

Jeonbuk Institute

기본연구

2022-09

# 뉴스 빅데이터를 활용한 전북지역경기지수 개발 연구

A Study on Developing Jeonbuk Economic Indicators by Utilizing News Big Data

이강진 최성환 김민경



## 설립목적

전라북도 및 전북지역 시군의 지역발전과 여성정책 등에 관련된 정책과제에 대하여 체계적인 조사·연구 활동을 통하여 지역단위의 정책개발 기능을 수행함으로써 지역 발전에 기여

## 주요기능

- 도정에 관한 중장기 개발계획 및 주요현안에 대한 조사 연구
- 지역경제, 지역발전, 지역여성에 관한 연구 및 정책 대안의 모색
- 정부, 전라북도 및 국내외의 연구기관과 민간단체의 각종 용역 수탁
- 연구 관련도서 및 간행물 출간
- 국내외 연구기관 간 공동연구 및 정보 교류 협력

## 연구진 소개

### 이강진

하와이대학교 경제학박사  
전북연구원 선임연구위원

### 최성환

한양대학교 도시공학(지역경제전공) 박사  
무한지식공작소 팀장

### 김민경

원광대학교 이학석사  
전북연구원 전문연구원

Jeonbuk Institute

기본연구

2022-09

# 뉴스 빅데이터를 활용한 전북지역경기지수 개발 연구

A Study on Developing Jeonbuk Economic Indicators by Utilizing News Big Data

이강진 최성환 김민경



## 연구진 및 연구 세부 분담

---

연구책임 이강진 | 선임연구위원 | 연구총괄, 제1장 ~ 5장  
공동연구 최성환 | 초빙연구위원 | 제3장 일부, 4장 일부, 5장 일부  
김민경 | 전문연구원 | 제2장, 3장 일부

---

자문위원 김주완 | 전) 전라북도 경제정책전문위원  
손영두 | 동국대학교 교수

---

연구관리 코드 : 22GI03

이 보고서의 내용은 연구자의 의견으로서  
전북연구원의 공식 입장과는 다를 수 있습니다.



## 1. 연구목적 및 방법

### 1) 연구 배경 및 목적

- 최근 빅데이터에 대한 관심이 고조되고 있는 가운데 매일 생산되는 미디어 데이터는 사회 다양한 분야에서 종합적으로 작성되고 그 양이 매우 방대하며 데이터 주기가 매우 빨라 경기현황을 파악하는데 매우 유용하게 활용될 수 있다.
- 각 신문 및 방송사들의 데이터가 인터넷 플랫폼에서 접근이 가능하도록 포털기능이 강화되어 데이터서비스가 용이하게 되어 분석을 위한 데이터베이스 구축이 가능하게 됨에 따라 뉴스 빅데이터를 활용한 지역경기지수 개발이 가능하게 되었다.
- 전라북도에서도 전라북도동행경기종합지수에 대한 활용도 제고 및 경기 판단의 고도화를 위해 보조적으로 활용될 경기지표 작성이 요구되고 있어, 본 연구에서는 매일 작성되는 신문, 방송 등 미디어의 뉴스 빅데이터를 활용하여 지역경기지수를 작성하고자 한다.

### 2) 연구 범위 및 방법

- 본 연구과제 수행을 위해서 데이터의 취득범위는 2011년 12월 1일부터 2021년 12월 31일로 설정하고, 분석 자료는 월간 데이터를 작성했다. 연구의 공간적 범위는 전라북도, 데이터 수집 범위 또한 전라북도 주요 언론사 기사를 대상으로 한다.
- 본 연구는 지역에서는 미개척 분야이고 지자체에서는 최초로 시도되는 연구로 국내외 연구사례 검토를 통해 연구범위를 선정하고 필요한 데이터는 Python을 이용한 web crawling 프로그램을 직접 만들어 지역 언론사에 접속하여 확보하였다. 뉴스경기종합지수 작성을 위해 BSI 방식과 통계청에서 제공하는 KOSTAT-CIS 프로그램을 이용한 Composite Index 작성방식을 활용하여 두 결과를 비교하였다. 유용성 검증을 위해 상관분석과 전환점 분석 등 다양한 계량분석을 수행한다.

## 2. 결론 및 정책제언

### 1) 결론

- 전북 뉴스경기종합지수 작성을 위해 새전북신문, 전북도민일보, 전북일보, 전북중앙신문, JTV전주방송, 전주MBC의 기사 데이터를 구축하였다. 기사는 Python을 활용한 뉴스기사 스크래핑 프로그램을 자체적으로 만들어 수집했다.
- 경기종합지수의 작성은 BSI 방식과 Composite Index 방식의 2가지를 사용하였다. 전북 뉴스경기종합동행지수 BSI 작성방법은 이금희(2017)의 연구와 같이 호황지표와 불황지표의 차를 호황지표와 불황지표의 합으로 나누는 BSI 방식을 적용했다.
- Composite Index 방식의 전북 뉴스경기종합지수 작성과정은 통계청의 지역경기종합지수 실무가이드를 따라서 작성했다. CI 작성은 구성지표의 계절 및 불규칙조정 → 전월비 증감률 산출 → 증감률 진폭의 표준화 → 종합증감률 산출 → 추세조정지수 산출 → 경기종합지수 산출 → 동행지수 순환변동치 산출 등의 과정을 통해 결과를 도출한다.
- 뉴스경기종합지수의 유용성을 판단하기 위해 순환변동치를 구하고 전환점 분석을 했다. 전북 경기동행종합지수 순환변동치와 비교해 볼 때 양 지수가 근접하게 움직이는 것을 볼 수 있다. 다만, 2013년 7월에서 2014년 1월까지의 이상치와 유사한 형태의 진폭, 2019년 10월에서 2020년 5월은 설명하기 어려울 만큼 변동폭이 큰 것을 볼 수 있다. 전환점 분석결과 전북 경기동행종합지수의 전환점이 가장 적게 나타났고 BSI는 경제순환주기가 가장 짧았다. 전북 경기동행종합지수와 비교할 때 2014년 2월의 정점은 7개월 선형, 2015년 5월의 저점은 5개월 선형하는 것을 볼 수 있었다.

