy in Aged Society", Korean Journal of Gerontological Social Welfare, Vol.64, pp.31-51, 2014. https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07430089

[6] E. Hyun, "The Impact of Active Aging of the Elderly on Life Satisfaction in the Aged Society", The Journal of Humanities and Social Sciences 21, Vol.11, No.4, pp.1491-1506, 2020. https://dx.doi.org/10.22143/HSS21.11.4.105

[7] Y. Park, "Effects of the Social Health of the Aged on their Quality of Life", The Journal of the Korea Contents Association, Vol.13, No.11, pp.726-739, 2013. https://dx.doi.org/10.5392/JKCA.2013.13.11.726

[8] M. Lee and H. Lee, "A Study on the Residential Services of Urban Type Senior Living for Active Senior", 2021 Spring Conference of the Korea Institute of Interior Design, Vol.10, No.1, pp.232-235, 2008. https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02183297

[9] J. Park, Y. Yang, and M. Kim, "The Effect of Active Senior's Career Orientation and Educational Entrepreneurship Satisfaction on Entrepreneurship intention and Entrepreneurship Preparation Behavior", Asia-Pacific Journal of Business Venturing and Entrepreneurship, Vol.15, No.1, pp.285-301, 2020. https://dx.doi.org/10.16972/apjbve.15.1.202002.285

[10] B. Choi and S. Moon, "Analysis of usage decision factors based on the satisfaction of smart seniors using smartphone delivery applications.", Journal of Digital Convergence, Vol.19, No.1, pp.199-209, 2021. https://dx.doi.org/10.14400/JDC.2021.19.1.199

[11] Y. Son and S. Park, "The Influence of Branded Contents-YouTube Creator's Characteristics on Users' Brand Attitude and Purchase Intention", The Korean Journal of Advertising, Vol.32, No.4, pp.89-116, 2021. https://dx.doi.org/10.14377/KJA.2021.5.30.89

[12] Y. Koh, "A Exploratory Study on the Digital Aging Policies as Solutions for a Aging Society", Journal of Digital Convergence, Vol.14, No.1, pp.115-123, 2016. https://dx.doi.org/10.14400/JDC.2016.14.11.115

[13] J. Park and C. Suh, "Analysis of Changes in the Housing Market Using TF-IDF Weight Model", Korea Real Estate Academy Review, Vol.63, No.0, pp.46-58, 2015. https://kiss.kstudy.com/thesis/thesis-view.asp?key=3647174

[14] J. Lee, M. Lee, and J. Kim, "A study on Korean language processing using TF-IDF", Korea Associationof Information Systems, Vol.28, No.3, pp.105-121, 2019. https://dx.doi.org/10.5859/KAIS.2019.28.3.105

[15] J. Lee and J. Rha, "Exploring Consumer Responses to the Cross-Border E-Commerce using Text Mining", Journal of Consumer Studies, Vol.26, No.5, pp.93-124, 2015. https://kiss.kstudy.com/thesis/thesis-view.asp?key=3373334

[16] J. Heo and E. Lee, "Trend Analysis of Fashion Brand Evaluation Using Big Data - Focusing on Gucci Brand -", Journal of the Korean Society of Costume, Vol.69, No.6, pp.38-51, 2019. https://dx.doi.org/10.7233/jksc.2019.69.6.038

[17] K. Han, "A Study on the Consumer's Perception of HiSeoul Fashion Show Using Big Data Analysis", Journal of Fashion Business, Vol.23, No.5, pp.81-95, 2019. https://dx.doi.org/10.12940/jfb.2019.23.5.81

[18] D. Blei, L. Carin, and D. Dunson, "Probabilistic topic models", IEEE signal processing magazine, Vol.27, No.6, pp.55-65, 2010. https://dx.doi.org/10.1109/MSP.2010.938079

[19] H. An and M. Park,"A Study on the User Perception in Fashion Design through Social Media Text-Mining", Journal of the Korean Society of Clothing and Textiles, Vol.41, No.6, pp.1060-1070, 2017. https://dx.doi.org/10.5850/JKSCT.2017.41.6.1060

[20] S. Park, "Analysis of Social Media Contents about Broadcast Media through Topic Modeling", Journal of Information Technology Services, Vol.15, No.2, pp.81-92, 2016. https://dx.doi.org/10.9716/KITS.2016.15.2.081

[21] K. Park and C. Lee, "A Study on the Research Trends for Smart City using Topic Modeling", Journal of Internet Computing and Services, Vol.20, No.3, pp.119-128, 2019. https://dx.doi.org/10.7472/jksii.2019.20.3.119

[22] M. Oh, J. Jung, and M. Lee, "A Review of the Research Trend Analysis on the Silver Consumer and Further Research Direction Suggestions", Journal of Corporation and Innovation, Vol.5, No.3, pp.103-119, 2012. http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02064111

[23] M. Oh, "A Topic Modeling Analysis from the 'Aging' Keyword of DomesticAcademic Research", korean management review, Vol.48, No.2, pp.515-532, 2019. https://dx.doi.org/10.17287/kmr.2019.48.2.515

◐ 저 자 소 개 ◑

이 동 우(Dongwoo Lee)
2019년 조선대학교 경영학부 졸업(학사)
2021년 조선대학교 경영학과, MIS Data Science 졸업(석사)
현재 알티데이터랩 Data Anlytics팀 연구원
관심분야 : Fraud detection, Text Mining, Topic Modeling etc.
E-mail : dw.lee@rtdata.co.kr

김 유 신(Yoosin Kim)
2000년 국민대학교 정보관리학과 졸업(학사)
2009년 국민대학교 경영정보학과 졸업(석사)
2013년 국민대학교 경영정보학과 졸업(박사) & 미국 텍사스 주립대 Post-Doctoral Research Fellow
현재 알티데이터랩 Founder & Chief Data Scientist
관심분야 : Real-Time Big Data Analytics & AI
E-mail : yoosin@rtdata.co.kr

신 은 정(Eunjung Shin)
2013년 이화여자대학교 컴퓨터공학과 졸업(학사)
2017년 연세대학교 의류환경학과 졸업(석사)
2019년 연세대학교 의류환경학과 졸업(박사)
현재 연세대학교 강사
관심분야 : Big Data Analysis, Artificial Intelligence, Consumer Behavior etc.
E-mail : ej_shin@yonsei.ac.kr