

# 빅데이터와 텍스트마이닝을 이용한 부동산시장 동향분석

전해정<sup>1,2</sup>

<sup>1</sup>상명대학교 경영대학원 글로벌부동산학과 조교수

<sup>2</sup>상명부동산연구소 책임연구위원

## Analysis of Real Estate Market Trend Using Text Mining and Big Data

Hae-Jung Chun<sup>1,2</sup>

<sup>1</sup>Department of Global Real Estate, Sangmyung University, Assistant Professor

<sup>2</sup>Sangmyung Institute for Real Estate Study, Research Fellow

요 약 본 연구는 빅데이터 분석방법인 텍스트마이닝을 이용한 부동산시장 동향분석에 관한 연구로 자료는 2016년 8월부터 2017년 8월까지의 포털사이트인 네이버에 게시된 인터넷 뉴스를 통해 수집하였다. TF-IDF 분석결과, 주택, 분양, 가구, 시장, 지역 순으로 빈도가 높게 나타났고 대출, 정부, 대책, 규제 등 정책과 관련된 단어들도 많이 추출되었으며 지역관련 단어는 서울의 출현빈도가 가장 많은 것으로 나타났다. 지역과 관련된 단어 조합은 '서울-강남', '서울-수도권', '강남-재건축', '서울-재건축'의 출현빈도가 많은 것으로 나타나 강남지역 재건축에 대한 사람들의 관심과 기대가 높은 것을 알 수 있다.

주제어 : 부동산시장, 빅데이터, 주택, 규제, 텍스트마이닝

**Abstract** This study is on the trend of real estate market using text mining and big data. The data were collected through internet news posted on Naver from August 2016 to August 2017. As a result of TF-IDF analysis, the frequency was high in the order of housing, sale, household, real estate market, and region. Many words related to policies such as loan, government, countermeasures, and regulations were extracted, and the region - related words appeared the most frequently in Seoul. The combination of the words related to the region showed that the frequencies of 'Seoul - Gangnam', 'Seoul - Metropolitan area', 'Gangnam - reconstruction' and 'Seoul - reconstruction' appeared frequently. It can be seen that the people's interest and expectation about the reconstruction of Gangnam area is high.

**Key Words** : Real Estate Market, Big Data, Housing, Regulation, Text Mining

## 1. 서론

### 1.1 연구의 배경 및 목적

비약적으로 기술이 발전하고 디지털 혁명이 이루어지면서 소셜 미디어(social media)의 등장으로 인해 쏟아지는 지식에서 다양한 의미를 추출해 내는 과정이 중요해지면서 빅데이터(big data)에 대한 관심이 날로 증가하고

있다. 또한 디지털 기술에 의해 모든 것이 서로 연결된 지능적인 사회로 정의되는 4차 산업혁명 사회에서의 빅데이터의 중요성은 좀 더 강조되고 있는 상황이다.

빅데이터는 데이터 기반의 합리적이고 과학적인 미래 예측, 의사결정지원 등을 통한 혁신의 기반으로 이용할 수 있는 기술로서 많은 국가와 국내외 기업은 기술개발과 활용의 선점을 위해 연구와 개발을 하고 있다. 빅데이

\*Corresponding Author : Hae-Jung Chun(hjchun6807@smu.ac.kr)

Received March 4, 2019

Accepted April 20, 2019

Revised April 5, 2019

Published April 28, 2019

터는 기존에 활용할 수 없었던 데이터까지 실시간으로 수집·저장·분석해 상호 의존성, 불확실성과 복잡성이 증가하는 사회변화를 예측하고 문제 해결에 필요한 가치와 정보를 제공할 것으로 기대하고 있다.

빅데이터의 분석기법은 텍스트마이닝(text mining), 데이터마이닝(data mining), 사회네트워크분석(social network analysis), 시스템다이나믹스(system dynamics) 등이 있다. 이 중에서 텍스트마이닝은 많은 정보 중에서 가치를 추출하는 방법이며, 이를 통해 관련된 키워드의 사회적 동향과 이슈를 파악하는데 이용되고 있다. 미국의 경우 구글에서 텍스트마이닝을 이용해 독감트렌드를 예측을 한 것이 질병통제예방센터의 독감예측보다 정확하고 신속하게 분석한 사례도 있다.