### 2) 정책제언

- BSI 모델과 Composite Index 모델을 통해서 알아본 뉴스빅데이터를 활용한 지역경지 지수 개발은 소기의 성과를 거두었다. 순환변동치의 국면전환의 유사성을 찾을 수 있었으며 뉴스경기종합지수의 선형성은 향후 지역경기 동향 파악의 중요한 자료로 활용될 수 있을 것이다. 통상적으로 경기지수개발 후 6개월에서 1년 정도는 모델의 안정성과 경기대응력을 판단하기 위한 테스트 및 모니터링 기간을 갖게 되므로 지속적인 업데이트를 통해 시계열을 확보하고 데이터의 안정성 방안을 모색하여 모델의 고도화가 필요하다.





---

# 차 례

## CONTENTS

---

요 약 ..... i

---

### 제1장 서 론

1. 연구 배경 및 목적 ..... 3  
가. 연구 배경 ..... 3  
나. 연구 목적 ..... 4  
2. 연구 내용 및 방법 ..... 5  
가. 연구 내용 ..... 5  
나. 연구 방법 ..... 9

---

### 제2장 뉴스 빅데이터의 활용방안

1. 뉴스 빅데이터를 활용한 경제분석 사례 ..... 13  
가. 정부 및 공공기관의 빅데이터 활용 사례 ..... 13  
나. 소셜네트워크 빅데이터의 경제 분석 활용 ..... 14  
다. 뉴스 빅데이터의 경제 분석 활용 ..... 15  
2. 뉴스 빅데이터의 특성 선행연구 검토 ..... 17  
가. 국내 연구 ..... 17  
나. 국외 연구 ..... 20

**제3장      뉴스 빅데이터 지역경기지수 작성**

- 1. 경기관련 뉴스 키워드 데이터 수집 ..... 23
  - 가. 뉴스 데이터 수집 대상 신문사 ..... 23
  - 나. 경기 관련 검색어의 선정 ..... 23
  - 다. 뉴스 데이터 수집 방법 ..... 24
- 2. 뉴스 데이터 수집 및 통계 분석 결과 ..... 26
  - 가. 뉴스 데이터 수집 결과 ..... 26
  - 나. 신문사별 뉴스기사 시계열 추이 ..... 34
  - 다. 뉴스 키워드 데이터 수집 및 결과 종합 ..... 39
- 3. 뉴스 빅데이터 지역경기지수 작성 ..... 40
  - 가. 개발 전체 ..... 40
  - 나. 개발 과정 ..... 41
  - 다. 뉴스 데이터와 경기지수 상관분석 ..... 42
  - 라. 시계열 조정 ..... 47
  - 마. 전북뉴스 지역경기(종합)지수 작성 ..... 50

**제4장      뉴스 빅데이터 지역경기지수 유용성 검정**

- 1. 경기지수의 순환변동치 작성 및 유용성 점검 ..... 61
  - 가. 경기순환 판단을 위한 보조지표로서의 순환변동치 작성 ..... 61
  - 나. 유용성 검증 ..... 62

---

## 차 례

### CONTENTS

---

|                           |    |
|---------------------------|----|
| 2. 주요 경기국면에서 전환점 분석 ..... | 65 |
| 가. 전환점 분석 방법 .....        | 65 |
| 나. 전환점 분석 결과 .....        | 65 |

---

## 제5장      결론 및 향후과제

|                                |    |
|--------------------------------|----|
| 1. 결론 .....                    | 69 |
| 2. 향후과제 .....                  | 71 |
| 가. 뉴스 데이터 구축 체계화 .....         | 71 |
| 나. 경기지표 개발 및 빅데이터 활용도 제고 ..... | 72 |

|            |    |
|------------|----|
| 참고문헌 ..... | 73 |
|------------|----|

|                      |    |
|----------------------|----|
| 영문요약 (Summary) ..... | 76 |
|----------------------|----|

|           |    |
|-----------|----|
| 부 록 ..... | 81 |
|-----------|----|

## 표 차례

### LIST OF TABLEES

|  |    |
|--|----|
| [표 1-1] 한국은행에서 작성하는 주요 조사 내용 .....             | 6  |
| [표 1-2] 뉴스 빅데이터 데이터베이스 구축을 위한 검색 연관어(예시) ..... | 7  |
| [표 2-1] 정부 및 공공기관의 빅데이터 활용 내용 .....            | 14 |
| [표 2-2] 빅데이터를 이용한 경제 분석 관련 선행연구 .....          | 15 |
| [표 2-3] 뉴스 빅데이터를 이용한 경기지수 분석 관련 선행연구 .....     | 16 |
| [표 3-1] 경기 관련 수집대상 경제용어 .....                  | 24 |
| [표 3-2] 6개 지역신문의 경기변동 관련 경제용어 웹 스크래핑 결과 .....  | 26 |
| [표 3-3] 새전북신문 불황 관련 용어 스크래핑 결과 .....           | 27 |
| [표 3-4] 새전북신문 호황 관련 용어 스크래핑 결과 .....           | 27 |
| [표 3-5] 새전북신문 경제 관련 용어 스크래핑 결과 .....           | 28 |
| [표 3-6] 전북도민일보 불황 관련 용어 스크래핑 결과 .....          | 28 |
| [표 3-7] 전북도민일보 호황 관련 용어 스크래핑 결과 .....          | 29 |
| [표 3-8] 전북도민일보 경제 관련 용어 스크래핑 결과 .....          | 29 |
| [표 3-9] JTV전주방송 불황 관련 용어 스크래핑 결과 .....         | 30 |
| [표 3-10] JTV전주방송 호황 관련 용어 스크래핑 결과 .....        | 30 |
| [표 3-11] JTV전주방송 경제 관련 용어 스크래핑 결과 .....        | 30 |
| [표 3-12] 전북일보 불황 관련 용어 스크래핑 결과 .....           | 30 |
| [표 3-13] 전북일보 호황 관련 용어 스크래핑 결과 .....           | 31 |
| [표 3-14] 전북일보 경제 관련 용어 스크래핑 결과 .....           | 31 |
| [표 3-15] 전북중앙신문 불황 관련 용어 스크래핑 결과 .....         | 32 |
| [표 3-16] 전북중앙신문 호황 관련 용어 스크래핑 결과 .....         | 32 |
| [표 3-17] 전북중앙신문 경제 관련 용어 스크래핑 결과 .....         | 32 |
| [표 3-18] 전주MBC 불황 관련 용어 스크래핑 결과 .....          | 33 |
| [표 3-19] 전주MBC 호황 관련 용어 스크래핑 결과 .....          | 33 |
| [표 3-20] 전주MBC 경제 관련 용어 스크래핑 결과 .....          | 33 |
| [표 3-21] 불황 관련 용어 상관관계 분석 결과 .....             | 35 |
| [표 3-22] 불황 관련 경제용어 상관성이 가장 높은 신문사 묶음 .....    | 35 |

