

인공지능의 사회적 수용도에 따른 키워드 검색량 기반 주가예측모형 비교연구*

조유정**

경희대학교 빅데이터응용학과
(yujung251@khu.ac.kr)

손권상**

경희대학교 경영학과
(miroo1215@khu.ac.kr)

권오병

경희대학교 경영학과
(obkwon@khu.ac.kr)

최근 주식의 수익률과 거래량을 설명하는 주요 요인으로서 투자자의 관심도와 주식 관련 정보 전파의 영향력이 부각되고 있다. 또한 인공지능과 같은 혁신 신기술을 개발보급하거나 활용하려는 기업의 경우 거시환경 및 시장 불확실성 때문에 기업의 미래 주식 수익률과 주식 변동성을 예측하기 어렵다는 문제를 가지고 있다. 이는 인공지능 활성화의 장애요인으로 인식되고 있다. 따라서 본 연구의 목적은 인공지능 관련 기술 키워드의 인터넷 검색량을 투자자의 관심 척도로 사용하여, 기업의 주가 변동성을 예측하는 기계학습 모형을 제안하는 것이다. 이를 위해 심층신경망 LSTM(Long Short-Term Memory)과 벡터자기회귀(Vector Autoregression)를 통해 주식시장을 예측하고, 기술의 사회적 수용 단계에 따라 키워드 검색량을 활용한 주가예측 성능 비교를 통해 기업의 투자수익 예측이나 투자자들의 투자전략 의사결정을 지원하는 주가예측 모형을 구축하였다. 또한 인공지능 기술의 세부 하위 기술에 대한 분석도 실시하여 기술 수용 단계에 따른 세부 기술 키워드 검색량의 변화를 살펴보고 세부기술에 대한 관심도가 주식시장 예측에 미치는 영향을 살펴보았다. 이를 위해 본 연구에서는 인공지능, 딥러닝, 머신러닝 키워드를 선정하여, 2015년 1월 1일부터 2019년 12월 31일까지 5년간의 인터넷 주별 검색량 데이터와 코스닥 상장 기업의 주가 및 거래량 데이터를 수집하여 분석에 활용하였다. 분석 결과 인공지능 기술에 대한 키워드 검색량은 사회적 수용 단계가 진행될수록 증가하는 것으로 나타났고, 기술 키워드를 기반으로 주가예측을 하였을 경우 인식(Awareness)단계에서 가장 높은 정확도를 보였으며, 키워드별로 가장 좋은 예측 성능을 보이는 수용 단계가 다르게 나타남을 확인하였다. 따라서 기술 키워드를 활용한 주가 예측 모델 구축을 위해서는 해당 기술의 하위 기술 분류를 고려할 필요가 있다. 본 연구의 결과는 혁신기술을 기반으로 기업의 투자수익률을 예측하기 위해서는 기술에 대한 대중의 관심이 급증하는 인식 단계를 포착하는 것이 중요하다는 점을 시사한다. 또한 최근 금융권에서 선보이고 있는 빅데이터 기반 로보어드바이저(Robo-advisor) 등 투자 의사 결정 지원 시스템 개발 시 기술의 사회적 수용도를 세분화하여 키워드 검색량 변화를 통해 예측 모델의 정확도를 개선할 수 있다는 점을 시사하고 있다.

주제어 : 인공지능, 사회적 수용도, 주가예측, 다변량 시계열, 키워드 검색량, LSTM, VAR

논문접수일 : 2020년 12월 30일 논문수정일 : 2021년 3월 2일 게재확정일 : 2021년 3월 8일

원고유형 : 일반논문 교신저자 : 권오병

1. 서론

제4차 산업혁명으로 기업 환경의 패러다임이 전환되고 있는 상황에서 제4차 산업혁명 핵심기

술인 인공지능의 변혁은 다양한 영역에서 폭넓게 이루어질 것으로 예상되고 있다(Boyd and Holton, 2018). 특히, 인공지능은 다른 기술들과 융합되며 기업의 디지털 전환(digital transformation)을 가속

* This research is supported by Kyung Hee University's scholarship.

** These authors contributed equally to this work

화하고, 기업은 이를 기반으로 새로운 비즈니스 모델을 창출함으로써 디지털 비즈니스 생태계를 조성하는 주요 동인으로 작용하고 있다(Paschek et al., 2017). 이에 따라 인공지능 기술 관련 기업에 대한 자본시장의 관심이 증가하고 있으며, 추후 디지털 비즈니스 생태계의 주축으로 성장할 것으로 기대되고 있다. 그러나 인공지능과 같은 혁신 기술의 경우 기업 또는 사회의 급속한 환경 변화를 야기하며, 이에 대한 사회 구성원의 다양한 견해가 상존하고 있다(Makridakis, 2017). 또한 신기술을 영위하는 기업의 경우 급변하는 환경에서 해당 기술을 통한 미래 이익 예측의 불확실성이 존재한다(Pastor and Veronesi, 2006). 따라서 인공지능에 대한 사회적 수용의 이해와 더불어 기업의 지속적이고 안정적인 성장 예측을 통한 시장의 불확실성 해소가 인공지능 기업 투자 활성화에 중요하다.

Gartner(2017)에 따르면 인공지능에 대한 사회적 수용도는 2017년을 기점으로 초기 채택자(Early adopters) 단계에 진입했으며, 2020년에는 후기 다수자(Early majority) 단계에 안착할 것으로 보인다. 한국의 경우에도 2016년 3월 ‘알파고 쇼크’ 이후 인공지능 기술 활용에 대한 대중의 관심이 급속히 증가하였으며, 다양한 제품이 출시되기 시작한 2017년부터 스마트 스피커의 확산을 통해 실생활에서도 인공지능 기술을 부분적으로 체감할 수 있게 되었다. 그러나 인공지능과 같은 신기술의 경우 초기 기술수용 단계를 거쳐 주류시장으로 안착하기 위해서는 캐즘(Chasm)을 거쳐야 하며, 이 과정에서 관련 기업의 성장과 역량의 확장을 위해서는 대중의 지속적인 관심과 투자 유치를 통한 원활한 자금조달이 필수적이다.

한편, 인터넷을 통한 정보검색의 결과를 기반

으로 의사결정을 하는 일은 일상화되고 있으며, 특히 검색엔진을 통해 특정 정보를 검색하는 행위는 해당 키워드에 대한 관심을 표출하는 행위이다(Choi and Varian, 2012). 즉, 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(Polgreen et al., 2008; Ginsberg et al., 2009), 검색 트렌드 분석을 이용한 자살 위험 예측(Solano et al., 2016) 등 사회적 현상에 대한 대중의 관심과 더불어, 구글 트렌드를 이용한 시장 분석(Ruohonen and Hyrynsalmi, 2017), 기업에 대한 검색량이 주가에 미치는 영향(Da et al., 2011; Bank et al., 2011) 등 광범위한 분야에서 인터넷 검색량을 활용하고 있다. 그러나 아직 기술 키워드 검색량을 사회적 수용 정도의 가늠자로 보고 해당 기술 기업의 주가예측에 활용한 연구는 거의 없다. 물론 주가예측의 경우 해당 기업과 관련한 온라인 문서에 대한 감성분석 기법 등을 활용한 연구는 진행 중이나(Huang et al., 2019; Alsmadi et al., 2019), 기술 키워드를 활용한 주가예측 연구는 거의 없다(Sarode et al., 2019).

따라서 본 연구의 목적은 인공지능 관련 키워드 검색량을 기준으로 인공지능 기술에 대한 대중의 관심도와 세부 기술에 대한 이해도의 변화를 인식하고, 이를 바탕으로 수용 단계를 구분하여 인공지능의 세부 기술 정보 검색량이 인공지능 관련 기업 주가 예측에 주요 요인으로 작용하는지를 검증하는 것이다. 이를 위해 본 연구는 첫째, 인공지능 자체에 대한 키워드 검색량과 머신러닝, 딥러닝 등 인공지능의 세부 기술에 대한 키워드 검색량을 분석함으로써 사회적 수용 단

계를 구분하고, 둘째, 다변량 시계열 데이터의 양방향 관계를 분석하는 모델 중 VAR(Vector Autoregression)과 최근 주가 예측 연구에서 좋은 성능을 보이고 있는 LSTM(Long Short-Term Memory) 알고리즘을 통해 키워드 검색량 기반의 주가 변화 예측모형을 구축하여 인공지능 기술의 사회적 수용 단계에 따른 주가 예측 정확도의 차이를 검증하며, 셋째, 인공지능 기술의 사회적 수용 단계에 따라 세부 기술 키워드 검색량이 주가 예측 정확도에 미치는 영향을 분석함으로써 기술에 대한 사회적 이해도의 중요성을 파악하고자 한다.

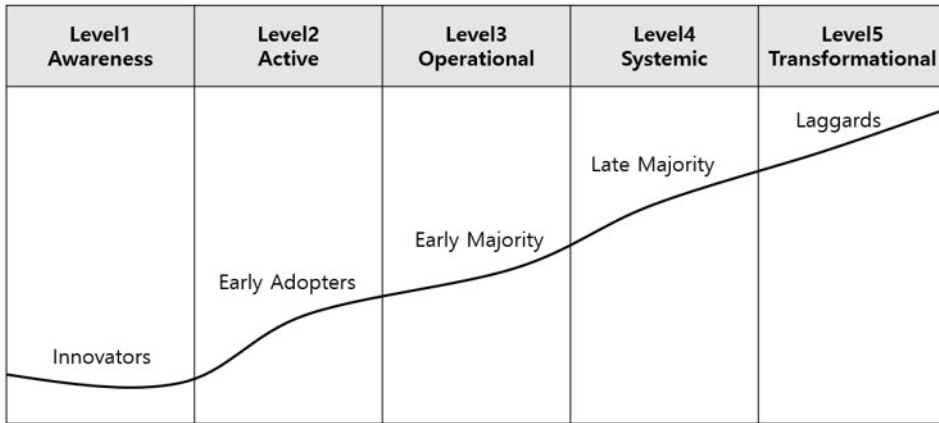
2. 관련 연구

2.1. 사회적 기술 수용도

새로운 기술이 사회에 도입되고 구성원들이 이를 수용하는 과정은 Rogers(1995)의 혁신확산이론(Innovation Diffusion Theory)에 따라 5단계로 구분될 수 있다. 혁신확산이론은 기술이 도입된 이후 사용자들이 기술을 사용하고 받아들이기 시작하는 수용 시점에 따라 그 단계를 구분할 수 있다는 것으로, 수용 시점은 신기술을 받아들이는 과정에서 해당 기술에 대한 새로운 가치나 필요성을 느끼고 기술에 노출되는 정도에 따라 다르게 나타날 수 있다(Rogers, 1995). 혁신 확산이론은 새로운 제품이나 기술을 받아들일 때 사용자들의 채택 요인을 분석하는데 주로 사용되어왔다(Cheong and Park, 2005; Tiong, 2020; Agag et al., 2019). Park and Chen(2007)은 의료 분야 종사자들의 스마트폰 사용의도에 영향을 주는 요인을 혁신확산이론과 기술 수용모델