부동산학은 다학제 융·복합 학문으로서 서로 복잡하게 상호작용을 하고 있으며 기존의 정형데이터(부동산가격, 거래량, 인구, 소득 등)를 이용해 이를 분석하고 해석하려는 연구들은 그동안 많이 있었다. 정형데이터 뿐 만 아니라 비정형데이터를 포함하고 있는 부동산 빅데이터를 이용해 새로운 방법론으로 부동산시장의 의미 있는 패턴을 찾아내는 것이 시대적으로 필요한 상황이다.

이에 본 연구는 비정형데이터인 부동산 관련 인터넷 뉴스 기사를 중심으로 빅데이터 분석방법인 텍스트마이닝을 이용해 부동산시장의 동향을 분석하고자 한다. 시간적 범위는 2016년 8월부터 2017년 8월까지의 1년간 포털사이트인 네이버에 게시된 뉴스를 통해 478,731건의 자료를 수집해 분석하였다.

본 연구의 구성은 다음과 같다. 2장은 분석방법으로 자료수집, TF-IDF(term frequency-inverse document frequency) 분석과 N-gram 순열에 대해 알아본다. 3장은 분석결과로 데이터의 워드 클라우드(word cloud)를 시각화하고 TF-IDF 분석과 N-gram 순열결과를 비교 분석한다. 이후 4장은 결론으로 연구결과를 요약하고 시사점을 제시하고자 한다.

## 1.2 선행연구 고찰

텍스트마이닝을 국내 주택·부동산 분야에 적용한 연구는 최근에서야 연구가 이루어지고 있으며 거의 전무한 상태이다.

장몽현·김한수(2018)은 텍스트마이닝을 이용해 인터넷 뉴스기사에 나타난 키워드 빈도변화와 주택가격이 어떤 관계성이 있는지를 분석하였다. 분석결과, 가격이 상

승하는 부동산 시장에서는 주택, 매매, 분양, 상승, 강남 등의 키워드가 자주 등장한다고 하였다[1].

경정익·이국철(2016)은 부동산 기사, SNS 등의 데이터로 텍스트마이닝을 이용해 부동산 의사결정에 도움이 되는 감성분석 모형을 도출하였다. 감성분석에 의해 산출한 데이터와 아파트 실거래가 데이터를 비교분석한 결과 상관관계가 높다고 하였다[2].

임준혁 외(2016)은 텍스트마이닝을 이용해 주거지재생과 관련된 학술연구논문과 신문기사를 분석하였다. 분석결과, 학술논문과 신문기사 모두에서 주거지재생의 필요성을 강조하고 있으나 신문기사의 경우는 학술연구논문과는 다른 주거지재생을 강조하고 있다고 하였다. 이는 현실에서 사업성이 높은 재건축, 재개발이 강조되었기 때문이라고 하였다[3].

박주섭 외(2016)는 2013년부터 2015년까지의 신문 기사를 텍스트마이닝을 이용해 해외의 스마트시티 구축 사례를 연구하였다. 연구결과, 플랫폼, 빅데이터, 클라우드, 세서 등이 스마트시티와 연관성이 높다고 하였다[4].

박종영·서충원(2015)은 부동산 뉴스 기사를 취합해 텍스트마이닝을 이용해 주택시장의 변화를 실증분석하였다. 분석결과, 양도세 중과 폐지와 취득세 감면이 침체된 주택시장을 회복하는데 기여하였고 바닥과 상승이라는 단어를 많은 기사들이 다루면서 주택시장이 상승분위기로 바뀌고 있다고 하였다[5].

최해옥(2015)은 소셜미디어의 지적측량 이슈 네트워크 분석을 텍스트마이닝을 이용해 분석하였다. 분석결과, 지적측량 행정과 처리에 관련된 사항은 국가 행정기관과 관련성이 깊고, 이에 대한 시행은 지자체와 연결이 되어 있다고 하였다[6].

