

# 빅데이터 기반의 국제거시경제 전망모형 개발 연구

**백예인** 대외경제정책연구원 국제거시금융실  
국제거시팀 부연구위원  
yaein.baek@gmail.com

**윤상하** 대외경제정책연구원 국제거시금융실  
국제거시팀 부연구위원  
syoon@kiep.go.kr

**김현학** 국민대학교 경제학과 부교수  
hyunhak.kim@kookmin.ac.kr

**이지윤** 대외경제정책연구원 국제거시금융실  
국제거시팀 연구원  
l jy08@kiep.go.kr



## 차 례

1. 연구의 배경 및 목적
2. 조사 및 분석 결과
3. 정책 제언

## 주요 내용

- ▶ 최근 글로벌 인플레이션, 코로나19 팬데믹 등으로 인한 경제 불확실성이 높아지면서 거시경제 전망의 정확성과 속보성에 대한 중요도가 높아지고 있음.
  - 빅데이터를 통해 잠정적으로 중요한 모든 정보를 전망에 활용할 수 있으므로 이를 사용한 거시경제 전망 연구가 필요
- ▶ 한국의 GDP 성장률 전망에 빅데이터를 활용하여 예측력이 향상되었으며, 이는 더 많은 데이터를 사용하면 잠재적으로 중요한 정보를 추가로 활용할 수 있다는 예상과 부합
  - 대량의 거시·금융 지표를 기반으로 미국과 한국의 GDP 성장률을 전망한 결과, 빅데이터를 활용한 머신러닝이 전통적 계량경제 모형에 비해 예측력이 향상
  - 네이버 검색 데이터와 동적모형 평균화 및 선택을 이용하여 한국의 경제성장률을 전망한 결과, 기존의 정형 데이터만 사용한 모형보다 개선된 예측력을 나타냈으며 경기변동 시점을 예측하는 데 유용할 것으로 보임.
- ▶ 향후 다양한 소스의 데이터를 발굴하여 거시경제 데이터베이스를 구축하고, 경제 전망에 활용할 수 있는 새로운 분석 기법을 개발하고 연구할 필요가 있음.

## 1. 연구의 배경 및 목적

- 전반적인 경제 상황을 판단하는 경제지표 예측의 중요성에 대한 인식은 빅데이터(big data)의 등장으로 전망에 사용할 새로운 데이터에 대한 기대를 높이고 있음.
  - 최근 글로벌 인플레이션, 코로나19 팬데믹 등으로 인한 경제 불확실성하에서 전망모형의 예측력을 높이기 위해서는 잠정적으로 중요한 모든 정보를 활용할 필요가 있음.
- 빅데이터가 경제 및 금융 상황에 대한 새로운 정보를 제공하고 이를 분석할 방법론으로 머신러닝이 주목 받으면서 수많은 연구가 수행되고 있음.
  - 소셜미디어, 온라인 검색, 뉴스 및 신문기사 등의 텍스트 데이터로부터 경제심리지수를 구축하여 사용 중
    - 네덜란드 통계청에서 구축한 소셜미디어지수(SMI: Social Media Index)가 소비자신뢰지수와 유의미한 관계를 보이는 것을 나타냈으며, 한국은행은 경제 분야 뉴스 기사에 나타나는 경제 심리를 지수화한 뉴스심리지수(NSI: News Sentiment Index)를 개발하여 주요 경제심리지표 및 실물경기지표 보다 선행함을 보임.
  - 머신러닝은 데이터 내의 패턴을 자동으로 인식하는 기법으로, 대량의 변수 및 모형을 사용할 수 있어 예측력을 높일 수 있음.
- 이에 따라 본 연구는 빅데이터 기반의 전망과 전통적 거시전망을 종합적으로 분석하여 향후 경제 전망 연구 방향에 대해 논의
  - 빅데이터를 활용하여 한국의 경기 상황을 대표하는 GDP 성장률을 단기 전망하고, 전통적 통계 모형 및 구조적 거시모형의 전망 결과와 비교하여 예측력이 향상되는지 분석

## 2. 조사 및 분석 결과

### 1) 빅데이터와 머신러닝을 이용한 경제성장률 전망

- 대량의 정형 데이터를 기반으로 한국과 미국의 실질 GDP 성장률을 전망한 결과, 빅데이터를 활용함으로써 예측력이 향상되었으며 이는 전통적 계량경제 모형보다 머신러닝을 통해 성과가 나타남.