---

## 표 차례

### LIST OF TABLES

---

|  |    |
|--|----|
| [표 3-23] 호황 관련 용어 상관관계 분석 결과 .....                       | 36 |
| [표 3-24] 호황 관련 경제용어 상관성이 가장 높은 신문사 묶음 .....              | 37 |
| [표 3-25] 경제 관련 용어 상관관계 분석 결과 .....                       | 38 |
| [표 3-26] 경제 관련 경제용어 상관성이 가장 높은 신문사 묶음 .....              | 38 |
| [표 3-27] 상관분석 결과 요약 .....                                | 43 |
| [표 3-28] 상관계수 0.4 이상인 경기지수별 단어 개수와 단어별 상관계수 .....        | 44 |
| [표 3-29] 조건에 부합하는 단어와 상관계수 .....                         | 45 |
| [표 3-30] 불황 중심 단어쌍 .....                                 | 45 |
| [표 3-31] 불황 단어와 호황 단어 상관계수 값 차이 .....                    | 46 |
| [표 3-32] 최종 선정 키워드 집합 .....                              | 46 |
| [표 3-33] 전북뉴스 지역경기지수(BSI) 산출 결과 .....                    | 50 |
| [표 3-34] 전북뉴스 지역경기지수(BSI)와 각 지수와의 상관관계 .....             | 51 |
| [표 3-35] 전북뉴스 지역경기종합지수 산출 결과 .....                       | 56 |
| <br>   |    |
| [표 4-1] 전북뉴스 지역경기종합지수 순환변동치 산출 결과 .....                  | 61 |
| [표 4-2] 전북뉴스 지역경기종합지수(CI)와 각 지수와의 상관관계 .....             | 63 |
| [표 4-3] NBER Bry-Bochan 방법에 의한 Turning Point 산출 결과 ..... | 65 |