이상훈(2014)은 현대 일본의 도시형 주택에 대한 설계자의 언설을 텍스트마이닝을 이용해 특징적인 면을 분석하였다. 분석결과, 주동수별 그룹마다 특징어의 출연경향이 차이가 존재하였으며, 주동수가 많아짐에 따라 주변 환경 고려나 외부공간 등과 관련성이 높은 키워드가 상위순위를 차지한다고 하였다[7].

김유신 외(2012)는 뉴스와 주가사이의 관계를 텍스트마이닝, 오피니언마이닝과 같은 빅데이터 분석을 이용해 투자 의사결정모형을 분석하였다. 분석결과, 전망, 시황, 해외뉴스가 주가지수변동을 가장 잘 반영한다고 하였다[8].

전해정(2017)은 빅데이터를 이용해 데이터마이닝으로 주택가격 결정요인을 분석하였다. 분석결과, 신경망분석

의 주택가격 예측력이 가장 높게 나타났다고 하였다[9].

채윤식·이상훈(2018)은 부산 및 기타지자체의 보고서로 텍스트마이닝을 이용해 스마트도시계획 수립에 필요한 전략을 도출하였다. 분석결과, ICT시스템과 도시통합 운영관리에 필요한 내용이 많이 있다고 하였다[10].

이준용 외(2007)은 데이터 마이닝을 주택가격 예측에 관한 연구를 하였다. 분석결과, 신경망분석이 결과의 해석이 어려우나 기존 통계기법보다 예측력이 높다고 하였다[11].

본 연구의 차별성은 수도권 지역의 주택가격이 상승하던 시기였던 2016-2017년까지 비정형데이터인 부동산 관련 뉴스기사를 수집해 빅데이터 분석기법인 텍스트마이닝을 적용해 부동산시장의 동향을 실증적으로 분석함에 있다.

## 2. 분석방법론

본 연구의 부동산시장 동향에 대한 인터넷 뉴스를 대상으로 한 텍스트마이닝을 이용한 연구절차는 Fig. 1.과 같다. 온라인 기사에서 수집된 비정형데이터에서 높은 단어빈도수를 가지는 단어들을 추출하였으며 조사, 접속사, 문장의 의미, 의미 없는 숫자, 특수 문자 등은 제거하여 데이터 전처리 과정(preprocessing process)을 거쳤다.

텍스트마이닝이란, 비정형화된 대규모 문서에서 필요한 정보를 찾아내는 방법으로 자연어 처리기술을 바탕으로 하고 있으며 컴퓨터 언어학과 통계학, 정보학, 대용량 데이터에서 규칙이나 패턴 등을 찾아내는 데이터마이닝 등이 결합된 연구 분야 이다[12].

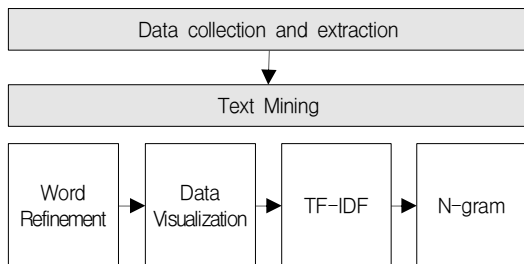


Fig. 1. Flow Chart of Analysis

연구의 순서와 방법은 아래와 같다. 첫째, 부동산 관련 대상을 알아보고 검색 키워드를 선정한다. 그리고 네이버에서 나타난 인터넷 뉴스를 R통계패키지를 이용해 수

집한다. 둘째, KoNLP 텍스트마이닝 패키지를 활용하여 형태소 분석 및 자연어 처리를 실시한다. 텍스트에 TF-IDF 분석으로 가중치를 산출하고 N-gram 순열을 이용하여 단어-단어 구조로 연결하여 추출된 단어가 표현하고자 하는 의미를 명확하게 파악한다.

### 2.1 자료수집

본 연구의 데이터 분석을 위해 자료는 포털사이트 네이버에 게시된 부동산 관련 인터넷 뉴스로 선정하였다. 자료 분석을 위해 주제 분야는 경제 분야를 선택하였으며 불필요한 기사들을 키워드를 통해 필터링하였다. 필터링 단어들은 아파트, 부동산, 분양, 주택 등으로 해당 단어가 포함된 기사들만 데이터 수집대상으로 간주하였으며 반면 분양정보와 같은 기사와 중복된 내용이 담긴 기사들은 수집대상에서 제외하였다.