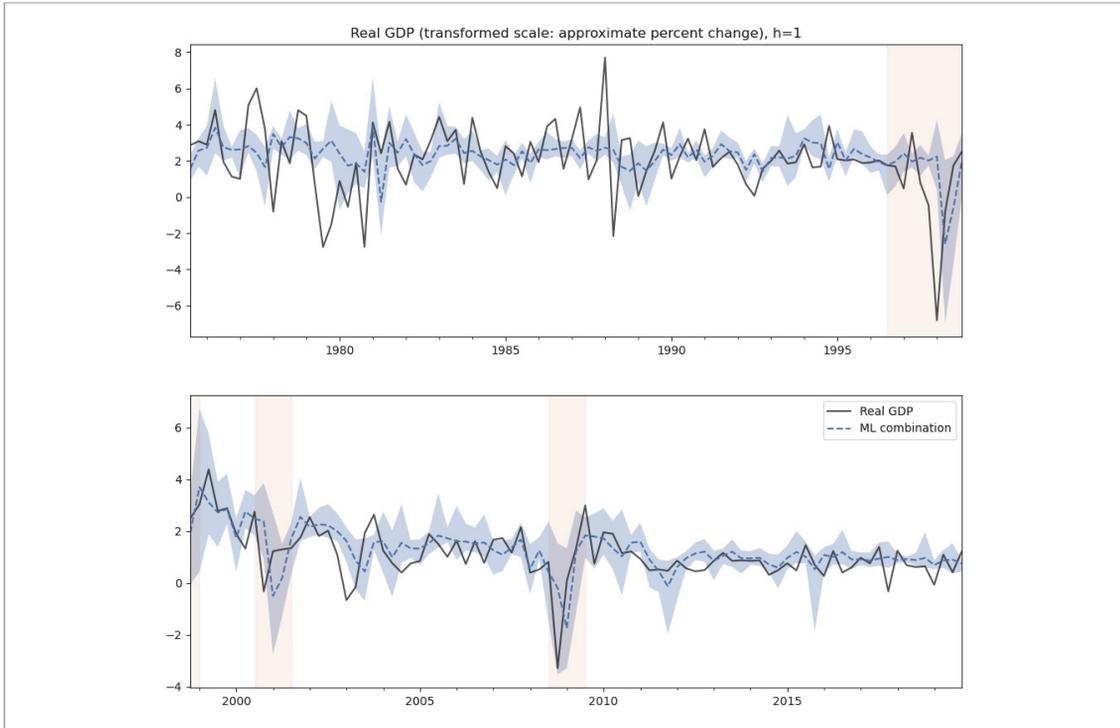
- 각국의 다양한 거시·금융 지표 200여 개를 가용하고 머신러닝 기법 세 가지, 혼합형 모형 다섯 가지, 머신러닝 예측조합과 전통적 계량경제 전망모형을 추정하여 GDP 성장률을 전망
  - 미국은 1959년 1/4분기부터 2023년 1/4분기까지 총 230개의 거시·금융 지표, 한국은 1960년 1/4분기부터 2023년 1/4분기까지 총 210개의 거시·금융 지표를 활용
  - 머신러닝은 선행연구에서 예측 성과가 좋았던 랜덤포레스트(Random Forest), 엑스지부스트(XGBoost)와 LSTM(Long Short-Term Memory)을 사용
  - 혼합형 모형은 두 가지 머신러닝 기법을 각각 예측변수 선별과 전망에 사용한 방법론으로 본 연구에서는 LASSO/OLS, LASSO/RF, LASSO/XGBoost, RF/OLS, RF/LASSO를 사용
  - 전통적 계량경제 전망모형은 자기회귀모형(AR: Autoregressive Model)과 대량의 예측변수를 사용할 수 있는 동적요인모형(DFM: Dynamic Factor Model), Stock and Watson(2002)의 Diffusion Index 모형을 사용
- 공통요인을 사용하여 대량의 예측변수를 다룰 수 있는 계량경제 모형보다 머신러닝이 대체로 좋은 성과를 보여줌.

● 미국의 GDP 성장률에 대한 머신러닝의 예측력이 전반적으로 전통적 계량경제 모형보다 뛰어난 데 비해, 한국의 경우 2000년 이전에는 예측 우위가 뚜렷하지 않음.