(Technology acceptance model)에 근거하여 분석하여 스마트폰에 대한 개인의 태도와 인지된 유용성(Perceived usefulness)이 사용의도에 많은 영향을 끼침을 밝혔다. 이와 유사하게 Han et al.(2019)의 연구에서는 AR안경(Augmented Reality Smart Glasses)의 사용의도에 영향을 주는 채택 요인을 혁신확산이론을 바탕으로 문화 관광 맥락에서 분석하였다. 장성희(2010)는 RFID (Radio-Frequency Identification)의 확산에 대한 다양한 관점을 제시하기 위해 혁신확산이론을 바탕으로 연구모형과 가설을 설정하고 RFID 기술 사용에 긍정적 영향을 미치는 요인을 분석하였다. 이처럼 신기술이 개발되고 소비자에게 사용되기 시작할 때, 혁신확산이론의 관점을 통해 기술이 어느 정도 사회에 수용되고 있고 이를 받아들이는 대중들의 사용의도를 파악할 수 있다.

최근에는 4차 산업혁명으로 인해 인공지능(AI), 빅데이터(Big Data), 사물인터넷(IoT)과 같은 새로운 개념의 기술들이 사회에 유입되고 있는데, 인공지능의 사회적 수용 단계의 경우 Gartner(2018)에서 발표한 인공지능 성숙도 모델(AI maturity model)이 가장 대표적으로 활용되고 있다. 혁신확산이론이 개인의 기술 수용 시점에 따라 분류한 것이라면, 인공지능 성숙도 모델은 기업 측면에서 인공지능 기술의 활용 정도에 따라 인식(Awareness), 활성화(Active), 운영(Operational), 전신(Systemic), 변형(Transformational)의 단계로 구분한 모델이다(Figure 1). 인식 단계는 인공지능 기술 사용에 대한 논의는 이루어지고 있지만 구체적인 프로젝트나 비즈니스 전략이 실행되지 않는 단계이고, 활성화 단계는 인공지능에 대한 개념이 정립되고 이를 활용한 파일럿 프로젝트 등이 시작되는 단계이다. 운영 단계에서는 적어도 하나 이상의 인공지능 관련 프로젝트가 진행



〈Figure 1〉 AI Maturity Model

되며 인공지능 관련 예산이 집행된다. 전신 단계는 기업의 새로운 제품과 서비스에 인공지능 기술이 내장되어 있고 직원들이 이 기술을 이해하고 있는 단계이며, 전환 단계는 기업의 모든 업무에 인공지능을 활용함으로써 업무 효율이 증대되는 단계이다.

인공지능은 제4차 산업혁명의 핵심 기술로서 다양한 분야와 융합되며 발전하고 있다. 인공지능 관련 연구는 2000년 340건에서 2017년 1,128건으로 3.3배 가량 증가하였고, 로봇 프로세스 자동화(Robotic Process Automation) 분야와 AI 최적화 하드웨어(AI Optimization H/W) 등 제반 분야에서도 지속적인 증가추세를 보이고 있다 (Cheong et al., 2018). 특히, 2016년 3월 ‘알파고 쇼크’ 이후 인공지능에 대한 대중의 관심이 급증하면서 인공지능과 관련된 다양한 기술과 방법론에 대한 관심도 함께 증가하고 있다. 이는 사용자와 사회에 실질적으로 도움을 줄 수 있는 방향으로 인공지능 관련 기술을 발전시키기 위한 움직임으로 볼 수 있다. 또한 인공지능은 금융, 헬스케어, 교육 등 산업 전반적인 분야에 빠르게

적용되고 있으며, 인공지능 스피커와 같은 제품을 통해 일상 생활에서도 소비자와 상호작용하는 기술들이 개발됨에 따라 다양한 조직의 비즈니스 전략으로서 여러 산업 분야에 활용되고 있다.

따라서 본 연구에서는 혁신확산이론과 인공지능 성숙도 모델에 기반하여 기술의 사회적 수용 단계를 나누어 분석하고자 한다. 이를 위해 대중의 관심도를 파악할 수 있는 지표인 인터넷 검색량을 토대로 인공지능 기술에 대한 사회적 수용 단계를 구분하였다. 검색량의 급격한 변화 시점을 기준으로 인공지능 관련 기술이 대중들에게 잘 알려지지 않았던 시기를 인식 전(Pre-awareness) 단계로, 대중의 관심이 점차 증가하는 시기를 인식(Awareness) 단계로 정의하였고 각 단계별 기술 키워드와 검색량의 변화가 인공지능 관련 기업들의 주가 예측에 함께 사용되었을 경우 예측 성능을 비교하였다.

2.2. 키워드 검색량과 주식시장

인터넷 검색 키워드는 주식시장에서 특정 대

상에 대한 투자자의 관심으로 해석할 수 있다 (Preis et al., 2013; Jeon et al., 2016). 또한 소비자의 미래 행동을 예측할 때 웹 검색량은 중요한 데이터로 활용될 수 있다. 이에 인터넷 검색량이 대중들의 관심 분야를 포괄하는 행위이며 이를 통해 대중들의 미래의 행동을 예측할 수 있음을 보여주는 연구들이 다양한 분야에서 시도되었다. 예를 들어 영화개봉 첫 주말의 박스오피스 매출, 비디오 게임 출시 첫 달 매출, 빌보드 핫 100 차트 곡 순위도 키워드 검색량을 통해 예측 가능하다(Goel et al., 2010). 또한 가까운 미래의 실업급여 청구 건수, 자동차부품 주문량, 소비자 신뢰지수 등을 Google Trends를 이용하여 예측하거나(Choi and Varian, 2012), Yahoo와 Google과 같은 검색 사이트에서 독감(Influenza) 관련 검색 결과와 실제 독감의 발생 건수의 관계를 조사하여 독감 유행을 예측한 바 있다(Polgreen et al., 2008; Ginsberg et al., 2009). 국내에서도 구글 트렌드를 활용하여 제약회사 제품명에 대한 검색량 데이터를 기반으로 바이오 약품 시장에서 기업의 시장 점유율 분석 및 추정(Bong and Lee, 2020), 검색량 트렌드를 통한 스마트 스피커 시장의 분석과 예측(Lee and Youn, 2018), 주식 및 채권과 관련된 구글 검색어 트렌드를 활용한 투자 포트폴리오 구축(Kim and Lee, 2020) 등 다양한 연구가 이루어지고 있다.

주식시장에서도 인터넷 검색량을 투자자들의 관심 척도로 간주하여 특정 기업이나 키워드에 대한 검색량의 변동에 따른 시장 예측이 가능하다. 과거에는 기업의 재무 정보나 기업 성과, 또는 과거의 주식 가격을 예측에 사용하는 통계적인 기법을 바탕으로 미래의 주식 가격을 예측하기도 하였으나(Mondal et al., 2014), 최근에는

SNS와 뉴스기사 등 다양한 비정형 데이터를 통해서도 주식시장 변동을 예측하고 있다(Kim et al., 2014). 인터넷 키워드 검색량을 활용한 주가 예측 연구도 활발히 진행되고 있는데, Preis et al.(2010)의 연구에서는 2004년부터 2010년까지의 구글 검색량 데이터를 이용하여 S&P500 기업의 주간 거래량이 해당 회사명의 주간 검색량과 상관관계가 있다는 것을 밝혔다. 또한 일부 연구에서는 특정 용어에 대한 인터넷 검색량을 기반으로 주식 투자 전략을 세우기 위한 방법론을 제시하고 있으며(Preis et al., 2013; Kim, 2013), 특정 주식에 대한 검색량을 바탕으로 투자자의 관심을 측정하고 주가변동성 간의 관계 분석을 통해 유의성을 입증하는 연구도 진행되었다(Da et al., 2011; Bordino et al., 2012). 국내에서도 인터넷 검색량과 코스피 및 코스닥 상장기업의 주가 변동성 간 상관관계를 실증적으로 분석하기 위한 시도가 있었다(Jeon et al., 2016; Kim, 2017; Lee et al., 2018). 선행연구에 따르면 특정 기술에 대한 키워드 검색량은 해당 기술과 관련된 기업들의 주가 예측에 주요한 요인으로 사용될 수 있고, 기술에 대한 사회적 관심과 수용 단계를 판단하는 기준이 될 수 있다.

이에 본 연구에서는 사회적 수용 단계에 따라 키워드 검색량을 기반으로 주식시장을 예측하고 수용 단계별 정확도를 비교하고자 한다. 추가적으로 한가지 키워드에 대한 검색량 뿐만 아니라 상위개념에 해당하는 키워드와 관련된 세부 기술의 검색량도 주가 예측에 영향을 미칠 것으로 가정하였으며, 세부 기술에 대한 검색량이 주가 예측에 미치는 영향을 수용 단계별로 비교하고 기술에 대한 사회적 이해도와 관련 주식시장 간의 관계를 살펴보고자 한다.

2.3. 주가예측 모델

시계열 데이터 예측은 주로 ARMA(Autoregressive Moving Average), ARIMA(Autoregressive Integrated Moving Average), VAR(Vector Autoregression) 등의 계량적 방법이 활용되어 왔다(Box et al., 2015; Da et al., 2011). 전통적 계량 모델은 주로 통계적 추론 방법을 통해 변수 간의 관계를 발견하거나, 예측변수의 자기 상관성을 검증하고 변수의 이전 값들을 예측에 활용한다. 그 중에서도 VAR 모델은 기본 회귀 모형과 시계열 분석 모형이 가지는 단점을 보완하여 가장 많이 사용되는 다변량 시계열 분석 접근법 중 하나로(Goel et al., 2010), 금융 시계열 데이터인 주가 예측에도 자주 사용되었다. 일례로 Suharsono et al.(2017)은 ASEAN의 주가지수들, SGE, IHSG, KLSE, PSE, SETI의 증가를 예측하는 VAR 모델과 계량적 분석 방법 중 하나인 VECM(Vector error correction models)의 예측 성능을 비교하여 VAR의 성능이 더 우수한 것을 밝혔고, Si et al.(2014)의 연구에서는 기업 증가의 움직임을 예측하는데 트위터와 같은 소셜 미디어를 통해 기업에 대한 감성분석 시계열 데이터를 활용하고, VAR 모델을 통해 사회적 정서에 근거한 주식 예측이 좋은 성능을 보일 수 있음을 주장하였다.