### 2.2 TF-IDF 분석

TF-IDF 분석은 단어가 한 문장이나 문서에서 차지하는 중요도를 평가하는 통계방법이다. 단어의 중요성은 한 문장에서의 출현횟수에 정비례하며, 단어가 포함되는 전체 문서수와 반비례 한다. TF-IDF 가중치는 여러 가지 형태로 사용자와 검색자료 간의 상관평도를 평가하는 방법으로 인터넷검색엔진 분야에서 많이 응용되고 있다.

TF-IDF는 구와 절 수준에서의 분석에서 많이 이루어지고 있으며, 문서d에서 단어w에 대한 weight값을 산출하는 방식을 아래의 수식으로 표현될 수 있다[13,14].

$$TFIDF(w,d) = TF(w,d) \times \log\left(\frac{N}{DF(w)}\right) \quad (식1)$$

TF(w,d) : 문서d에 단어w가 나타나는 횟수  
 DF(w) : 단어w가 들어가는 문서의 총수  
 N : 전체 문서의 총수

log를 취한 이유는 문서의 수집량이 많아질 때 N 값이 커지는 것을 조절하기 위함이다. 이러한 TF-IDF 방식을 이용하면 하나의 문서 중에서 가장 weight값이 높은 단어가 그 문서에 키워드로 채택된다[15].

### 2.3 N-gram 순열

N-gram은 문자열에서 N개의 연속된 요소를 추출하는 방법이며 확률학, 통신이론, 컴퓨터언어학 및 데이터 압축 영역에서 널리 사용됩니다. N-gram의 두 가지 이

점은 단순성과 확장성입니다. 빈도분석을 통해 나온 단어들은 원래의 자기 자리에 있던 순서를 완전히 무시하고 처리되므로 단어 자체가 가지는 의미를 제대로 표현하지 못할 수 있다. 이를 보완하기 위해 문서에서 인접한 단어들을 쌍으로 묶어 표현하는 N-순열 기법을 적용한다[5]. n=1의 N-gram은 일반적으로 “Unigram”이라고 하며 n=2는 “Bigram”(또는“Digram”)이며, n=3은 “Trigram”라고 한다.

### 3. 분석결과

본 연구에서 이용된 자료는 R 프로그램을 활용하여 2016년 8월부터 2017년 8월까지의 1년간 포털사이트인 네이버[16]에 게시된 인터넷 뉴스를 통해 수집하였다. 분석기간에 경제 분야에 해당되는 인터넷 기사는 총 478,731건이며, 부동산 관련 기사는 총 11,293건인 것으로 나타났다. Table 1은 각 언론사별 기사빈도를 보여주고 있다.

Fig. 2.를 보면 부동산 관련 인터넷 기사 수집 추이는 2016년 10월에서 2017년 4월까지 기사 양이 감소하기 시작하여 2017년 6월부터 다시 증가하는 것으로 나타났다. 아파트·주택에 대하여 게시된 기사가 증가하는 것으로 나타난 시점은 여름철인 2017년 6월1일부터 6월30일까지로 1167건이었으며, 상대적으로 감소하는 경향을 나타낸 시점은 2017년 4월1일에서 4월30일까지로 400건이다.

본격적으로 분석하기 전에 어떠한 단어가 많이 노출되고, 서로 어떤 관계성을 가지고 있는지 살펴보고자 원자료에 대해 시각화를 시도해 보았다. 먼저 전체적으로 어떤 단어들이 보여지는지 확인하고자 워드 클라우드(word cloud)를 그렸다. Fig. 3.에 의하면 주택, 분양, 가구, 시장, 서울과 같은 단어들이 많이 언급된 것으로 나타났다. 워드 클라우드는 단어의 출현 빈도가 많을수록 단어가 크게 나타나며, 단어의 출현 빈도가 비슷하면 같은 색, 같은 크기로 나타나게 된다. 따라서 대략적으로 어떤 단어들이 많이 나타나는지 특별한 단어는 어떤 것이 있는지 워드 클라우드를 통해 쉽게 확인할 수 있다[5].