---

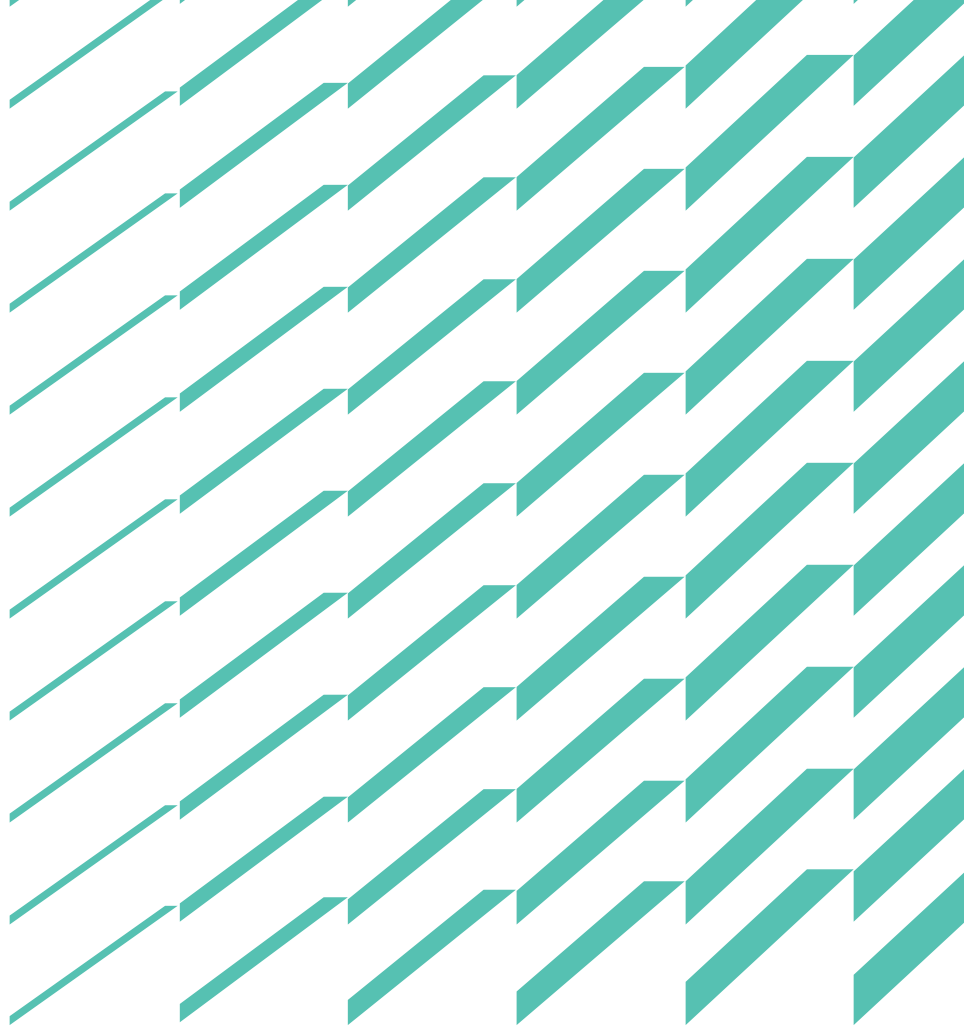
## 그림 차례

### LIST OF FIGURES

|  |    |
|--|----|
| [그림 1-1] 연구의 흐름 .....                            | 8  |
| [그림 3-1] Jupyter notebook을 이용한 웹 스크래핑 프로그램 ..... | 25 |
| [그림 3-2] 6개 신문사 불황 관련 기사 수 추이 .....              | 34 |
| [그림 3-3] 6개 신문사 호황 관련 기사 수 추이 .....              | 36 |
| [그림 3-4] 6개 신문사 경제 관련 기사 수 추이 .....              | 37 |
| [그림 3-5] 전북뉴스경기종합지수 개발 FrameWork .....           | 41 |
| [그림 3-6] 상관분석을 위한 데이터 구축 방식(일부 표현) .....         | 42 |
| [그림 3-7] 경기침체 변화율 .....                          | 47 |
| [그림 3-8] 해고 변화율 .....                            | 47 |
| [그림 3-9] 인플레이션 변화율 .....                         | 48 |
| [그림 3-10] 금리인하 변화율 .....                         | 48 |
| [그림 3-11] 채용 변화율 .....                           | 48 |
| [그림 3-12] 경기회복 변화율 .....                         | 49 |
| [그림 3-13] 디스플레이션 변화율 .....                       | 49 |
| [그림 3-14] 금리인상 변화율 .....                         | 49 |
| [그림 3-15] 전북뉴스 지역경기지수(BSI) .....                 | 50 |
| [그림 3-16] 전북뉴스 지역경기지수(BSI)와 동행지수 순환변동치 .....     | 51 |
| [그림 3-17] 전북뉴스 지역경기종합지수 작성 흐름도 .....             | 53 |
| [그림 3-18] 전북뉴스 지역경기종합지수 .....                    | 57 |
| [그림 4-1] 전북뉴스 지역경기종합지수와 동행지수 순환변동치 .....         | 62 |
| [그림 4-2] 뉴스경기종합지수와 상관계수 분석 .....                 | 63 |



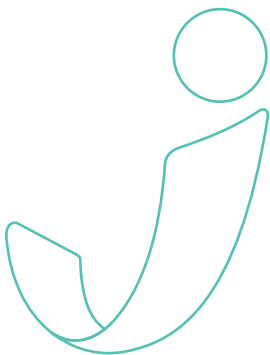




# 제 1 장

## 서론

1. 연구배경 및 목적
2. 연구내용 및 방법





---

# 제1장 서론

## 1. 연구 배경 및 목적

### 가. 연구 배경

#### 1) 지역 경기판단의 보조지표 작성 요구

각 지자체에서는 지역경기종합지수 작성을 통해 지역경제 현황을 진단하는 지표로 활용하고 있으며 지자체마다 편차가 있지만 지역경기종합지수를 지역경제의 단중기 정책의 기초자료로 활용하고 있다.

매월 작성되는 지역경기종합지수는 지역의 주요 경제지표의 composite index로써 획득 가능한 자료를 최대한 활용하여 작성하고 있으나 국가차원에서 작성하는 경기종합지수에 비해 활용 가능한 데이터가 부족하여 경기종합지수의 고도화를 위한 대안 모색이 요구되고 있다.

지역경기 현황을 파악하기 위해서는 다양한 부문에서 진단이 필요하나 지역에서는 단기 지역경기현황을 파악할 수 있는 지표로 지역경기종합지수가 유일하여 이를 보완할 수 있는 지수 개발에 대한 필요성이 요구된다.

#### 2) 뉴스 빅데이터에 대한 활용도 제고 필요성 대두

2019년 신문 산업 현황을 보면 문체부 등록 간행물은 12,702개(일간신문 535개, 주간신문 3,279개, 인터넷신문 8,888개)이며 실제 발행여부 확인된 신문사는 5,960개(일간신문 262개, 주간신문 1,416개, 인터넷신문 4,282개)로 나타났다.

최근, 디지털 사회로의 전환이 가속화됨에 따라 다양한 데이터가 디지털화 되고 이에 대한 관리가 체계화되고 있으며 사회의 복잡성이 증가함에 따라 작성되는 데이터양이 급격히 증가하고 있어 빅데이터의 작성, 관리, 활용 등 다양한 부문에서 중요성이 부각되고 있다.

---

빅데이터의 특성은 일반적으로 저장, 관리, 분석을 기존 방식으로 수행하기 어려울 정도로 그 규모가 크며, 데이터 작성 주기 또는 순환 속도가 빠르고, 데이터 형식이 숫자, 문자, 사진, 음성 및 영상 등으로 다양하여 기존 방식으로 관리 및 분석이 어려운 데이터를 의미한다(이공희 외, 2014).

다양한 빅데이터 가운데 지역경기지수에 활용할 수 있는 부문이 신문, 방송 등 미디어에서 생산되는 빅데이터로 매일 작성되는 미디어 데이터는 사회 다양한 분야에서 종합적으로 작성되고 그 양이 매우 방대하며 데이터 주기가 매우 빨라 경기현황을 파악하는데 매우 유용한 것으로 나타났다.

각 신문 및 방송사들의 데이터가 인터넷 플랫폼에서 접근이 가능하도록 포털기능이 강화되어 데이터서비스가 용이하게 되어 분석을 위한 데이터베이스 구축이 가능하게 됨에 따라 뉴스 빅데이터를 활용한 지역경기지수 개발이 가능하게 되었다.