한편, 금융 시계열 데이터의 비선형적이고 시간 의존적인(time-dependent) 문제 해결에 우수한 성능을 보이는 딥러닝(Deep Neural Network) 기반의 예측 모형 연구가 시도되고 있다(Yu and Yan, 2020; Patel et al., 2015). Rather et al.(2015)은 딥러닝 모델 중 RNN(Recurrent Neural Network)과 전통적 시계열 예측 방법인 ARMA 모델을 결합하여 주식시장 예측에 강건성을 지닌 하이브리드 모델을 제안하였고, Khashei and

Hajirahimi(2019)는 딥러닝의 기본 개념이 될 수 있는 MLP(Multi-Layer Perceptron)와 ARIMA 모델의 예측 성능을 비교하고 ARIMA와 MLP를 결합한 모델을 만들어 더 좋은 성능을 가진 예측 모델을 개발하였다. 이처럼 최근에는 금융 시계열 데이터의 예측 성능을 개선하기 위한 방법으로 VAR과 GRU(Gated Recurrent Unit)네트워크가 결합된 분석 모델을 제안하여 딥러닝 분석 방법보다 더 좋은 예측 성능을 나타내는 모델을 개발하는 등(Munkhdalai et al., 2020) VAR 모델의 성능을 개선하기 위해 인공지능 기반 모델과 융합하는 연구가 등장하고 있다.

특히 딥러닝 모델들 중에서도 LSTM(Long Short-Term Memory)이 비선형 시계열 데이터 예측에 좋은 성능을 보이고 있는데, Simi-Namini et al.(2018)과 Qian et al.(2019)의 연구에서 전통적 시계열 기법인 ARIMA 모델과 LSTM을 사용하여 주식 시장을 예측한 결과 LSTM의 성능이 더 우수한 것을 확인하였다. Simi-Namini et al.(2018)은 주가 지수 예측에 ARIMA모델과 LSTM모델을 비교하여 LSTM이 ARIMA모델보다 평균 85%~87%의 오류율을 감소시킨 것을 확인하였으며, 이와 비슷하게 주식의 등락을 예측하는 판별문제에 MLP(Multi-Layer Perceptron), Random Forest, Pseudo-random 방식을 적용해 보았을 때, LSTM의 예측 성능이 가장 좋은 것으로 나타났다(David et al., 2017). 또한 LSTM과 다른 시계열 분석 모형과의 결합을 통해 예측성능을 더욱 올릴 수 있는 하이브리드 모델(Hybrid model)에 대한 다양한 연구가 수행되고 있다. Kim and Won(2018)은 3개의 GARCH모형을 사용하여 매개변수를 선택하고, 선택한 변수들을 입력 데이터로 LSTM에 학습시키는 하이브리드 모델을 통해 단일 계량모형(single econometric models)보다

〈Table 1〉 Literature review

Reference	Method	Data	Comments
Da et al. (2011)	VAR	Russell 3000	Investor interest measurement based on Google search query
Suharsono et al. (2017)	VAR, VECM	SGE, IHSG, KLSE, PSE, SETI	Comparison of ASEAN's forecasting performance of VAR and VECM models
Si et al. (2014)	VAR	goog, amzn, ebay, tgt, wmt, qcom	Propose a notion of Semantic Stock Network (SSN) to improve the prediction of the stock market.
Munkhdalai et al. (2020)	VAR, GRU	S&P 500 index, DJI, NASDAQ et al	Development of a combined model for improving predictive performance of financial multivariate time series data
Patel et al. (2015)	ANN, SVM, RF, Naive-Bayes	CNX Nifty, BSE SENSEX	Predicting direction of movement of stock and stock price index for Indian stock markets
Rather et al. (2015)	RNN, ARMR	NSE	Propose a robust and novel hybrid model for prediction of stock returns
Khashei and Hajirahimi (2019)	ARIMA, MLP	SZII, DJIAI, N225	The performance hybrid models is evaluated for predicting stock prices
Moghar and Hamiche (2020)	LSTM, RNN	Google, NKE opening price	Forecasting the stock market using LSTM and RNN
Kim and Won (2018)	LSTM, GARCH	KOSPI200 index,	Development of a model that is robust to the volatility of the stock market using LSTM and GARCH models
Qian et al. (2019)	ARIMA, LSTM	Chinese Stock Market	Forecasting the stock market using LSTM and comparison of ARIMA model
Wang et al. (2018)	LSTM	Chinese Stock Market	Optimizes the parameters of the LSTM model

하이브리드 모델의 성능이 우수한 것을 확인하였다. 그러나 Goel et al.(2017)의 연구에 따르면 LSTM 모델이 기존의 계량적 시계열 예측 모델에 비해 항상 우수한 성능을 보이지 않으며, 사용되는 데이터의 특성에 따라 VAR 모델의 성능이 더 높게 나타날 수도 있음을 주장하고 있다.

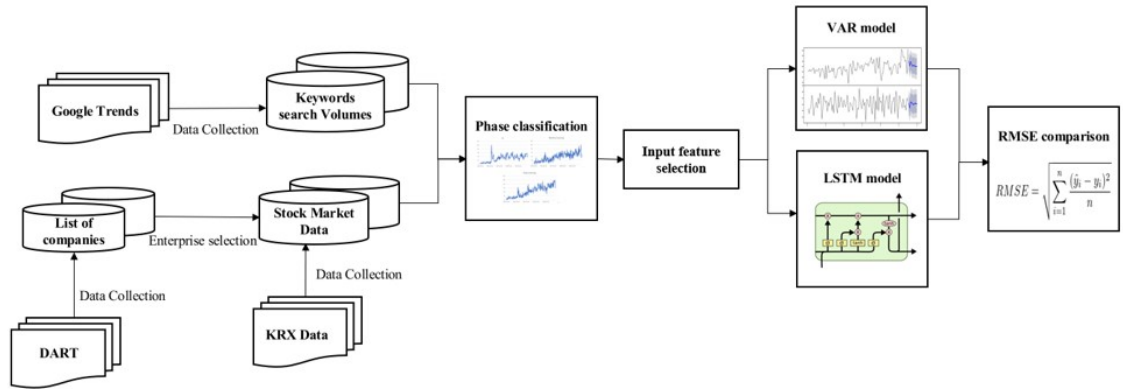
따라서 본 연구에서는 딥러닝 기법 중에서도 시계열 예측에 우수한 성능을 보이는 LSTM 모델과 VAR 모델을 이용하여 인터넷 검색량을 기반으로 주가 예측을 수행하고, 인공지능 기술에 대한 사회적 수용 단계 및 세부 기술 키워드에 따른 예측 성능의 차이를 살펴보고자 한다. 또한

ARIMA 모델을 벤치마크 모델로 선정하여 본 연구에서 제안하는 분석 모델의 성능을 검증하고자 한다.

3. 연구 방법

3.1. 분석절차

본 연구에서 수행한 분석과정은 <Figure 2>와 같다. 먼저 구글 트렌드와 한국 거래소에서 키워드 검색량 데이터와 주가 데이터를 수집한 뒤 키워드 검색량의 변화가 가장 큰 시점을 기준으로



〈Figure 2〉 Analysis Process

기술의 사회적 수용 단계를 나누고, 각 구간별로 VAR 과 LSTM 모델을 사용하여 주가를 예측한다. 평가지표는 예측모델을 성능을 평가하는데 주로 사용되는 RMSE(Root Mean Squared Error)를 사용하였고(Simi-Namini et al., 2018), 각 구간별로 세부 기술에 대한 키워드 검색량이 주가 예측 성능에 미치는 영향을 살펴보았다.

(KOSDAQ)에 상장되어 있는 기업들이며 수집대상 기간 내에 코스닥에 상장되지 않은 2개의 기업을 제외하고 최종적으로 12개의 기업을 분석에 사용하였다. 분석에 사용된 데이터는 한국거래소 정보데이터시스템을 통해 각 기업의 일별 시가, 고가, 저가, 종가 거래량 등의 정보를 수집하였다. 분석 대상 기업 목록은 <Table 2>와 같다.

3.2. 데이터 수집

3.2.1. 주가 데이터

본 연구에서 사용한 주가 데이터는 금융감독원 전자공시 시스템(DART)의 기업정보를 바탕으로 인공지능 관련 사업을 영위하고 있는 14개의 기업을 선정하였으며, 해당 기업들의 2015년 1월1일부터 2019년 12월 31일 까지의 주식데이터를 사용하였다. 기업 선정 기준은 1차적으로 금융감독원 전자공시 시스템의 기업 정보를 활용하여 기업이 제공하는 서비스와 보유 기술을 살펴보았고, 2차적으로 각 기업들의 홈페이지와 전문가 의견을 참고하여 인공지능 관련 기업들을 선정하였다. 선정된 기업들은 모두 코스닥

3.2.2. 키워드 검색량 데이터

키워드 검색량 데이터는 2015년 1월 1일부터 2019년 12월 31일까지 5년 간의 데이터를 구글에서 제공하는 구글 트렌드(Google Trends, <https://trends.google.co.kr/trends/?geo=KR>)를 통해 수집하였다. 구글 트렌드에서 제공되는 데이터는 검색 기간 중 검색 횟수가 가장 많았던 일자를 100으로 산정한 후, 다른 시점의 검색량을 상대적인 수치로 계산하여 나타낸다. 따라서 검색건수에 대한 절대적인 수치가 아닌 수집기간 내에서 각 키워드에 대한 대중들의 관심 변화를 파악하기 수월하다는 장점이 있다. 인터넷 검색량 데이터를 얻기 위한 키워드로는 ‘인공지능’과 함께 대표적인 인공지능 관련 세부 기술이라

〈Table 2〉 Company list

Company	Listing Date	Representative services and technologies	Number of listed shares	Notes
KAONMEDIA	2005.07.12	AI/OTT, Network HW/SW, Mobile platform solutions and development	15,074,293	-
Datasolution	2017.08.03	Bigdata, Predictive analysis, Machine learning, Public Cloud, Azure, Migration	15,828,580	-
Dreamus Company	2003.12.19	Online music service, Distribution of records and digital content, Performance business	56,859,847	In 2019, the company changed its name from Iriver to Dreamus Company
Mobiis	2016.09.08	Machine learning, Accelerator, Products required for control of Big Science instrument	32,171,314	In 2017, the company changed its name from Hana Financial Group No. 8 to Mobiis
SelvasAI	2009.12.10	Smart device, Healthcare, Smartcar, Home IoT	22,055,945	In 2016, the company changed its name from Deotech to SelvasAI
Syntekabio	2019.12.17	New drug development using AI, Anti-cancer bio-marker	13,362,016	-
RFsemi	2007.11.20	Manufacture and sale of semiconductor devices	9,494,724	-
WISE iTech	2018.04.02	Machine Learning, Bigdata Analysis, Application Software Development	4,453,611	Listed on the KOSDAQ from the KONEX in 2020
ESTsoft	2008.07.01	Software business, Artificial intelligence-based business	9,899,400	-
INFINITT Healthcare	2010.05.26	Sales of medical software and maintenance	24,396,458	-
JLK	2019.12.11	AI Healthcare Platform, Remote Healthcare Platform	15,094,489	-
ZUM internet	2016.06.09	Various Internet-based services	26,881,862	-

고 할 수 있는 ‘딥러닝’, ‘머신러닝’ 키워드를 선정하였다.