Table 1. Frequency of Articles by Media

Ranking	Division	Number	%	Ranking	Division	Number	%
1	Newsis	1389	12.30	30	Asia Business	69	0.61
2	Yonhap News	1375	12.17	31	Hankook Business	54	0.48
3	Dong-A Ilbo	840	7.44	32	Financial News	53	0.47
4	Seoul Shinmun	744	6.59	33	Chosun Ilbo	48	0.43
5	Digital Times	692	6.13	34	Maeil Shinmun	47	0.42
6	KBS	671	5.94	35	The Daily Sports	40	0.35
7	Jungang Ilbo	628	5.56	36	Hankook Business TV	38	0.34
8	Busan Ilbo	492	4.36	37	Kookmin Ilbo	35	0.31
9	Jose Ilbo	456	4.04	38	Ohmynews	33	0.29
10	Segye Ilbo	403	3.57	39	Economist	33	0.29
11	MBN	323	2.86	40	Electronic Times	32	0.28
12	YTN	318	2.82	41	Sports Kyunghyang	31	0.27
13	Maeil Business Newspaper	253	2.24	42	Chosung Biz	31	0.27
14	Yonhap News TV	216	1.91	43	SBS CNBC	28	0.25
15	Edaily	167	1.48	44	Pressian	24	0.21
16	Hankook Ilbo	166	1.47	45	Kangwon Ilbo	13	0.12
17	MBC News	166	1.47	46	Money S	13	0.12
18	Sports Chosun	148	1.31	47	Sports Donga	13	0.12
19	News 1	135	1.20	48	Women Shinmun	9	0.08
20	Munhwa Ilbo	134	1.19	49	inews24	7	0.06
21	Money Today	130	1.15	50	Hankyung Business	5	0.04
22	SBS News	115	1.02	51	TV Chosun	5	0.04
23	Kyunghyang Shinmun	109	0.97	52	Jungang SUNDAY	4	0.04
24	Seoulbusiness	105	0.93	53	Channel A	4	0.04
25	Hankyoreh Shinmun	98	0.87	54	Weekly Kyunghyang	3	0.03
26	Nocut News	95	0.84	55	Dailian	2	0.02
27	Herald Business	89	0.79	56	Herald POP	2	0.02
28	Maeil Business Economy	81	0.72	57	Media Today	1	0.01
29	JTBC	77	0.68	58	Weely Chosun	1	0.01

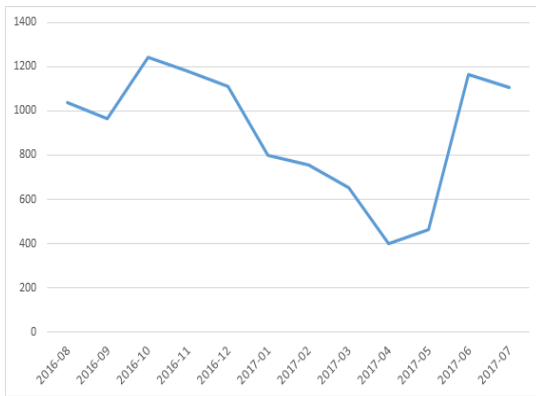


Fig. 2. Trend of Real Estate Articles

출현빈도 높은 상위 20개의 단어를 대상으로 TF-IDF 분석을 활용하여 단어의 가중치를 도출하고 부동산·주택과의 연관성을 분석하였다.

TF-IDF 분석은 용어의 가중치를 부여하여 계산하는 방법으로 구와 절 수준에서의 분석이 이루어지며 개별적인 문서에서 단어에 대한 중요도를 표현하는 방법으로 텍스트마이닝에서 많이 사용되는 기법이다[12].