- 양국 모두 1분기 후 GDP 성장률에 대한 머신러닝 전망이 기준모형보다 유의미하게 예측력이 뛰어나지만, 전망시계가 길어지면서 예측력이 떨어지는 양상을 나타냄.
  - 미국의 경우 금융위기 기간을 제외한 시기의 1분기 후 GDP 성장률에 대한 머신러닝 전망치가 AR(4)보다 유의미하게 예측력이 뛰어남.
- 한국의 경우 랜덤포레스트, LASSO/RF와 머신러닝 예측조합의 예측 성과가 우수한 것으로 나타남.
  - 한국은 머신러닝과 DFM의 예측력이 비슷하나 미국은 DFM이 머신러닝보다 후행 전망하는 모습을 보임.

● 빅데이터를 활용한 머신러닝의 예측 성과가 우수한 편이나 본 연구에서 채택한 롤링윈도(rolling window)와 훈련 및 검증 집합의 표본 수에 대해 구한 결과이며, 전망시계, 국가, 시기에 따라 예측력이 달라질 수 있음을 주의

그림 1. 머신러닝 예측조합의 한국 GDP 성장률 예측



주: 점선은 머신러닝 및 혼합형 모형(Random Forest, XGBoost, LSTM, LASSO/OLS, LASSO/RF, LASSO/XGBoost, RF/LASSO) 7개의 전망치 평균을 나타내며, 점선을 포함한 구간은 7개 전망치의 최솟값과 최댓값을 나타냄.  
 자료: 한국은행, 통계청, 에너지경제연구원, 국토교통부, BIS, IMF, OECD, Bloomberg, CEIC(검색일: 2023. 7. 24)를 이용해 추정한 결과에 기반하여 저자 작성. 자료에 관한 자세한 내용은 연구보고서의 [부록 표 3]을 참고.

## 2) 비정형 데이터를 이용한 경제성장률 전망

- 본 연구에서는 네이버 검색 데이터를 비정형 데이터로 정의하고, Koop and Onorante(2019)의 동적모형 평균화 및 선택(DMA 및 DMS)을 이용하여 한국의 GDP 성장률을 전망
  - Choi and Varian(2012)의 선구적인 연구 이후 인터넷 검색 데이터를 이용한 현재 예측(nowcasting)은 시의성 있는 거시경제지표를 찾는 정책 입안자들 사이에서 주목받고 있음.
  - Koop and Onorante(2019)는 구글 검색 데이터를 기존 거시변수들과 함께 사용하면 거시경제에 대한 전체적인 예측이 개선됨을 보여줌.
  - 본 연구에서는 네이버 검색량이 한국 경제의 국면에 대한 집단적 지식을 제공한다고 가정하고, GDP 성장률 예측변수로 통상적으로 사용되었던 거시·금융 지표 8개를 기준으로 각 변수와 관련된 네이버 검색량을 표준화하여 검색지수를 구축
    - 2016년 1월 1일부터 2023년 6월 30일까지 수집된 네이버 검색 데이터를 주성분 분석을 통해 8개의 거시변수에 매칭하여 총 8개의 거시변수와 8개의 네이버 검색지수를 예측변수로 사용

- 온라인 검색 데이터가 특정 시점의 경제성장률 예측에 가장 기여도가 높은 변수를 선정하는 데 유용하다는 아이디어를 기반으로 하여 동적모형의 선택 확률로 네이버 검색지수를 사용
- 한국 GDP 성장률을 전망하는 데 네이버 검색지수를 사용한 DMA와 DMS가 검색지수를 사용하지 않은 OLS, AR(2), 임의보행 모형들보다 예측력이 향상
  - 또한 거시변수를 보조하기 위해 네이버 검색지수를 추가한 DMA와 DMS가 거시변수만을 사용한 결과보다 예측력이 개선되어 온라인 검색 데이터가 경제전망에 유용한 정보를 제공할 수 있음을 나타냄.
  - 기존의 전망모형들이 평균 회귀적인 성향을 보이는 데 비해 네이버 검색지수를 사용한 DMA와 DMS는 변곡점들을 잘 예측하는 것으로 나타나 경기변동 시점에 유용하게 사용될 것으로 보임.
  - 네이버 검색지수 8개 중 DMA 모형에 꾸준히 높은 확률로 기여하는 지수는 실업률, 금리 스프레드와 산업생산임.
- 온라인 검색 결과 자체가 정확한 경제 정보를 반영하는 것은 아니지만, 기존의 거시경제 데이터가 추적하지 못하는 여론의 흐름 혹은 소비자들의 경제심리 등을 반영할 수 있으므로 경제 전망모형에 보조적인 지표로 활용된다면 예측력 향상을 기대할 수 있음.