3.3. 데이터 전처리

수집된 키워드 검색량 데이터는 주 단위 데이터로 총 260건을 수집하였으며, 일별 단위로 수집된 기업의 주가 데이터는 주별 단위로 변경하여 해당 주의 키워드 검색량 값이 매칭되도록 수정하였다. 각 기업이 코스닥에 상장된 일시가 상이하고, 수집 기간 내에 거래가 중지된 이력이

있는 기업도 존재하기 때문에 주가 데이터를 통합하여 분석에 사용하기 위해 주가 지수와 주식 회전율로 계산함으로써 데이터 희소성 문제(data sparsity problem)를 최소화 하였다. 주가 지수와 주식 회전율이 계산된 식은 다음과 같다.

$$\text{Revised Stock Index(RSI)} = \frac{TMC_t/NC_t}{TMC_1/NC_1} \times 100$$

* TMC : Total Market Cap, NC : The number of companies

$$\text{Turnover(T)} = \frac{\text{Weekly transaction volume}}{\text{Total number of listed shares}}$$

<Table 3> Number of training and testing data set

Time period	Definition	Train	Test
Pre-awareness	Only a few people know about AI technology.	48	13
Awareness	AI experimentation, mostly in a data science context	159	40

또한 분석에 사용되는 변수인 주가 지수, 주식 회전율, 키워드 검색량은 값의 범위가 각각 다르기 때문에 정규화 작업을 통해 0과 1사이의 값으로 변환하였다.

3.4. 사회적 수용 단계 분석

본 연구에서는 인공지능 기술의 사회적 수용 단계를 구분하기 위해 인터넷 검색량의 변화율을 사용하였다. 분석 기간에 해당하는 2015년 1월 1일부터 2019년 12월 31일까지 ‘인공지능’ 키워드에 대해 구글 트렌드에서 t시점의 검색량 데이터를 t-1 시점의 값으로 나눈 후 로그 변환을 하였으며, 로그 값의 크기가 큰 폭으로 변화하는 시점을 수용 단계의 변화 시점으로 정하였다 (Jeon et al., 2016).

$$svi_change = \log(svi_t / svi_{t-1})$$

분석 결과 2016년 3월 ‘알파고 쇼크’ 사건이 있었던 시기에 인공지능에 대한 검색량이 폭발적으로 급증하였고, 이 시점을 기준으로 ‘머신러닝’, ‘딥러닝’의 키워드 검색량 또한 점차 증가하는 양상을 보였다. 따라서 2015년 1월부터 2016년 3월 13일까지를 소수의 사용자들에 의해 인공지능 기술이 인식되고 있는 인식 전(Pre-awareness) 단계, 2016년 3월 이후를 인공지능 기술에 대한 대중의 관심이 급증하고 사회적 수용이 확산되는 인식(Awareness) 단계로 구분하여 사용하였다.

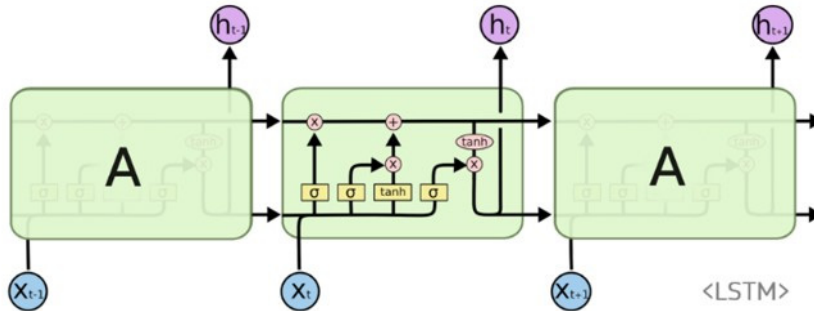
각 구간의 데이터는 주가 예측 검증에 위해 8:2의 비율로 훈련데이터와 검증데이터로 <Table 3>과 같이 나누었다.

3.5. 분석모델

3.5.1. LSTM (Long Short-Term Memory)

LSTM은 RNN이 가지고 있는 데이터의 시퀀스(Sequence)가 장기적으로 길어질수록 점차 학습이 이루어지지 않는 기울기 소실(Vanishing gradient)의 문제를 극복하기 위해 개발된 모델로, 현재 신경망을 통한 주가변동성 예측 연구에 많이 적용되는 모델이다. 또한 장기적인 메모리가 필요한 학습에서 우수한 성능이 나타나는 것이 입증되었으며 RNN 구조로는 불가능한 인공지능 문제들을 해결하였다(Graves, 2012). LSTM은 cell state를 통해 지속적으로 학습할 부분, 삭제할 부분, 새롭게 읽어 들일 부분을 학습하는데, 이는 각각의 입력 게이트(input gate layer), 망각 게이트(forget gate layer)를 통해 이루어지고 최종적으로 어떠한 정보를 출력 게이트(output gate layer)로 내보낼지 결정된다. LSTM의 구조는 <Figure 3>과 같다.

본 연구에서는 LSTM을 이용하여 키워드 검색량을 통해 관련 기업의 주가를 예측하고 RMSE 값을 살펴봄으로써 사회적 기술 수용 단계에 따라 구분한 구간별로 더 정확한 예측성능을 나타내는 세부기술에 대해 알아보하고자 한다.



<Figure 3> LSTM structure

실험에 사용된 LSTM의 구조는 64개의 메모리 셀(memory cell)과 2개의 Dense layer로 구성하였다. Dense layer는 각각 32개의 노드(node)로 이루어져 있으며 ReLU를 활성화 함수로 사용한다. Epoch은 100, Early stopping을 20으로 설정하였고, Batch size는 1, 5, 10, 20으로 각각 수행하여 효율적인 학습과 최적의 예측값을 찾도록 하였다. 손실 함수는 Adam, 입력 변수는 주식 회전을 과 키워드 검색량, 목표변수는 주가 지수로 설정하였다. 또한 학습이 이루어지는 과정에서 4주간의 데이터 변화량이 다음 데이터를 예측하도록 모델을 구축하여 1개월의 추세로 주가 지수와 키워드 검색량의 변화가 예측값에 반영되도록 하였다. 본 연구에서 설계한 LSTM 모델의 파라미터 개수와 층의 구조는 <Table4> 와 같다.

3.5.2. VAR (Vector Autoregressive Model)

VAR 모델은 다변량 시계열 데이터 간의 상호 의존성을 파악하고 변수 간 영향을 분석하는데 사용된다(Munkhdalai et al., 2020; Harivigneshwar et al., 2019). 또한 목표변수(Target variable)를 예측할 때 목표변수에 대한 과거값을 예측에 사용할 뿐만 아니라 다른 변수의 값을 분석에 사용한다는 특징이 있으며, ARIMA 모델과 같이 미래를 예측하면서도 변수 간 상호작용에 대한 영향이 무시되는 상황을 보완해줄 수 있는 모델이다(Moon, 2017). VAR 모델을 나타내는 선형식은 아래와 같이 표현될 수 있다.

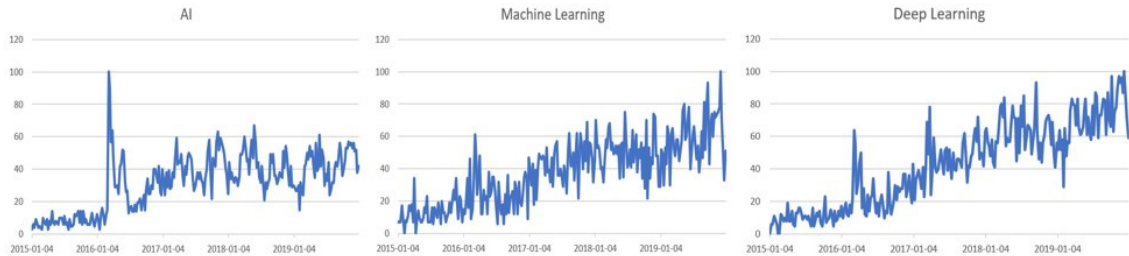
$$Y_t = C + \theta_1 Y_1 + \dots + \theta_p Y_{t-p} + \varepsilon_t$$

<Table 4> LSTM model summary

Layer	Output Shape	Parameters
lstm_1(LSTM)	(None, 64)	17,152
Dense_3(Dense)	(None, 32)	2,080
Dense_4(Dense)	(None, 32)	1,056
Dense_5(Dense)	(None, 1)	33

Note: Total params: 20,321, Trainable params: 20,321

Y 는 N 개의 변수를 포함하는 다변량 벡터를 의미하고, p 는 시차를 나타낸다. 따라서 시차의 적절한 설정에 따라 예측 결과가 상이하게 나타날 수 있는데, 본 연구에서는 LSTM 모델과 동일한 조건에서 비교하기 위해 LSTM의 파라미터 설정과 동일한 4주 동안의 관측값으로 목표변수를 예측하도록 시차(p)를 4로 설정하였다.