Fig. 3. Analysis of Word Cloud

Table 2의 TF-IDF 분석결과, 주택이 가장 높은 빈도가 나타났으며 TF-IDF 지수는 101991.60이었다. 다음으로 빈도가 높았던 분양의 TF-IDF 지수는 88191.95로 나타났으며 가구의 TF-IDF 지수는 81706.73로 나타났다. 추출된 단어를 살펴보면 주택(22805회), 분양(18461회), 가구(16560회), 시장(13723회), 지역(11067회) 등과 같이

부동산과 밀접한 관련된 명사들이 가장 많이 추출되었을 알 수 있다. 그리고 대출(9204회), 정부(9015회), 대책(7248회), 규제(5746회) 등 정책과 관련된 단어들도 많이 추출되었으며 지역관련 단어는 서울(14340회)의 출현빈도가 가장 많은 것으로 나타났다.

앞서 원 자료에서 추출해온 단어들은 순서와 단어들의 연결 구조를 알 수 없기 때문에 단어 자체가 문장 중에서 가지는 의미를 제대로 보여주지 못한다. 따라서 본 연구에서는 N-gram기법을 통해 문서에서 인접한 단어들을 묶어서 추출된 단어가 표현하고자 하는 의미를 파악하였다.

Table 3은 빈도수가 높은 주요단어가 어떤 의미로 다른 단어와 연결되어 있는지를 결합 빈도수로 보여준다. ‘주택-담보대출’, ‘전매-제한’, ‘대출-규제’, ‘분양-물량’, ‘분양권-전매’, ‘정부-대책’, ‘청약-경쟁률’, ‘과열-지구’, ‘대책-발표’, ‘가격-상승’, ‘입주-물량’, ‘규제-강화’, ‘입주-예정’, ‘서울-강남’순으로 단어 간 조합 빈도를 보여준다.

Table 2. TF-IDF Weights

Ranking	Content	Frequency	TF-IDF
1	Housing	22805	101991.60
2	Housing Sale	18461	88191.95
3	Household	16560	81706.73
4	Seoul	14340	73731.10
5	Market	13723	71429.42
6	District	11067	61039.15
7	Mortgage	9204	53211.56
8	Government	9015	52388.73
9	Housing Complex	8950	52104.43
10	Expectation	8595	50539.58
11	Price	8315	49290.45
12	Supply	7667	46346.63
13	Measurement	7248	44401.46
14	Housing Offer	6696	41785.13
15	Investment	6302	39877.81
16	Area	6147	39117.85
17	Price Rise	5991	38347.28
18	Business	5950	38143.80
19	Regulation	5746	37125.22
20	Scale	5340	35066.56

Note : The top 20 texts were presented based on the frequencies among the words with the frequency of occurrence of 0.1% or more.

출현빈도가 높은 단어조합을 살펴보면 ‘전매-제한’, ‘대출-규제’, ‘정부-대책’, ‘대책-발표’, ‘금리-인상’과 같은 정책관련 내용들이 많은 것을 알 수 있으며, ‘과열-지구’, ‘규제-강화’, ‘시장-과열’과 같은 우려를 담은 내용들의 출현빈도도 높은 것으로 알 수 있다. 지역과 관련된 단어

조합은 ‘서울-강남’, ‘서울-수도권’, ‘강남-재건축’, ‘서울-재건축’의 출현빈도가 많은 것으로 나타나 서울과 수도권, 특히 강남지역 재건축에 대한 관심과 기대가 많다는 것을 알 수 있다. 이는 본 연구의 분석기간(2016.8-2017.8)을 고려하면 박근혜 정부(2013.2-2017.3)가 주택시장 활성화 종합대책(4.1)에서 DTI와 LTV 규제를 완화하고 2014년 규제합리화를 통한 주택시장 활력회복 및 서민 주거안정 강화방안(9.1)에서 재건축 연한을 준공 후 30년으로 완화해 이후 강남을 중심으로 한 서울과 수도권의 주택가격이 다른 지역에 비해 급상승하였다. 이에 주택시장의 이상과열 현상이 있다고 판단한 문재인 정부가 들어서면서 금융규제 강화, 재건축 규제 연한 강화, 전매 제한과 투기지역 지정 등 다양한 논의가 이루어졌고 실

제 본 연구의 분석기간 이후인 2017년 실수요 보호와 단기 투자수요 억제를 통한 주택시장 안정화 방안(8.2)과 가계부채 종합대책(10.24)에서 다양한 규제강화 정책이 수립 집행되었다.

