

# 기술 키워드 검색량 기반 주가변동성 예측 모형 비교 연구

조유정

경희대학교 빅데이터응용학과  
yujung@khu.ac.kr

손권상

경희대학교 경영학과  
miroo1215@khu.ac.kr

권오병

경희대학교 경영학과  
obkwon@khu.ac.kr

**Abstract** - 최근 주식의 수익률과 거래량을 설명하는 주요 요인으로 투자자의 관심과 주식 관련 정보 전파의 영향력이 부각되고 있다. 본 연구의 목적은 제4 차산업혁명 관련 기술 키워드의 인터넷 검색량을 투자자의 관심 척도로 사용하여, 기업의 주가 변동성을 예측하는 모형들의 예측 성능을 비교하고자 한다. 이를 위해 본 연구에서는 인공지능, IoT, 로봇, 빅데이터 등 4 차산업혁명 관련 기술 키워드에 대한 일일 검색량 데이터와 코스닥 상장 기업의 주가 정보를 활용하였다.

**Key Terms** - 기술 키워드, 주가변동성, 주가 예측 모형

이 논문 또는 저서는 2017 년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A3A2066740)

## I. 서론

인터넷 보급의 확산과 정보통신 기술의 발달로 인해 정보검색의 결과를 기반으로 의사결정을 하는 일은 일상화되고 있으며, 특히 인터넷 사용자 검색 엔진을 통해 특정 정보를 검색하는 행위는 해당 키워드에 대한 관심을 표출하는 행위이다. 특정 키워드의 인터넷 검색은 의사결정을 위한 정보검색 단계에 해당하기 때문에 인터넷 검색량은 대중의 관심을 측정하는 척도로 다양하게 이용되어 왔다. 가장 대표적인 예로는 감기 증상과 관련된 특정 키워드의 검색량 데이터를 활용한 독감 유행 예측(Ginsberg et al., 2009), 검색 트렌드 분석을 이용한 자살 위험 예측(김병철, 2015), 구글 트렌드를 이용한 시장 분석 및 예측(이건창&윤성욱, 2018) 등 광범위한 분야에서 인터넷 검색량을 활용하고 있다.

특히 최근부터 주가시장 예측을 위한 정보로써 인터넷 검색, 소셜 네트워크 서비스, 온라인 커뮤니티 등을 활용하는 경우가 많아지면서, 특정 산업과 기술에 대한 키워드 검색량을 통해 주가의 움직임을 분석하려는 연구가 시도되고 있다. 일례로 특정 주식에 대한 검색량을 바탕으로 투자자의 관심을 측정하고 주가변동성 간의 관계 분석을 통해

유의성을 입증하는 연구가 있었으며(Da et al., 2011; Bordino et al., 2012), 국내에서도 인터넷 검색량과 코스피 및 코스닥 상장기업의 주가변동성 간 상관관계를 실증적으로 분석하기 위한 시도가 있었다(이병욱 외, 2018).

그러나 인터넷 검색량과 주가변동성의 상관관계가 유의한 것으로 다수의 선행연구를 통해 입증되었음에도 불구하고, 대부분의 연구는 금융 시계열 데이터 예측의 전통적인 방법론인 Auto-Regressive Moving Average(ARMA), AutoRegressive Integrated Moving Average(ARIMA), Vector Auto-Regression(VAR) 등과 같은 계량적 방법을 활용하였다(Box et al., 2015). 전통적인 계량 모형의 경우 통계적 추론을 통해 변수 간 관계를 설명하고 평가하는데 용이하지만, 선형모형 구조를 가정하기 때문에 비선형 움직임을 포착할 수 없는 단점이 있다.

한편, 최근 주가 변동성 예측에 대한 계량적 기법의 한계를 보완할 수 있는 기계학습 기반의 예측 모형 연구가 시도되고 있으며, 향상된 예측 성능을 보여주는 결과가 발표되고 있다(Rather et al., 2015; 주일택&최승호, 2018). 그러나 일부 연구에서는 시계열 데이터의 특성이 선형에 가까울 경우 VAR 모형의 예측력이 더 높다는 결과가 제시되는 등(Ouhame & Hadi, 2019) 연구성과의 비밀관성을 보이고 있다.

따라서 본 연구에서는 제 4 차산업혁명 관련 기술의 기술 성숙도에 따른 키워드 검색량 변화를 기준으로 기업의 주가 변동성 예측을 위한 기존 모형들의 예측 성능을 비교하고, 그 결과를 토대로 예측 모형 선정에 대한 시사점을 도출하고자 한다. 특히 아이디어 단계에서는 해당 기술에 대한 검색량이 비교적 일정하고, 개발 및 실용화 초기 단계에서는 기술에 대한 검색량의 변동이 심하다가, 성장 단계에서는 해당 기술이 분화되어 가면서 세부기술의 검색량이 증가하는 대신 집합명사로서의 기술명칭은 다시 검색량이 일정해질 것을 예상하여, 검색량이 기술의 성숙을,