〈Figure 4〉 Time series plots of weekly keyword search volumes

4. 실험 및 결과

4.1. 실험 환경

본 연구에서 예측 모델을 구축하기 위해 사용된 하드웨어 환경은 다음과 같다. Intel Core i7-6700K CPU와 32GB DDR4 RAM을 사용하였으며, LSTM 모델 학습에는 GeForce GTX1070 GPU를 사용하였다. 실험 환경은 Ubuntu 16.04 OS에서 Python 3.7.7 버전을 사용하였으며, VAR 모델 구축에는 Statsmodel 0.12.1 버전, LSTM 모델 구축에는 Tensorflow-gpu 1.15.0 버전과 Scikit-learn 0.23.1 버전을 사용하였다.

4.2. 키워드 검색량 분석

2015년 1월 1일부터 2019년 12월 31일까지 구글 트렌드의 주별 키워드 검색량 증가 추이를 분석한 결과는 〈Figure 4〉와 같다. ‘인공지능’ 키워드의 경우 알파고 쇼크가 있었던 2016년 3월에 가장 많은 검색량이 나타났고, 인공지능 세부 기술에 해당하는 ‘머신러닝’과 ‘딥러닝’은 이 시점을 기준으로 점차 증가하는 추세를 보여주고 있다.

추가적으로 ‘인공지능’의 세부 기술로 선정한 ‘머신러닝’, ‘딥러닝’의 상관분석을 통해 ‘인공지

능’과의 상관관계 분석을 진행하였으며, 결과는 〈Table 5〉와 같다. ‘머신러닝’과 ‘딥러닝’ 키워드 검색량은 모두 ‘인공지능’ 키워드 검색량과 0.7 이상의 높은 상관관계가 나타났다. 또한 각 키워드 검색량과 목표변수인 주가 지수의 경우, 수집된 기간 내에서는 ‘딥러닝’ 검색량의 상관계수가 -0.705, ‘인공지능’ 검색량의 상관계수가 -0.679, ‘머신러닝’의 상관계수는 -0.669로 3가지 키워드 모두 주가지수와 음의 상관관계가 나타났다.

〈Table 5〉 Correlation analysis

	AI	ML	DL
AI	1		
ML	.725**	1	
DL	.758**	.822**	1
Stock index	-0.679	-0.669	-0.705

Note: * p < 0.05, ** p < 0.01

4.3. 기술 키워드 기반 예측 결과

본 연구에서는 기술 키워드 검색량을 기반으로 통계적 시계열 분석 기법인 VAR 모델과 딥러닝 기법인 LSTM 모델을 이용하여 인공지능 기술의 사회적 수용 단계에 따른 주가 지수 예측 성능을 비교하였다. 분석은 인공지능(AI), 딥러닝(DL), 머신러닝(ML) 각각의 검색량을 주식 회

(Table 6) The predictive performance of using single keyword search volumes

Time period	Dataset	Model	Batch Size	RMSE (Total execution time)		
				AI, Turnover	ML, Turnover	DL, Turnover
Pre-awareness	61	VAR	-	0.07034*	0.05776*	0.06199*
		LSTM	1	0.13224 (15.64s)	0.17348 (5.25s)	0.09260 (9.39s)
			5	0.16357 (2.59s)	0.16606 (2.54s)	0.17446 (2.60s)
			10	0.15635 (2.24s)	0.17970 (2.14s)	0.18568 (2.26s)
			20	0.19083 (2.04s)	0.20182 (2.02s)	0.19239 (2.17s)
Awareness	199	VAR	-	0.08447	0.08919	0.08954
		LSTM	1	0.06196* (41.88s)	0.06093* (36.83s)	0.07783* (32.59s)
			5	0.06665 (5.09s)	0.07600 (7.25s)	0.09525 (14.37s)
			10	0.06261 (6.46s)	0.07929 (4.34s)	0.07947 (6.18s)
			20	0.07010 (3.13s)	0.06810 (3.64s)	0.08879 (4.08s)

Note: * on RMSE values indicate that best predicting performance in the phase.

전율(Turnover)과 함께 입력변수로 사용하고, 주가 지수를 목표변수로 설정하여 진행하였다. 또한 입력변수의 조합을 다르게 하여 정보의 양 뿐만 아니라 정보의 범위를 확대하였을 때 주가 지수 예측에 어떠한 영향을 끼치는지 살펴보았다. <Table 6>은 각 하나의 키워드 검색량과 주식 회전율을 입력변수로 사용한 결과를 나타낸다.

단일 키워드 검색량으로 주가를 예측한 결과 인식 전(Pre-awareness) 단계에서는 전체적으로 VAR 모델이 LSTM 모델보다 더 좋은 성능을 보였다. VAR 모델에서는 ‘머신러닝’ 검색량과 주식 회전율을 변수로 사용한 경우 RMSE 값이 0.05776으로 예측 성능이 가장 높게 나타났으며, ‘딥러닝’ 검색량과 주식 회전율이 변수일 경우

(RMSE=0.06199), ‘인공지능’ 검색량과 주식 회전율이 변수일 경우(RMSE=0.07034)의 순서로 예측 성능이 높은 것으로 나타났다. 반면, 인식(Awareness) 단계에서는 전체적으로 LSTM 모델이 VAR보다 예측 성능이 더 높은 것으로 나타났다. LSTM 모델의 입력변수에 따른 차이를 살펴보면 ‘머신러닝’ 검색량과 주식 회전율이 사용되었을 때 예측 성능(RMSE=0.06093)이 가장 우수했고, ‘인공지능’ 검색량과 주식 회전율(RMSE=0.06196), ‘딥러닝’ 검색량과 주식 회전율(RMSE=0.07783)의 순서로 성능이 높게 나타났다. 또한 LSTM모델의 학습 속도와 성능의 차이를 살펴보기 위해 batch size를 다르게 설정하였으나, batch size가 증가할수록 학습시간이 감

<Table 6> The predictive performance of using mixed keyword search volumes

Time period	Dataset	Model	Batch Size	RMSE (Total execution time)		
				AI, ML, Turnover	AI, DL, Turnover	ML, DL, Turnover
Pre-awareness	61	VAR	-	0.09817*	0.10818*	0.09068*
		LSTM	1	0.15481 (9.02s)	0.18603 (7.03s)	0.22004 (10.75s)
			5	0.18985 (2.72s)	0.16800 (2.61s)	0.18879 (2.37s)
			10	0.18763 (2.20s)	0.17360 (2.35s)	0.22190 (2.07s)
			20	0.18006 (2.12s)	0.20392 (2.05s)	0.18715 (1.94s)
Awareness	199	VAR	-	0.08598	0.08613	0.09136
		LSTM	1	0.06775* (30.84s)	0.10083 (54.87s)	0.07311* (26.84s)
			5	0.07855 (7.59s)	0.07066* (6.31s)	0.09470 (8.18s)
			10	0.07672 (5.46s)	0.07211 (5.70s)	0.08489 (3.93s)
			20	0.07572 (4.12s)	0.07887 (3.47s)	0.07419 (4.53s)

Note: * on RMSE values indicate that best predicting performance in the phase.

소된다는 차이점 외에 뚜렷한 성능 개선을 보이지 못하였다.

다음으로 각 키워드 검색량을 조합하여 인공지능 기술에 대한 관련정보의 범위가 끼치는 영향을 살펴보았다. 먼저 상위 개념인 ‘인공지능’과 세부 기술인 ‘머신러닝’ 및 ‘딥러닝’ 키워드 검색량을 각각 조합하였으며, 세부 기술 키워드 검색량을 조합하여 분석을 진행하였다. 분석결과는 <Table 6>과 같다.

두 가지의 키워드 검색량을 복합적으로 추가지수 예측에 사용하였을 경우 인식 전(Pre-awareness) 단계에서는 단일 키워드 검색량과 동일하게 VAR 모델의 성능이 전체적으로 높게 나타났다. 키워드 조합에 따른 차이를 살펴보면 ‘머신러닝’

과 ‘딥러닝’ 키워드 검색량을 조합하여 변수로 사용한 경우(RMSE=0.09068)가 가장 높은 성능을 보여주었으며, ‘인공지능’과 ‘머신러닝’ 키워드 검색량의 조합(RMSE=0.09817), ‘인공지능’과 ‘딥러닝’ 키워드 검색량의 조합(RMSE=0.10818)의 순서로 예측 성능이 높게 나타났다. 인식(Awareness) 단계에서도 단일 키워드 분석 결과와 유사하게 LSTM의 성능이 전체적으로 VAR 모델보다 더 좋은 결과를 보였다. 키워드 조합에 따른 차이를 살펴보면 ‘인공지능’과 ‘머신러닝’ 키워드 검색량을 사용한 경우(RMSE=0.06775, batch size=1)가 가장 높은 성능을 보여주었으며, 다음으로는 ‘인공지능’과 ‘딥러닝’ 키워드 검색량의 조합(RMSE=0.07066, batch size=5), ‘딥러

(Table 7) The predictive performance of using all keyword search volumes

Time period	Dataset	Model	Batch size	RMSE
				AI, ML, DL, Turnover
Pre-awareness	61	VAR	-	0.23353
		LSTM	1	0.19209 (11.75s)
			5	0.16656* (2.40s)
			10	0.20166 (2.17s)
			20	0.16864 (2.11s)
Awareness	199	VAR	-	0.08776
		LSTM	1	0.06923* (42.54s)
			5	0.10735 (6.99s)
			10	0.09127 (5.36s)
			20	0.08472 (4.64s)

Note: * on RMSE values indicate that best predicting performance in the phase.

닝'과 '머신러닝' 검색량의 조합(RMSE=0.07311, batch size=1)의 순서로 예측 성능이 높게 나타났다. 두 가지 키워드 검색량 조합을 분석에 사용한 예측 모델의 성능은 단일 키워드 검색량 조합을 사용했을 경우보다 전반적으로 RMSE가 높게 나타나며 예측 성능이 다소 하락하는 것으로 나타났다.

단일 키워드 검색량 기반 예측 모델과 복합 키워드 검색량 기반 예측 모델의 분석 결과를 살펴보면 공통적으로 '머신러닝' 키워드 검색량을 포함할 경우 주가 예측 모델의 성능이 높게 나타났다. 다른 키워드 검색량의 경우 인식 전(Pre-awareness) 단계에서는 인공지능의 세부 기술 키워드인 '딥러닝' 키워드 검색량을 포함한

예측 모델의 성능이 비교적 높게 나타났으며, 인식(Awareness) 단계에서는 포괄적 개념인 '인공지능' 키워드 검색량을 포함한 예측 모델의 성능이 '딥러닝' 키워드 검색량을 사용한 예측모델에 비해 비교적 높게 나타났다.