## 나. 연구 목적

전라북도를 포함한 지역연구원에서 경기동행종합지수를 작성하고 있으며 지역별로 특성이 있으나 통계청에서 공표하고 있는 승인통계 중심으로 작성되고 있어 자료의 신뢰성은 있으나 데이터의 속보성, 다양한 지역지표의 발굴에 어려움을 겪고 있어 최적의 지수 인지에 대한 고민이 필요한 상황이다.

지역의 경기상황을 보다 정확하게 측정하기 위한 다양한 노력이 필요한 상황에서 최근에 연구가 많이 진행되고 있는 빅데이터를 활용하여 지역의 경기지수를 개발하는 것은 하나의 큰 방향이 될 것을 보인다.

전북에서도 전북 동행경기종합지수에 대한 활용도 제고 및 경기 판단의 고도화를 위해 보조적으로 활용될 경기지표 작성이 요구되고 있어, 본 연구에서는 매일 작성되는 신문, 방송 등 미디어의 뉴스 빅데이터를 활용하여 지역경기지수를 작성하고자 한다.

---

## 2. 연구 내용 및 방법

### 가. 연구 내용

#### 1) 연구범위

본 연구의 시간적 범위 설정에 있어 고려할 사항은 경기지수를 작성하기 위한 충분한 시계열의 확보와 더불어 지역 언론사의 기사를 취득할 수 있는 기간, 그리고 지역 언론사를 얼마만큼 포함해야 할 것인가를 동시에 고려해야 한다.

경기지수 작성을 위해서는 최대한 많은 데이터를 확보해야 하나 프로그래밍과 같은 기술적인 한계와 지역 언론사에서 자료 취득이 불가능한 경우와 같은 불가항력적인 요인, 그리고 연구과제를 수행하는데 자료 취득에 드는 시간과 비용 등을 고려해 결정하였다.

본 연구과제 수행을 위해서 데이터의 취득범위는 2011년 12월 1일부터 2021년 12월 31일로 설정하고, 분석 자료는 월간 데이터를 작성하는 것으로 하였다.

본 연구의 공간적 범위는 연구목적이 뉴스기사를 이용한 전라북도 경기지수를 작성하는 것이므로 전라북도 전역으로 설정하였고, 데이터 수집 범위 또한 전라북도 주요 언론사 기사를 대상으로 한다.

#### 2) 주요 연구내용

본 연구의 내용적 범위는 지역경기지수 작성에 필요한 뉴스 빅데이터의 취득, 데이터 가공, 정제 및 분석, 경기지수 작성, Simulation을 통한 적정성 검토 등을 주요 연구내용으로 한다.

빅데이터를 활용한 경기지수 작성 사례 검토를 통해 다양한 데이터 취득방법 중 지역에서 활용할 수 있는 사례를 도출하고 현실적으로 적용 가능한 범위에서 데이터를 취득한다. 취득된 비정형데이터를 정형화하는 과정에서 가공 및 정제, 그리고 분석을 통해 데이터의 정합성을 확보한다. 이후 경기지수 개발과 개발된 지수의 비교, 분석 및 simulation을 통해 활용가능성을 검토한다.

## ■ 뉴스 빅데이터를 활용한 지역경기지수 작성

뉴스 빅데이터는 표본조사와 같이 모집단을 대표하도록 설계되어 있지 않아서 모집단을 대표해서 구체적으로 파악하기 어렵지만 조사 응답률 하락 등 조사환경이 열악해지는 현 상황에서 표본조사를 보완할 수 있는 한 대안이 될 수 있다.

단기적인 경기 현황 분석을 위해 일반적으로 활용하는 방법 중 하나는 경제주체의 판단을 물어보는 조사로 대표적으로 한국은행에서 조사하고 있는 기업경기조사와 소비자동향조사가 있다.

[표 1-1] 한국은행에서 작성하는 주요 조사 내용

| 조사명                                    | 작성 내용   |
|--|---|
| 기업경기조사                                 | · 전국 2,862개 법인기업을 대상으로 전기에 비해 업황, 재고, 인력 등의 사정이 나아졌는지 나빠졌는지를 파악하는 조사로, 조사된 결과를 기업경기실사지수(Business Survey Index : BSI)로 요약정리하고 있으며 이 지수는 긍정 응답 업체 수의 비중에서 부정 응답 업체수의 비중을 차감하여 비율로 표현한 후 100을 더해서 작성 |
| 소비자동향조사                                | · 전국 도시 2,200 가구를 대상으로 전기에 비해 생활형편, 가계수입 전망 등이 나아졌는지 파악하는 조사로, 조사된 결과를 소비자동향지수(Consumer Survey Index : CSI)로 요약정리   |
| 경제심리지수(ESI : Economic Sentiment Index) | · 기업과 소비자 모두를 포함한 인간의 경제상황에 대한 심리를 종합적으로 파악하기 위하여 BSI 및 CSI 지수(각각 40개 및 24개) 중 경기 대응성이 높은 7개 항목을 선정하여 이들의 표준화지수를 기중평균한 지수로 월별로 작성   |

뉴스 빅데이터를 활용한 지역경기지수 작성은 구글 또는 네이버 검색엔진을 활용하는데 한계가 있어 뉴스 빅데이터를 활용하여 경기지수를 작성하고자 한다. 네이버 검색엔진을 활용한 경기판단지표 개발(이궁희, 황상필, 2014)은 국가적인 차원에서 가능성을 보여주었으나 지역적 차원에서의 경기지수 개발에는 검색의 지역적 분리, 데이터의 습득 부문에서 한계를 가진다. 검색엔진을 활용한 데이터베이스에서 지역을 분리하는 것은 가능할 것이나 기업의 협조를 얻기가 어려운 상황이며 지역의 상황만을 지역에서 검색한다고 보기에는 한계가 있는 반면 지역 뉴스는 지역의 경기에 초점이 맞추어져 있어 데이터 적합성에서 더 우위에 있을 것으로 보인다. 자료의 습득 부문에서 뉴스 빅데이터는 개별 연구자가 Web crawling 프로그램을 통해 데이터베이스를 구축하는 방법을 사용할 수 있는 반면 검색엔진을 활용한 데이터베이스 구축은 기업으로부터 데이터를 구입해야 하는 비용 부담을 초래하기 때문에 뉴스 빅데이터를 활용한 지역경기지수 작성이 지수를

작성유지관리 하는데 유리하다.