표 1. DMA와 DMS의 예측 결과

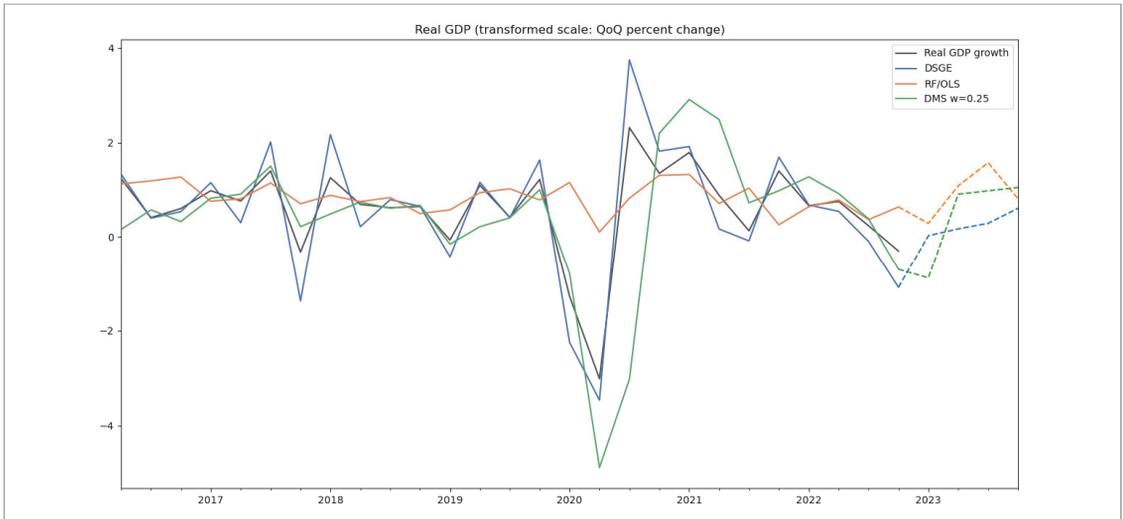
		$\lambda = \alpha = 0.95$		$\lambda = \alpha = 0.97$		$\lambda = \alpha = 0.99$	
		DMA	DMS	DMA	DMS	DMA	DMS
검색지수 이용 안함	AR(2)	32.84	-	-	-	-	-
	OLS	1.57**	-	-	-	-	-
	R.W.	0.91	-	-	-	-	-
	$\omega = 1$	0.79*	0.76**	0.73**	0.66**	0.69**	0.65**
검색지수 이용	$\omega = 0.75$	0.84*	0.74**	0.78*	0.71**	0.73**	0.69**
	$\omega = 0.5$	0.80*	0.74**	0.75**	0.71*	0.71**	0.69**
	$\omega = 0.25$	0.77**	0.71**	0.73**	0.68**	0.69**	0.65**
검색지수만 이용	$\omega = 0$	0.75**	0.71**	0.72**	0.68**	0.68**	0.65**
	OLS	1.82*	-	-	-	-	-

자료: 한국은행(검색일: 2023. 12. 20); 통계청(검색일: 2023. 12. 20); 네이버 데이터랩(검색일: 2023. 9. 9)을 이용하여 추정된 결과에 기반하여 저자 작성.