마지막으로 인공지능 기술에 대한 관련정보의 범위를 가장 넓게 설정하여 모든 키워드 검색량을 조합한 주가 예측 모델의 성능을 살펴보았으며, 결과는 <Table 7>과 같다. 분석 결과 단일 키워드 검색량 및 복합 키워드 검색량을 사용한 예측 모델과는 달리 인식 전(Pre-awareness) 단계에서 VAR 모델(RMSE=0.23353)에 비해 LSTM 모델의 성능(RMSE=0.16656, batch size=5)이 더 우수한 것으로 나타났으며, 인식(Awareness) 단

<Table 8> The predictive performance of ARIMA model

Time period	Dataset	Model	RMSE
			Revised Stock Index
Pre-awareness	61	ARIMA	0.14891
Awareness	199		0.10127

Note: p = 4, q = 4

계에서도 LSTM 모델의 성능(RMSE=0.06923, batch size=1)이 가장 높은 것으로 나타났다.

본 연구에서 사용한 예측 모델의 성능을 검증하기 위해 전통적 시계열 분석 모델인 ARIMA (Autoregressive Integrated Moving Average) 모델을 이용하여 주가 예측 분석을 수행하였다. ARIMA 모델은 과거 데이터가 지니고 있는 추세 (Momentum)를 반영하여 자기회귀와 이동평균을 고려하는 모델로, 시계열 예측 분석에 가장 널리 사용되고 있다(Mondal et al., 2014). 본 연구에서 사용한 모델의 조건과 동일하게 맞추기 위해 ARIMA 모델의 시차는 모두 4로 설정하였으며, 결과는 <Table 8>과 같다.

ARIMA 모델의 예측 성능은 인식 전 (Pre-awareness) 단계에서 RMSE는 0.14891, 인식 (Awareness) 단계에서 RMSE는 0.10127로 나타났다. 이는 본 연구에서 기술 키워드 검색량을 기반으로 주가 지수를 예측한 VAR 모델과 LSTM 모델에 비해 큰 차이를 보이고 있다. 따라서 주가 변동에 대한 예측을 위해서는 다변량 시계열 모델을 통해 키워드 검색량을 활용하는 것이 예측 성능에 유의미한 영향을 미친다는 것을 보여주고 있다.

5. 토의 및 결론

5.1. 주요 발견

본 연구에서 도출된 주요 결과는 다음과 같다. 첫째, 인공지능 기술에 대한 키워드 검색량은 사회적 수용 단계가 진행될수록 증가하는 것으로 나타났다. 특히 알파고 쇼크를 기점으로 인공지능 자체에 대한 키워드 검색량 뿐만 아니라 세부 기술인 머신러닝과 딥러닝의 키워드 검색량 또한 증가하는 양상이 나타났으며, 다양한 인공지능 기반 제품이 상용화와 함께 지속적인 키워드 검색량 상승세를 보여주고 있다. 이는 대중의 혁신 인지도를 높이고 제반 정보를 신속하고 효율적으로 전달하는 커뮤니케이션 전략의 중요성을 시사한다. Rogers(1995, p.195)에 따르면 혁신의 확산은 사회 구성원 사이에서 특정 채널을 통해 커뮤니케이션이 되는 과정으로, 매스미디어와 대인 채널 등 다양한 커뮤니케이션 채널의 중요성을 제시하고 있다. 인공지능 기술의 사회적 수용과정도 알파고 쇼크와 스마트 스피커 등 인공지능 기반 제품의 상용화를 통해 대중은 간접적 경험과 사회에 미칠 영향을 매스 미디어를 통해 자주 접하게 되었으며, 대인 커뮤니케이션 및 인터넷 등 상호작용 커뮤니케이션이 결합됨으로써 혁신 확산 속도에 영향을 미치는 인지된 혁신의 특성 중 적합성(compatibility), 시험가능성(trialability)

과 관찰가능성(observability)에 대한 논의가 활발히 이루어질 수 있었다(Kang and Kim, 2018; Shim and Chae, 2019; Ko et al., 2020).

둘째, 본 연구에서 분류한 인공지능 기술의 사회적 수용단계별 키워드 검색량 기반의 주가예측 결과, 인식 전(Pre-awareness) 단계의 예측 정확도가 가장 높은 것으로 나타났다. 이는 인공지능 기술에 대한 대중의 관심이 확산되기 이전 해당 분야에 대한 지식이 있는 일부 초기 수용자의 정보 검색량이 기업의 주가 현황을 보다 더 잘 반영한다는 점을 보여준다. 현재 인공지능 분야의 기술 혁신 속도가 급속하게 증가하고 있는 상황에서 대중은 인공지능 기반의 다양한 제품과 서비스를 대면하게 된다. 이에 따라 인공지능 기술에 대한 지식을 습득하고자 하는 호기심과 새로운 변화에 대한 흥미가 동반되며(Sohn and Kwon, 2020), 인공지능 기술에 대한 대중의 관심이 증가할수록 관련 기업의 향후 성과와 이를 통해 창출할 수 있는 미래가치에 대한 낙관적인 기대치가 높아짐으로써 실제 투자행위로 이어지게 된다(Liang, 2006; Kim, 2018). 그러나 본 연구의 분석대상인 인공지능 기술 기반 기업의 경우 KOSDAQ 상장 초기 기업이며, 대기업 위주의 인공지능 관련 시장 구조와 아직까지 상용화 기술의 한계가 존재하고 있다(Park, 2018). 따라서 인공지능 기술을 통한 미래 이익 예측의 불확실성은 상존하고 있으며(Pastor and Veronesi, 2004), 대중의 지속적인 관심 증가와 실질적인 기업의 주가는 일치하지 않을 수 있음을 의미한다.

셋째, 인공지능 기술에 대한 사회적 수용단계별로 주가예측 모델에 사용된 키워드에 따라 예측 정확도가 다른 양상을 보이는 것으로 나타났다. 인식 전(Pre-awareness) 단계에서는 인공지능

기술에 대한 사회적 인식이 높지 않은 상황이기 때문에 일부 전문적 지식을 지닌 초기 수용자의 정보 검색이 주를 이루며 인공지능의 가장 대표적인 기술인 ‘머신러닝’과 ‘딥러닝’ 키워드의 성능이 가장 높게 나타났으나, 인식(Awareness) 단계에서는 ‘딥러닝’ 키워드보다 상위 개념인 ‘인공지능’ 키워드를 사용한 모델의 예측 성능이 더 높게 나타났다. 또한 키워드 간 조합을 사용한 경우보다 ‘머신러닝’ 단일 키워드를 사용한 예측 모델의 성능이 가장 높은 것으로 나타났는데, 이는 세부 기술과 같은 특정 정보에 대한 검색량이 주가변동성 예측에 유의하게 작용할 수 있다는 점을 시사한다(Bank et al., 2011; Da et al., 2011). 따라서 기술 키워드를 활용한 주가예측 모델 구축 시 해당 기술의 사회적 수용도와 하위 기술 분류에 대한 고려가 필요하다.

5.2. 시사점

본 연구의 결과는 다음과 같은 학술적 시사점을 제공한다. 첫째, 혁신 기술을 기반으로 하는 기업에 대한 투자 수익률 예측을 위해서는 해당 기술의 사회적 수용 과정에서 대중의 관심이 급증하는 인식 단계를 포착하는 것이 가장 중요하다. 일반적으로 주가예측 모델은 변동성, 거래량, 경제지표 등 변수를 활용하여 예측의 정확도를 높이는 방법으로 분석 방법의 개선을 추구해왔다(Box et al., 2015). 그러나 일부 전통산업을 제외할 경우 현재 상장되는 대부분의 기업은 특정 기술을 기반으로 하고 있으며, 기술에 대한 사회적 수용은 기술의 특징과 국가에 따라 다른 양상을 보이며 수용 단계도 상이하다(Zhu et al., 2006; Erumban and De Jong, 2006). 따라서 기술 기반 기업의 주가예측 모델을 구축하기 위해서

는 키워드 검색량, 뉴스 보도량 등 커뮤니케이션 채널에 대한 분석을 바탕으로 산업과 기업의 성장과 더불어 사회적 수용 현상에 대한 이해가 병행되어야 한다.

둘째, 키워드 검색량을 예측 모델의 변수로 사용할 경우 해당 키워드가 내재한 정보의 범위를 고려해야 한다. 구글 트렌드 키워드는 주식 시장 예측 정확도에 상당한 영향을 미치고 있으나 (Pai et al., 2018), 어떠한 키워드를 선택하느냐에 따라 예측 정확도의 결과는 크게 변동될 수 있다 (Naccarato et al., 2018). 본 연구에서는 보편적이고 광의의 정보를 포함하는 ‘인공지능’ 키워드와 특정 기술에 대한 정보를 포함하는 ‘머신러닝’ 및 ‘딥러닝’ 키워드를 사용하였으며, 각 키워드 간 상관관계는 매우 높게 나타났다. 그러나 분석 결과 ‘머신러닝’ 키워드를 사용한 경우 예측 성능이 가장 우수하게 나타났으며, 이는 기업의 주가 예측에는 범용적인 정보를 포함하는 세부 기술 키워드를 사용하는 것이 유효하다는 점을 시사한다. 따라서 키워드 검색량을 활용한 시장 예측 연구를 수행할 경우 목적 시장의 정보를 반영할 수 있는 키워드의 선택이 필요하다.

또한 본 연구의 결과는 실무적 측면에서도 활용될 수 있으며, 다음과 같은 시사점을 제공한다. 첫째, 본 연구에서 검증한 기술의 사회적 수용도에 따른 키워드 검색량의 변화 추이와 예측 모델의 정확도 변화는 최근 금융권에서 도입하고 있는 빅데이터 기반 인공지능 로보어드바이저의 투자 의사 결정 시스템에 활용될 수 있을 것으로 기대된다. 최근 주식 시장의 차트 이미지를 활용하는 CNN(Hoseinzade and Haratizadeh, 2019; Liu et al., 2020) 기반의 모델과 시계열 특성을 지닌 주가 데이터를 활용하는 LSTM(Roondiwala et al., 2018; Ding and Qin, 2020) 기반의 모델 등 다

양한 인공지능경망을 융합하는 주가예측 연구가 진행되고 있다. 이러한 상황에서 기술 키워드 검색량을 포함하는 예측 모델을 통해 미래 특정 기간에 대한 리스크를 보다 합리적인 방법으로 예측하고, 이에 맞는 투자전략 수립과 효율적 수행이 가능할 것으로 예상된다.