뉴스 빅데이터를 활용한 지역경기지수 작성은 기업경기조사의 BSI와 비슷한 방식으로 뉴스 빅데이터의 불황지수와 호황지수를 결합하여 BSI 방식의 뉴스 빅데이터 경기지수를 작성한다. 뉴스 빅데이터 경기지수의 유용성을 파악하기 위해 불황지수와 호황지수를 월별 지수로 작성하고 BSI, CSI, ESI와 비교하여 유용성을 검증한다.

먼저, 뉴스 빅데이터 데이터베이스 구축을 위한 경기 순환 사이클 관련 연관어 통계 검토 및 연관어 검색을 통한 데이터베이스를 구축한다. 뉴스 빅데이터 데이터베이스 작성을 위한 가장 기초가 되는 경기변동 관련 용어를 경기불황, 경기호황, 경제연관 단어로 구분하여 web-crawling 방법을 활용하여 데이터베이스를 구축하고, 향후 상관분석을 통해 데이터를 정제한다.

[표 1-2] 뉴스 빅데이터 데이터베이스 구축을 위한 검색 연관어(예시)

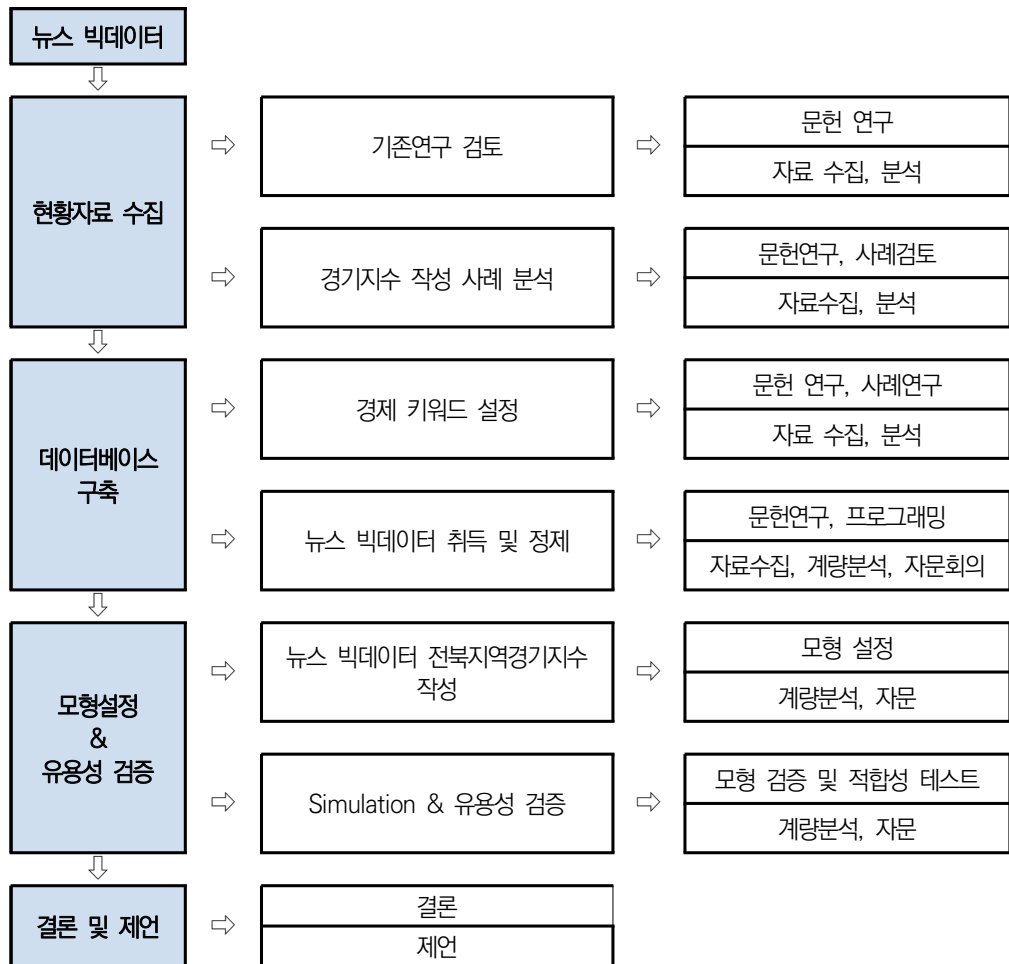
| 경기 불황 용어   | 경기 호황 용어  | 경제 연관 용어   |
|--|---|--|
| 불경기, 불황, 경기침체, 경기불황, 경제불황, 위기, 금융위기, 경제위기, 외환위기, 적자, 부도, 파산, 폭락, 하한가, 신용불량, 정리하고, 해고, 부채, 금리인하, 디플레이션, IMF, 재고증가 | 호경기, 호황, 경기회복, 경기호황, 경제호황, 흑자, 채용, 폭등, 상한가, 거품, 금리인상, 인플레이션, 재고감소, 활황, 수출증가, 고용률 상승, 건조 | 환율, 주가, GDP, 실업률, 물가, 경제성장, 통화정책, 재정정책, 한국은행, 외환보유고, 경상수지, 신용등급, 유가, FRB |

경제관련 단어 53개 키워드 중에서 경기와 관련성이 높은 키워드를 선정하기 위해 다양한 선택 방법론에 대한 검토를 수행할 예정이다. 그 중 하나의 방법은 선행연구와 같이 ESI와의 상관계수 절댓값이 0.5보다 크고 경기선행지수 순환변동치의 상관계수 절댓값이 0.4보다 큰 키워드를 선택하고, 선택된 키워드에 대응되는 키워드, 서로 대비되는 후보 키워드 중 상관계수 값의 차이가 큰 키워드 쌍을 선택하는 등의 방법론과 다양한 선택 방법론에 대한 검토를 수행한다.