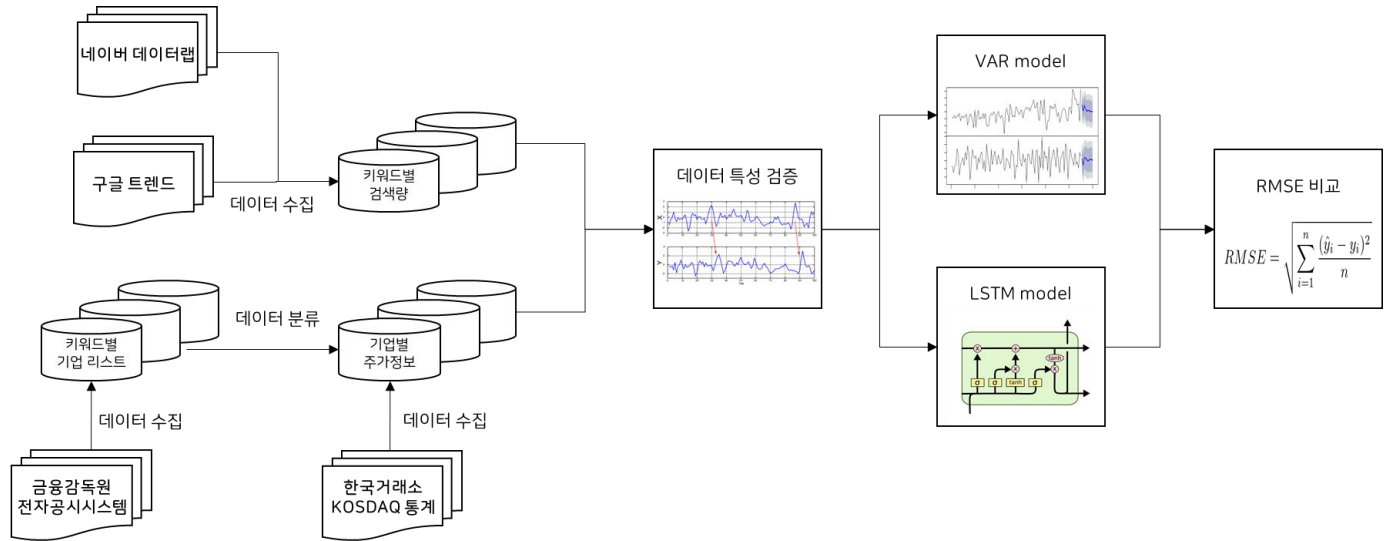
그리고 해당 기술을 사업에 활용하는 기업의 성과에 영향을 줄 것으로 예상된다.

## II. 방법

제 4차 산업혁명 관련 기술 키워드를 선정한 후 해당 키워드를 사업내용에 포함하는 기업 리스트를 확보하였다. 기업의 주가 관련 데이터는 한국거래소(KOSDAQ)에서 확보하였다. 수집한 데이터는 키워드 검색량과 기업 주가 거래량 간

인과성 검증 및 선형성 검증을 위해 데이터 셋으로 구축하였다. 키워드 검색량 데이터는 구글 트렌드와 네이버 데이터랩에서 2016년 1월 1일부터 2019년 12월 31일까지의 일별 데이터를 수집하였다.

다음으로 수집된 기술 키워드 검색량과 기업의 주가정보에 대한 데이터 특성 검증을 실시하며, 이를 바탕으로 예측 모델을 구축한다. 전체적인 연구 프로세스는 <그림 1>과 같다.



<그림 1> 연구 프로세스

## III. 참고문헌

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data. Nature", Vol.457, No.7232, 1012-1014.

김병철, "인터넷 검색 트렌드 분석을 이용한 자살 위험 예측에 관한 연구," 커뮤니케이션학 연구, 제 23권, 제 2호, (2015), 99-120.

이건창, 윤성욱, "구글트렌드를 이용한 스마트 스피커 시장 분석과 예측," 정보과학회 컴퓨팅의 실제 논문지, (2018), 제 24권, 제 11호, 596-602.

Da, Z., J. Engelberg, and P. Gao, "In search of attention", The Journal of Finance, Vol.66, No.5(2011), 1461-1499.

Bordino, I., S. Battiston, G. Caldarelli, M. Cristelli, A. Ukkonen, and I. Weber, "Web search queries can predict stock market volumes", PloS one, Vol.7, No.7(2012), 1-17.

이병욱, 김준호, 우종필, "구글 트렌드를 이용한 기업의 매출액 및 주가 예측",

예술인문사회융합멀티미디어논문지, (2018), 제 8권, 제 10호, 491~501

Box, G.E., G.M. Jenkins, G.C. Reinsel and G.M. Ljung, "Time series analysis: forecasting and control", John Wiley & Sons, 2015

Rather, A.M., A. Agarwal, and V.N. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns", Expert Systems with Applications, Vol.42, No.6(2015), 3234-3241.

주일택, 최승호, "양방향 LSTM 순환신경망 기반 주가예측모델,"

한국정보전자통신기술학회논문지, (2018), 제 11권, 제 2호, 204~208.

Ouhame, S. and Y. Hadi, "Multivariate workload prediction using Vector Autoregressive and Stacked LSTM models", In Proceedings of the New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Society, (2019), 1~7