둘째, 본 연구에서 사용한 VAR 모델과 LSTM 모델의 예측 성능은 인공지능 기술에 대한 사회적 수용 단계에 따라 다른 양상으로 나타났다. 비교적 데이터의 길이가 짧은 인식 전 (Pre-awareness) 단계에서는 VAR 모델의 성능이 우수했으며, 데이터의 길이가 증가한 인식 (Awareness) 단계에서는 LSTM 모델의 성능이 우수하게 나타났다. LSTM 모델은 자체적으로 시계열 데이터에서 높은 수준의 시간적 패턴을 학습하는데 탁월한 성능을 보이는 특성이 있기 때문에, 정보량이 증가함에 따라 변동성 예측에 필요한 기능을 효과적으로 학습할 수 있다(Kim and Won, 2018). 따라서 다변량 시계열 예측을 위한 모델 구축 시 최신의 기계학습 알고리즘이 항상 우수한 것은 아니며, 전통적인 계량적 접근 방식과 적절한 결합을 통해 정보량에 따라 성능의 변동에 대처할 수 있는 효과적인 모델을 개발할 수 있음을 시사한다.

5.3. 한계점

본 연구는 주가 예측을 위해 인공지능 기술에 대한 사회적 수용 단계를 기준으로 키워드 검색량을 활용한 VAR 모델과 LSTM 모델을 구축하였다는 의의가 있음에도 불구하고 몇 가지 연구의 한계점도 존재한다. 첫째, 본 연구에서 수집한 구글 트렌드의 키워드 검색량은 인공지능, 머신러닝, 딥러닝으로 한정되어 있으나, 해당 데이

터만으로는 인공지능 기술에 대한 대중의 관심 정도를 정확히 유추하기가 부족할 수 있다. 따라서 추후 연구에서는 인공지능에 대한 뉴스 기사 및 댓글 등의 토픽 데이터를 통해 보다 구체화된 키워드 검색어를 사용함으로써 대중의 관심도를 보다 정확히 도출할 수 있을 것이다.

둘째, 본 연구에서 구분한 인공지능 기술에 대한 사회적 수용 단계의 기준점은 연구자 해석에 따라 차이가 있을 수 있다. 알파고 쇼크를 기준으로 인공지능 기술에 대한 사회적 수용이 인식(Awareness) 단계로 넘어간 것에는 선행연구(Kang and Kim, 2018; Shim and Chae, 2019)와 키워드 검색량 변화의 로그 계산을 통해 도출할 수 있었으나, 활용(Active) 단계로 전환되는 시점에 대해서는 논의가 필요하다. 특히 사회적 수용 단계의 경우 대중의 인식뿐만 아니라 관련 산업의 기술 동향에 대한 반영도 필요하며, 이에 대한 관련 키워드와 시장 반응 등에 대한 고려가 종합적으로 필요하다. 따라서 추후 연구에서는 보다 구체적인 사회·경제적 지표를 통해 인공지능 기술에 대한 사회적 수용 단계를 도출할 수 있을 것이다.

셋째, 본 연구의 분석대상인 인공지능 기술 기반 기업의 경우 현재 상장 초기 기업이 대부분이며, 거래정지 등으로 인한 데이터의 불연속성이 존재하기 때문에 주가지수를 별도로 산출하여 분석을 진행했다. 이 경우 산업에 대한 전반적인 주가 등락 추세를 관찰할 수 있으나, 개별 기업의 기술력 차이, 사업의 특수성 등이 고려되지 않는다는 한계가 존재한다. 따라서 향후 연구에서는 분석의 범위를 개별 기업으로 구체화하여 보다 세분화된 주가예측 모델을 구축할 수 있을 것이다.

참고문헌(References)

- [1] Agag, G. M., M. A. Khashan, and M. H. ElGayaar, "Understanding online gamers' intentions to play games online and effects on their loyalty: An integration of IDT, TAM and TPB", *Journal of Customer Behaviour*, Vol.18, No.2(2019), 101-130.
- [2] Alsmadi, I., M. Al-Ayyoub, M. Alsmirat, and Y. Jararweh, "Using popular search terms in stock price prediction", In 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), 279-285).
- [3] Bank, M., M. Larch, and G. Peter, "Google search volume and its influence on liquidity and returns of German stocks", *Financial Markets and Portfolio Management*, Vol.25, No.3(2011), 239-264.
- [4] Bong, K. T., and H. S. Lee, "Analysis and estimation for market share of biologics based on Google trends big data", *Journal of the Society of Korea Industrial and Systems Engineering*, Vol.43, No.2(2020), 14-24.
- [5] 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.
- [6] Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.
- [7] Boyd, R., and R. J. Holton, "Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation", *Journal of Sociology*, Vol.54, No.3(2018), 331-345.

- [8] Cheong, J. H., and M. Park, "Mobile internet acceptance in Korea", *Internet Research*, Vol. 15 No. 2(2005), 125-40.
- [9] Choi, H., and H. Varian, "Predicting the present with Google Trends". *Economic record*, Vol.88, (2012), 2-9.
- [10] Chung, M. S., and J. Y. Lee, "Systemic Analysis of Research Activities and Trends Related to Artificial Intelligence(A.I.) Technology Based on Latent Dirichlet Allocation (LDA) Model)", *Journal of the Korea Industrial Information Systems Research*, Vol.23, No.3(2018), 87-95.
- [11] Chung, M. S., S. H. Park, B. H. Chae, and J. Y. Lee, "Analysis of major research trends in artificial intelligence through analysis of thesis data", *Journal of Digital Convergence*, Vol.15, No.5(2017), 225-233.
- [12] Da, Z., J. Engelberg, and P. Gao, "In search of attention", *The Journal of Finance*, Vol.66, No.5(2011), 1461-1499.
- [13] Ding, G., and L. Qin, "Study on the prediction of stock price based on the associated network model of LSTM", *International Journal of Machine Learning and Cybernetics*, Vol.11, No.6(2020), 1307-1317.
- [14] Erumban, A. A., and S. B. De Jong, "Cross-country differences in ICT adoption: A consequence of Culture?", *Journal of World Business*, Vol.41, No.4(2006), 302-314.
- [15] Gartner, *Gartner's 2017 Hype Cycle for Artificial Intelligence*, 2017. Available at: <https://www.gartner.com/doc/3770467/hype-cycle-artificial-intelligence-> (accessed 28 November 2020)
- [16] Gartner, "Applying Artificial Intelligence to Drive Business Transformation: A Gartner Trend Insight Report", 2-7. 2018.
- [17] 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(2009), 1012-1014.
- [18] Goel, H., I. Melnyk, N. Oza, B. Matthews, A. Banerjee, "Multivariate aviation time series modeling: VARs vs. LSTMs", *In Proceedings of the SIAM International Conference on Data Mining (SDM)*, (2017), 27-29.
- [19] Goel, S., J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts, "Predicting consumer behavior with Web search", *Proceedings of the National academy of sciences*, Vol.107, No.41(2010), 17486-17490.
- [20] Graves, A., *Supervised sequence labelling with recurrent neural networks* (pp. 37-45). Springer, Berlin, Heidelberg.2012
- [21] Han, D. I. D., M. C. Tom Dieck, and T. Jung, "Augmented Reality Smart Glasses (ARSG) visitor adoption in cultural tourism", *Leisure Studies*, Vol.38, No.5(2019), 618-633.
- [22] Harivigneshwar, C. J., K. B. Dharmavenkatesan, R. Ajith, and R. Jeyanthi, "Modeling of Multivariate Systems using Vector Autoregression (VAR)", *In 2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vol.1, (2019), 1-6.
- [23] Hoseinzade, E., and S. Haratizadeh, (2019). "CNNpred: CNN-based stock market prediction using a diverse set of variables", *Expert Systems with Applications*, Vol.129, (2019), 273-285.

- [24] Huang, T. C., R. N. Zaeem, and K. S. Barber, "It is an equal failing to trust everybody and to trust nobody: Stock price prediction using trust filters and enhanced user sentiment on Twitter", *ACM Transactions on Internet Technology (TOIT)*, Vol.19, No.4(2019), 1-20.
- [25] Jang, S. H., "A Study on the Factors Influencing RFID Diffusion: In the Perspective of Innovation Diffusion Theory", *Journal of the Korea society of computer and information*, Vol.15, No.11(2010), 173-183.
- [26] Jeon, S. M., Y. J. Chung, and D. Y. Lee, "The Relationship between Internet Search Volumes and Stock Price Changes: An Empirical Study on KOSDAQ Market", *Journal of Intelligence and Information Systems*, Vol.22, No.2(2016), 81-96.
- [27] Kang, W. K., and B. R. Kim," Consideration of Human Emotions about Artificial Intelligence - Focused on the Analysis of Newspaper Articles on AlphaGo VS Lee Sedol", *Journal of Korean Ethics Studies*, Vol.1, No.123(2018), 181-201.
- [28] Khashei, M., and Z. Hajirahimi, "A comparative study of series arima/mlp hybrid models for stock price forecasting", *Communications in Statistics-Simulation and Computation*, Vol.48, No.9(2019), 2625-2640.
- [29] Kim, D. Y., J. W. Park, and J. H. Kim, "A Comparative Study between Stock Price Prediction Models Using Sentiment Analysis and Machine Learning Based on SNS and News Articles". *Journal of Information Technology Services*, Vol.13, No.3(2014), 211-233.
- [30] Kim, E. C., and D. W. Lee, "A study on asset allocation strategy using Google trends", *Journal of the Korean Data & Information Science Society*, Vol.31, No.1(2020), 173-186.
- [31] Kim, H. Y., and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models", *Expert Systems with Applications*, Vol.103, (2018), 25-37.
- [32] Kim, M. S., and H. J. Kwon, "The Effect of Portal Search Intensity on Stock Price Crash", *The Journal of Society for e-Business Studies*, Vol.22, No.2(2017), 153-168.
- [33] Kim, M. S., and P. H. Koo, "A Study on Big Data Based Investment Strategy Using Internet Search Trends", *Journal of the Korean Operations Research and Management Science Society*, Vol.38, No.4(2013), 53-63.
- [34] Kim, R. M., "An Empirical Study on the Relation between Search Volume, Investors Trading, and Stock Returns", *The Korean Journal of Financial Engineering*, Vol.17, No.2(2018), 53-85.
- [35] Ko, H. S., D. H. Park, and N. R. Lee, "Challenges of Establishing Ethics Principles and a Governance Regime for Artificial Intelligence", *Journal of Law & Economic Regulation*, Vol.13, No.1(2020), 7-36.
- [36] Lee, B. W., J. H. Kim, and J. P. Yu, "Forecasting Company Sales and Stock Price Using Google Trend: Focusing on the Keywords of BMW and Mercedes-Benz", *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, Vol.8, No.10(2018), 491-501.
- [37] Lee, G., and S. Youn, "Smart speaker market analysis and forecast using Google trends", *KIISE Transactions on Computing Practices*, Vol.24, No.11(2018), 596-602.