#### 4. 결론

본 연구는 텍스트마이닝을 이용한 부동산시장 동향분석에 관한 연구로 자료는 2016년 8월부터 2017년 8월까지의 1년간 포털사이트인 네이버에 게시된 인터넷 뉴스를 통해 수집하였다. 분석기간에 경제 분야에 해당되는 인터넷 기사는 총 478,731건이며, 부동산 관련 기사는 총 11,293건인 것으로 나타났다.

부동산 관련 인터넷 기사 수집 추이는 2016년 10월에서 2017년 4월까지 기사 양이 감소하기 시작하여 2017년 6월부터 다시 증가하는 것으로 나타났다. 아파트·주택에 대하여 게시된 기사를 통해 노출이 증가하는 것으로 나타난 시기는 여름철인 6월이었고, 상대적으로 감소하는 경향을 나타난 시기는 4월인 것으로 나타났다.

TF-IDF 분석결과, 주택(22805회), 분양(18461회), 가구(16560회), 시장(13723회), 지역(11067회) 순으로 빈도가 높게 나타났다. 그리고 대출(9204회), 정부(9015회), 대책(7248회), 규제(5746회) 등 정책과 관련된 단어들도 많이 추출되었으며 지역관련 단어는 서울(14340회)의 출현빈도가 가장 많은 것으로 나타났다.

N-gram 분석결과, 출현빈도가 높은 단어조합을 살펴보면 ‘전매-제한’, ‘대출-규제’, ‘정부-대책’, ‘대책-발표’, ‘금리-인상’과 같은 정책관련 내용들이 많은 것을 알 수 있으며, ‘과열-지구’, ‘규제-강화’, ‘시장-과열’과 같은 우려를 담은 내용들의 출현빈도도 높은 것으로 알 수 있다. 지역과 관련된 단어 조합은 ‘서울-강남’, ‘서울-수도권’, ‘강남-재건축’, ‘서울-재건축’의 출현빈도가 많은 것으로 나타나 서울과 수도권, 특히 강남지역 재건축에 대한 기대와 관심이 많은 것을 알 수 있다.

이는 박근혜 정부에서 주택시장 활성화를 시키고자 DTI와 LTV 규제를 완화하고 재건축 연한을 완화해 강남을 중심으로 한 서울과 수도권의 주택가격이 상승하였고 이후 문재인 정부는 주택시장이 과열되어 있다고 판단해 각종 규제정책에 대한 논의가 이루어졌기 때문인 것으로 판단된다.

Table 3. Key Word N-gram

Content	Frequency	Probability
Housing-Mortgage	2883	0.57%
Resale-Regulation	1542	0.30%
Mortgage-Regulation	1419	0.28%
Housing Sale - Housing Volume	1415	0.28%
Housing Rights - Resale	1291	0.26%
Government-Measurement	1260	0.25%
Housing Sale - Competition Rate	1246	0.25%
Overheating-Region	1170	0.23%
Measurement-Announcement	1151	0.23%
Price-Rise	1147	0.23%
Moving-Housing Volume	1099	0.22%
Regulation-Reinforcement	1062	0.21%
Moving-Expectation	1030	0.20%
Seoul - Gangnam	1010	0.20%
interest rate - Rise	995	0.20%
Housing Sale-Expectation	988	0.20%
Market-Overheating	851	0.17%
District-Appointed	826	0.16%
Housing Volume-Household	797	0.16%
New-Housing Sale	783	0.15%
Household-Supply	782	0.15%
Housing-Supply	779	0.15%
Gangnam-Reconstruction	699	0.14%
Housing Sale - Household	685	0.14%
Household-Housing	683	0.13%
Guarantee-Mortgage	671	0.13%
Market-Stability	639	0.13%
Seoul-Reconstruction	630	0.12%
Seoul-Metropolitan	616	0.12%
Government-Regulation	598	0.12%
Housing Price-Drop	570	0.11%
Government-Seoul	567	0.11%
Reconstruction-District	565	0.11%
Mortgage-Housing	539	0.11%

Note : The probability of appearance by N-gram is 0.1% or more.