### 3) 구조적 거시모형과 빅데이터 기반의 전망 취합

- 본 연구에서는 전망뿐만 아니라 경제 정책 효과를 분석할 수 있는 구조적 거시모형인 동태확률일반균형 (DSGE: Dynamic Stochastic General Equilibrium) 모형을 사용해 한국의 GDP 성장률을 예측
  - 이론적 일관성이 높은 DSGE 모형과 달리 데이터 일관성이 높은 머신러닝, DMA와 DMS 모형은 경제 정책 변화에 따른 충격 반응 분석을 할 수 없다는 한계점이 있음.
  - 다만 본 연구에서 추정한 DSGE 모형의 GDP 성장률 예측값은 표본 내 전망치이기 때문에 빅데이터 기반의 표본 외 전망치와 직접적인 비교를 하기 어려움.
- 2016년 2/4분기부터 2022년 4/4분기까지의 유사 표본 외 자승평방근오차(RMSE)를 통해 예측력을 비교하면 대량의 정형 데이터를 사용한 머신러닝이 네이버 검색지수를 사용한 DMA 및 DMS보다 우수해보이나 통계적으로 유의미하지 않음.
  - 실제 경제성장률과 각 전망모형의 전망치를 그림으로 표시하면 DMA 및 DMS 전망치가 2020년에 급락한 GDP 성장률과 비슷한 움직임을 보이지만 머신러닝 전망치는 후행하는 모습을 나타냄.
  - DSGE 모형의 전망치는 실젯값과 비슷하지만, 표본 내 전망치이기 때문에 예측력이 더 뛰어나다고 단언할 수 없음.

그림 2. 2016Q2~2022Q4 한국 GDP 성장률(QoQ) 전망 결과



주: 전망모형 추정에 사용한 데이터의 구성 및 시작 시점은 각각 다르지만 끝나는 시점은 2022년 4/4분기로 동일함. 실선은 유사 표본 외 전망치(RF/OLS, DMS)와 표본 내 전망치(DSGE)이며, 점선은 2023년 1/4분기부터 4/4분기까지 분기별 한국 GDP 성장률의 표본 외 전망치를 나타냄. 표본 내 전망치는 과대 적합 문제로 인해 실젯값과 잘 맞는 모습을 보일 수 있음을 감안해야 함.

자료: 한국은행, e-나라지표, BEA, FRED(모든 자료의 검색일: 2023. 8. 30), 한국은행, 통계청, 에너지경제연구원, 국토교통부, BIS, IMF, OECD, Bloomberg, CEIC(모든 자료의 검색일: 2023. 7. 24), 한국은행, 통계청(모든 자료의 검색일: 2023. 12. 20), 네이버 데이터랩(검색일: 2023. 9. 9)을 이용해 추정된 결과에 기반하여 저자 작성.

### 3. 정책 제언

- 본 연구 결과, 경제성장률 전망에 빅데이터를 활용하면 예측력이 향상되어 다양한 소스의 데이터를 발굴하여 거시경제 데이터베이스를 구축할 필요가 있음을 시사
  - 거시변수만 사용한 전망모형보다 네이버 검색지수를 이용한 전망치가 경제성장률의 급락 및 반등하는 움직임을 잘 예측하는 것으로 나타나 새로운 소스의 데이터가 경제지표 예측에 중요한 요소임을 짐작할 수 있음.
  - 온라인 검색 데이터는 소비자들의 경제심리를 실시간으로 반영하여 급변하는 경기 상황을 전망하는 데 유용하므로 다양한 소스의 자료를 발굴하여 적극적으로 활용할 필요가 있음.
- 경제 전망에 쓰일 데이터베이스 구축과 더불어 전망에 활용할 수 있는 새로운 분석 기법을 개발하고 연구할 필요가 있음.
  - 대량의 데이터를 사용할 수 있는 전통적 계량경제 방법론(DFM, Diffusion Index 등)이 이미 널리 사용되고 있지만, 본 연구에서는 같은 데이터를 사용한 머신러닝의 예측력이 상대적으로 뛰어난 것을 확인
  - 머신러닝으로 도출한 경제전망이 기존의 전망모형 및 경제 전망 전문가의 정성적 판단을 완전히 대체하기 어렵지만, 최종 전망 또는 경기 상황을 참고하는 보조지표를 구축하는 데 활용할 수 있음.
- 본 연구와 관련하여 향후 추진할 수 있는 경제 전망 후속 연구를 제안
  - 단기 과제로 첫째, 본 연구에서 사용한 비정형 및 정형 데이터를 결합한 데이터베이스를 기반으로 전망 또는 경기 상황을 판단할 수 있는 지표를 구축
  - 두 번째 단기 과제는 구조적 거시모형과 빅데이터 활용 방법론을 결합한 전망 모형에 관한 연구
  - 장기 과제로 검색 엔진뿐만 아니라 소셜미디어, 온라인 뉴스 등의 데이터에 텍스트 마이닝 기법을 사용하여 거시경제 전망에 필요한 데이터베이스 및 플랫폼을 구축하면 의미 있는 기초자료가 될 것임. **KIEP**