- [38] Liang, X., “Mining associations between web stock news volumes and stock prices”, *International Journal of Systems Science*, Vol.37, No.13(2006), 919-930.
- [39] Liu, P., J. Liu, and K. Wu, “CNN-FCM: System modeling promotes stability of deep learning in time series prediction”, *Knowledge-Based Systems*, (2020), 106081.
- [40] Makridakis, S., “The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms”, *Futures*, Vol.90, (2017), 46-60.
- [41] Mondal, P., L. Shit, and S. Goswami, “Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices”, *International Journal of Computer Science*, Vol.4, No.2(2014), 13-29.
- [42] Moon, K. S., “Vector Autoregressive Model: VAR”, *Journal of The Korean Official Statistics*, Vol.2, No.1(1997), 23-56.
- [43] Munkhdalai, L., M. Li, N. Theera-Umpon, S. Auephanwiriyakul, and K. H. Ryu, “VAR-GRU: A Hybrid Model for Multivariate Financial Time Series Prediction”, In *Asian Conference on Intelligent Information and Database Systems*, (2020), 322-332.
- [44] Naccarato, A., S. Falorsi, S. Lorig, and A. Pierini, “Combining official and Google Trends data to forecast the Italian youth unemployment rate”, *Technological Forecasting & Social Change*, Vol.130, (2018), 114-122.
- [45] Pai, P. F., L. C. Hong, and K. P. Lin, “Using internet search trends and historical trading data for predicting stock markets by the least squares support vector regression model”, *Computational Intelligence and Neuroscience*, Vol.2018, (2018), 6305246.
- [46] Park, S. U., “AI technology and market trends”, *The magazine of KIICE*, Vol.19, No.2(2018), 11-22.
- [47] Park, Y., and J. V. Chen, “Acceptance and adoption of the innovative use of smartphone”, *Industrial Management & Data Systems*, Vol.107, No.9, (2007), 1349-1365.
- [48] Paschek, D., C. T. Luminosu, and A. Draghici, “Automated business process management—in times of digital transformation using machine learning or artificial intelligence”, In *MATEC Web of Conferences 121*, (2017), 04007.
- [49] Pástor, L., and P. Veronesi, “Was there a Nasdaq bubble in the late 1990s?”, *Journal of Financial Economics*, Vol.81, No.1(2006), 61-100.
- [50] Patel, J., S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques”, *Expert systems with applications*, Vol.42, No.1(2015), 259-268.
- [51] Polgreen, P. M., Y. Chen, D. M. Pennock, F. D. Nelson, and R. A. Weinstein, “Using internet searches for influenza surveillance”, *Clinical infectious diseases*, Vol.47, No.11, (2008), 1443-1448.
- [52] Preis, T., D. Reith, and H.E. Stanley, “Complex dynamics of our economic life on different scales : insights from search engine query data”, *Philosophical Transactions of the Royal Society*, Vol.368, (2010), 5707-5719.
- [53] Preis, T., H. S. Moat, and H. E. Stanley, “Quantifying trading behavior in financial markets using Google Trends”, *Scientific Report*, Vol.3, No.1, (2013), 1-5.

- [54] Qian, F., and X. Chen, “Stock prediction based on lstm under different stability”, In *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 483-486, (2019).
- [55] 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.
- [56] Rogers, E. M. *Diffusion of innovations*, 4th edition, Free Press, New York. 1995.
- [57] Roondiwala, M., H. Patel, and S. Varma, “Predicting stock prices using LSTM”, *International Journal of Science and Research*, Vol.6, No.4(2017), 1754-1756.
- [58] Ruohonen, J., and S. Hyrynsalmi, “Evaluating the use of internet search volumes for time series modeling of sales in the video game industry”, *Electronic Markets*, Vol.27, No.4(2017), 351-370.
- [59] Sarode, S., H. G. Tolani, P. Kak, and C. S. Lifna, “Stock price prediction using machine learning techniques”. In *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, 177-181, (2019).
- [60] Shim, J. W., and S. G. Chae, “Seeking Possibility of Ethical Issues Based on Public Attitude Toward Artificial Intelligence Through Analysis of Social Network Data”, *The Journal of Humanities and Social science*, Vol.10, No.3(2019), 1337-1347.
- [61] Si, J., A. Mukherjee, B. Liu, S. J. Pan, Q. Li, and H. Li, “Exploiting social relations and sentiment for stock prediction”, In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (2014), 1139-1145.
- [62] Siami-Namini, S., N. Tavakoli, and A. S. Namin, “A comparison of ARIMA and LSTM in forecasting time series”, In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1394-1401, (2018).
- [63] Sohn, K., and O. Kwon, “Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products”, *Telematics and Informatics*, Vol.47, (2020), 101324.
- [64] Solano, P., M. Ustulin, E. Pizzorno, M. Vichi, M. Pompili, G. Serafini, and M. Amore, “A Google-based approach for monitoring suicide risk”, *Psychiatry Research*, Vol.246, (2016), 581-586.
- [65] Suharsono, A., A. Aziza, and W. Pramesti, “Comparison of vector autoregressive (VAR) and vector error correction models (VECM) for index of ASEAN stock price”, In *AIP Conference Proceedings*, Vol.1913, No.1(2017), 020032.
- [66] Tiong, W. N., “Factors Influencing Behavioural Intention towards Adoption of Digital Banking Services in Malaysia”, *International Journal of Asian Social Science*, Vol.10, No.8(2020), 450-457.
- [67] Yu, P., and X. Yan, “Stock price prediction based on deep neural networks. Neural Computing and Applications”, *Neural Computing and Applications*, Vol.32, No.6(2020), 1609-1628.

Abstract

Comparison of Models for Stock Price Prediction Based on Keyword Search Volume According to the Social Acceptance of Artificial Intelligence

Yujung Cho* · Kwonsang Sohn** · Ohbyung Kwon***

Recently, investors' interest and the influence of stock-related information dissemination are being considered as significant factors that explain stock returns and volume. Besides, companies that develop, distribute, or utilize innovative new technologies such as artificial intelligence have a problem that it is difficult to accurately predict a company's future stock returns and volatility due to macro-environment and market uncertainty. Market uncertainty is recognized as an obstacle to the activation and spread of artificial intelligence technology, so research is needed to mitigate this. Hence, the purpose of this study is to propose a machine learning model that predicts the volatility of a company's stock price by using the internet search volume of artificial intelligence-related technology keywords as a measure of the interest of investors. To this end, for predicting the stock market, we using the VAR(Vector Auto Regression) and deep neural network LSTM (Long Short-Term Memory). And the stock price prediction performance using keyword search volume is compared according to the technology's social acceptance stage. In addition, we also conduct the analysis of sub-technology of artificial intelligence technology to examine the change in the search volume of detailed technology keywords according to the technology acceptance stage and the effect of interest in specific technology on the stock market forecast. To this end, in this study, the words artificial intelligence, deep learning, machine learning were selected as keywords. Next, we investigated how many keywords each week appeared in online documents for five years from January 1, 2015, to December 31, 2019. The stock price and transaction volume data of KOSDAQ listed companies were also collected and used for analysis. As a result, we found that the keyword search volume for artificial intelligence technology increased as the social acceptance of artificial intelligence technology increased. In

* Department of Big Data Analytics, Kyung Hee University

** School of Management, Kyung Hee University

*** Corresponding author: Ohbyung Kwon

School of Management, Kyung Hee University

26-6, Kyunghedae-ro, Dongdaemun-gu, Seoul, Republic of Korea

Tel: +82-02-961-2148, E-mail: obkwon@khu.ac.kr

particular, starting from AlphaGo Shock, the keyword search volume for artificial intelligence itself and detailed technologies such as machine learning and deep learning appeared to increase. Also, the keyword search volume for artificial intelligence technology increases as the social acceptance stage progresses. It showed high accuracy, and it was confirmed that the acceptance stages showing the best prediction performance were different for each keyword. As a result of stock price prediction based on keyword search volume for each social acceptance stage of artificial intelligence technologies classified in this study, the awareness stage's prediction accuracy was found to be the highest. The prediction accuracy was different according to the keywords used in the stock price prediction model for each social acceptance stage. Therefore, when constructing a stock price prediction model using technology keywords, it is necessary to consider social acceptance of the technology and sub-technology classification. The results of this study provide the following implications. First, to predict the return on investment for companies based on innovative technology, it is most important to capture the recognition stage in which public interest rapidly increases in social acceptance of the technology. Second, the change in keyword search volume and the accuracy of the prediction model varies according to the social acceptance of technology should be considered in developing a Decision Support System for investment such as the big data-based Robo-advisor recently introduced by the financial sector.

Key Words : Artificial Intelligence, Social Acceptance, Stock Price Prediction, Multivariate Time Series, Keyword Search Volume, LSTM, VAR

Received : December 30, 2020 Revised : March 2, 2021 Accepted : March 8, 2021

Corresponding Author : Ohbyung Kwon

저 자 소개



조유정

현재 경희대학교 일반대학원 빅데이터 응용학과에서 석사과정에 재학중이다. 수원대학교 응용통계학과에서 학사학위를 취득하였다. 관심 분야는 딥러닝, 데이터 분석, 컴퓨터 비전, 데이터 마이닝 등이다.



손권상

인하대학교에서 국제통상학 학사학위와 경영학 석사학위를 취득하였고, 경희대학교 일반대학원 경영학과에서 박사과정을 수료하였다. 현재 인하대학교와 광운대학교 경영학과 강사로 재직 중이며, 관심분야는 AI 기반 의사결정지원, 비즈니스 애널리틱스, 제로 샷 학습, 추천시스템 등이다.



권오병

현재 경희대학교 경영대학 교수로 재직 중이다. 서울대학교 경영학과에서 학사학위를 한국과학기술원에서 석사 및 박사학위를 취득하였고, 카네기멜론대학 ISRI연구소에서 유비쿼터스 컴퓨팅 프로젝트를 수행한 바 있다. 관심분야는 AI비즈니스, 텍스트 분석, 휴먼로봇 인터페이스, 상황인식 서비스, 의사결정지원시스템 등이다.