뉴스 빅데이터를 활용한 지역경기지수는 월별 뉴스 호황 및 불황 데이터베이스를 바탕으로 경기실사지수(BSI)와 같은 형태로 호황 검색통계의 평균에서 불황 검색통계의 평균을 차감한 후 100을 더한 지수로 작성한다. 경기지수의 계절조정은 호황 검색통계와 불황 검색통계를 각각 나누어서 계절조정 후 결합하는 간접법보다는 일관성 확보를 위해 통계청 프로그램을 이용한 계절조정 방법을 활용한다.

## ■ 뉴스 빅데이터 전북지역경기지수 simulation 및 유용성 검증

전북지역경기지수 순환변동치는 뉴스 빅데이터 경기지수에 H-P 필터를 적용하여 산출한다<sup>1)</sup>. 순환변동치의 유용성은 교차상관분석을 통해 경기지수 순환변동치를 BSI, CSI, 전라북도경기동행지수 순환변동치간 시차 상관관계를 분석하고 경기기준 순환일 기준 전환점 분석을 수행한다.



[그림 1-1] 연구의 흐름

1) H-P 필터와 이와 관련된  $\lambda$ 값은 OECD 및 우리나라 경기종합지수 작성할 때 이용하는 방법이다.



---

## 나. 연구 방법

### 1) 기초자료 수집 및 선행연구 검토

본 연구는 지역에서는 미개척 분야이고 지자체에서는 최초로 연구되는 내용으로 국내외 연구사례 검토를 통해 지역에서 활용할 수 있는 범위를 설정하였다. 경기종합지수 작성을 위한 매뉴얼 검토를 통해 연구의 신뢰도를 높이고자 하였다.

### 2) 데이터 수집 및 경기지수 개발 위한 프로그래밍

본 연구 수행의 가장 기초가 되는 것은 빅데이터 접근성 확보이다. 언론사 데이터베이스에 접속하여 필요한 자료를 취득하는 과정은 시간이 많이 소요되고 기술적인 부분에서 해결해야 할 부문이 많았다.

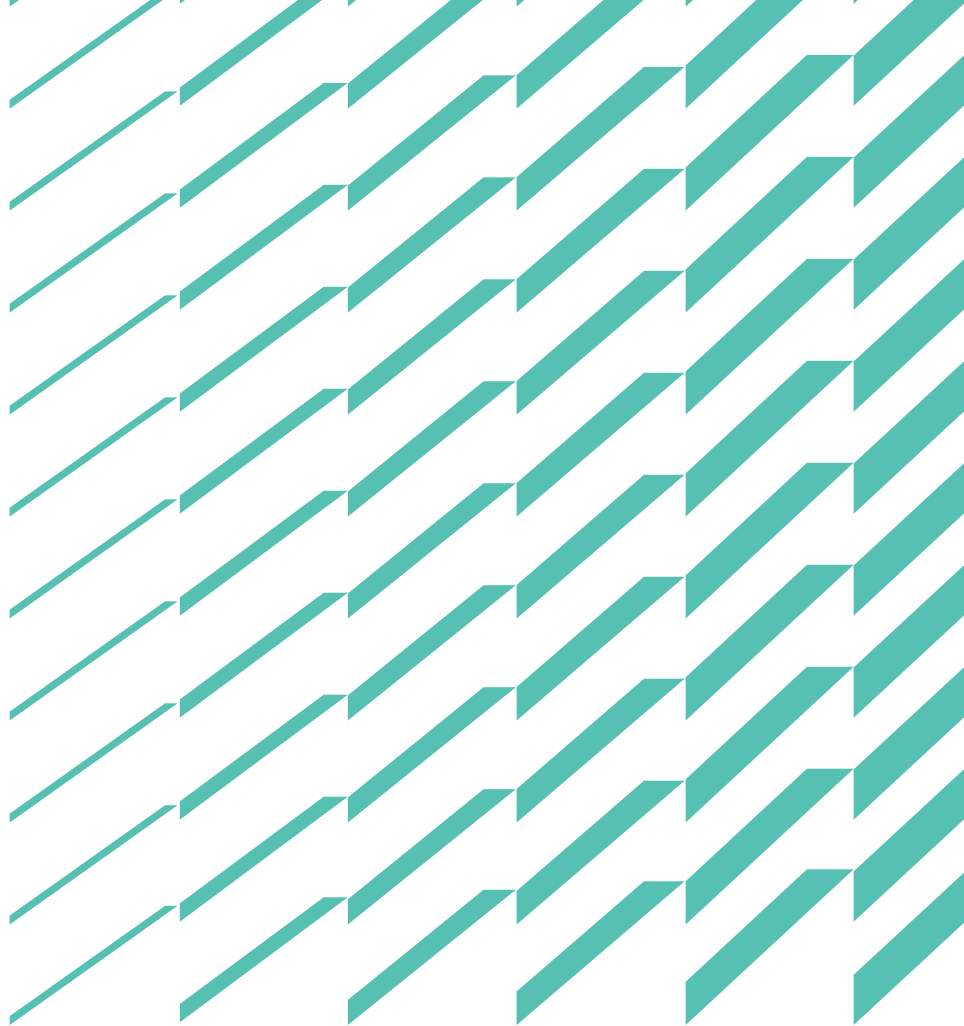
타 연구에서는 기관의 협조 및 데이터 구입으로 빅데이터의 접근이 쉬울 수 있으나 지역 차원에서 지속적 유지관리를 위해서는 추가비용 없이 데이터를 접근할 수 있는 방법이 필요했다. 이를 위해 최근 선호되고 있는 프로그래밍인 파이썬(python)을 활용하여 Web crawling program을 작성하였다. 프로그래밍을 통해 수집한 자료를 정제하여 활용가능한 지표로 전환하였다

수집된 데이터는 2가지 방법을 사용하여 전북뉴스 지역경기(종합)지수를 작성하였다. 하나는 BSI이고 다른 하나는 Composite Index 방법이다. Composite Index 방법은 통계청에서 제공하는 KOSTAT-CIS 프로그램을 활용하여 지수를 개발하고 이에 대한 Simulation을 통해 경기대응력 등을 평가하였다.