본 연구결과에 따르는 정책적 시사점은 빅데이터를 이용해 부동산시장의 동향을 파악하면 부동산시장의 움직임을 신속하게 포착할 수 있으므로 정부정책당국자는 부동산 빅데이터와 관련된 기술개발과 이를 정책적으로 응용할 수 있는 다양한 지원책을 수립·집행해야 한다. 이를 통해 부동산정보 선진화에 기여할 수 있고 부동산 시장을 좀 더 체계적이고 신속하게 알 수 있을 것으로 기대된다.

인터넷 뉴스 기사 외 다른 비정형 데이터를 이용하고 텍스트마이닝 외에 인공지능(AI)을 적용한 방법을 적용해 연구를 확대하는 것은 이후 연구과제로 남긴다.

## REFERENCES

- [1] M. H. Jang & H. S. Kim. (2018). A Study on Text Mining and Housing Market, *Proceeding of Autumn Annual Conference of KHA*, 30(2), 403-404.
- [2] J. I. Kyung & K. C. Lee. (2016). Development of Sentiment Analysis of Real Estate Big Data by Using Textmining, *Housing Studies Review*, 24(4), 115-136.
- [3] J. H. Lim, J. D. Yoon & J. M. Choi. (2016). Analysis of trends related to the regeneration of residential area using textmining, *Proceeding of Autumn Annual Conference of KHA*, 28(2), 197-202.
- [4] J. S. Park, S. G. Hong & N. R. Kim. (2016). A Development Plan for Co-creation-based Smart City through the Trend Analysis of Internet of Things, *Journal of the Korea Industrial Information Systems Research*, 21(4), 67-78.
- [5] J. Y. Park & C. W. Seo. (2015). Analysis of Changes in the Housing Market Using TF-IDF Weight Model, *Korea Real Estate Academic Review*, 64, 46-58.
- [6] H. O. Choi. (2015). Analysis of cadastral issues through text mining analysis of social media, *Journal of the Korean Urban Management Association*, 23(3), 147-161.
- [7] S. H. Lee. (2014). A Study on the Characteristics of Small-scale Urban Housing Planning in Modern Japan, *Journal of the Architectural Institute of Korea Planning & Design*, 30(9), 129-136.
- [8] Y. S. Kim, N. G. Kim & S. R. Jeong. (2012). Stock-Index Invest Model Using News Big Data Opinion Mining, *Journal of Intelligent Information System*, 18(2), 143-156.
- [9] H. J. Chun. (2017). A Study on the Determinants of Housing Price Using Data Mining. *Journal of the Residential Environment Institute of Korea*, 15(3), 35-46.
- [10] Y. S. Chae & S. H. Lee. (2019). Identification of Strategic Fields for Developing Smart City in Busan Using Text Mining, *Journal of Digital Convergence*, 16(11), 1-15.
- [11] J. Y. Lee, M. H. Choi & S. Y. Lee. (2007). A Study on the Forecasting Model of Apartment Price Based on Data Mining, *Journal of Korea Planners Association*, 42(4), 135-148.
- [12] H. K. Park. (2015). *Patent Trend Analysis for Internet of Things Technologies of Korea and U.S.A Using Text Mining Method*, Master Dissertation, Yonsei University, Seoul.
- [13] G. Salton & C. Buckley. C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), 513-523.
- [14] J. Han, J. Pei & M. Kamber (2017). *Data mining: concepts and techniques*, Elsevier.
- [15] Y. S. Shim & H. B. Kim. (2016). A Study of Destination Image and Measurement Using Text Mining, *Journal of Tourism Sciences*, 40(7), 221-245.
- [16] Naver [www.naver.com](http://www.naver.com)

전 해 정 (Chun, Hae Jung)

[정회원]



- 2003년 8월 : 연세대학교 생활과학(이학사, 경영학부전공)
- 2008년 2월 : 연세대학교 경제학(경제학석사)
- 2012년 8월 : 중앙대학교 도시계획부동산학과(도시및지역계획학

박사)

- 2019년 3월 ~ 현재 : 상명대학교 경영대학원 글로벌부동산학과 조교수
- 관심분야 : 주택, 부동산, 계량경제모형
- E-Mail : [hjchun6807@smu.ac.kr](mailto:hjchun6807@smu.ac.kr)