### 3) 전문가 자문 수행

프로그램 개발을 위해 사전적으로 프로그램 작성에 대한 학습을 수행하였으며 적절한 프로그램 선정을 위한 전문가 의견 수렴과 프로그램 오류를 해결하기 위한 자문회의를 진행하였다. 통계청에서 제공하는 프로그램에 대한 학습과 통계청 담당자를 비롯한 전문가 의견 수렴을 진행하였다.

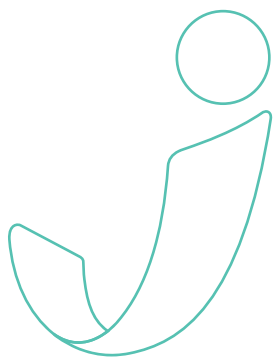




## 제 2 장

### 뉴스 빅데이터의 활용방안

1. 뉴스 빅데이터를 활용한 경제분석 사례
2. 뉴스 빅데이터의 특성 선행연구 검토





---

## 제 2 장 뉴스 빅데이터의 활용방안

### 1. 뉴스 빅데이터를 활용한 경제분석 사례

---

#### 가. 정부 및 공공기관의 빅데이터 활용 사례

뉴스 빅데이터는 통신사, 카드사 등을 중심으로 기업의 수익창출 영역으로 급속하게 확대되고 있고, 정부 및 공공기관들 또한 빅데이터의 생성, 구축, 활용을 위한 활발한 연구로 실용적인 결과를 창출하기 위해 노력하고 있다.

정부는 디지털정부를 표방하고 행정시스템의 디지털화 뿐 아니라 공공데이터의 활용도 제고와 데이터산업을 육성하는 복안을 가지고 있다. 정부의 디지털 관련 산업 육성은 하나의 거스를 수 없는 큰 흐름으로써 향후에도 지속적으로 추진될 것으로 보인다.

빅데이터를 가장 잘 활용하는 분야는 카드사와 통신사로 볼 수 있다. 카드사는 카드 거래 데이터를 보유하고 있어 소비자의 행태의 변화, 신시장의 개척에 중요 분석 자료로 활용하고 카드의 부정사용 방지에도 중요한 역할을 하고 있다.

통신사 또한 통신 빅데이터를 활용하여 유망업종 분석, 소비자 이동 경로, 특정지역 관광객 추산 등을 분석하여 중소벤처기업부와 문화체육관광부 관련 사업을 수행하는데 중요한 분석 자료를 제공하고 있다. 최근 연구 트렌드가 관광객 대상 설문조사에서 빅데이터를 활용한 분석으로 전환되어가고 있는 추세이다.

구글 등에서 사용되는 검색엔진의 검색 키워드를 이용하여 최근 코로나19와 같은 전염병 예측, 민심동향 분석 등 키워드간의 상관분석을 통한 인과관계분석 등을 수행한다.

서울시의 경우에는 2013년도 심야노선버스를 선정하는 과정에서 통신데이터와 택시 승하차 데이터를 활용하는 등 적극적인 빅데이터 활용에 앞장서고 있다.

[표 2-1] 정부 및 공공기관의 빅데이터 활용 내용

| 구분  | 빅데이터 활용 내용   |
|-----|--|
| 정부  | · 정부 및 공공기관이 보유하고 있는 자체 데이터 공개, 데이터 산업 육성                              |
| 카드사 | · 카드 거래 데이터를 활용하여 소비자들의 소비행태 및 시장을 분석하고 신상품 개발, 결제패턴 분석을 통한 카드 부정사용 예방 |
| 통신사 | · 통신 내역 데이터를 활용하여 유망업종 분석, 소비자들의 이동, 특정지역 관광객 추산 등 분석                  |
| 구글  | · 검색엔진의 검색 키워드를 이용한 독감 예측, 소셜네트워크 데이터를 이용한 민심 동향분석                     |
| 서울시 | · 2013년 통신데이터와 택시 승하차 데이터를 활용하여 심야노선버스 선정                              |

온라인 활동에서 얻을 수 있는 빅데이터는 크게 뉴스, 소셜네트워크 텍스트와 검색으로 구분할 수 있다. 매일 작성되는 뉴스는 각 언론사별로 인터넷 서비스를 제공하고 있어 다양한 분석 자료로 활용이 가능하며 전복에서는 전복일보를 비롯한 5개 언론, 방송사가 네이버 기반 서비스를 제공하고 있다. 소셜네트워크 텍스트로는 대표적으로 트위터, 페이스북, 블로그, 카카오톡, 라인 등을 통해 사용자간 주고받거나 게시하는 텍스트가 있고, 이러한 텍스트는 그대로 분석되기 보다는 텍스트마이닝 기법을 적용하여 주요 단어를 추출집계하게 된다.

## 나. 소셜네트워크 빅데이터의 경제 분석 활용

빅데이터를 이용한 경제 분석은 분석 대상 데이터와 분석방법이 기존 분석과 다른 점이 있으며 활용 예로는 구글 검색통계를 이용한 경제 분석과 트위터 게시글을 이용한 경제 분석 등이 있다. 분석 대상 데이터로는 표본조사 등으로 작성된 경제통계보다는 구조화되지 않은 구글 검색 데이터, 트위터 데이터, 거래 데이터 등 경제주체의 온라인 활동 결과로 축적된 데이터가 이용되며, 분석방법은 기존 인과관계 중심의 계량모형보다는 상관관계 분석, 기계학습 분석 방법이 이용된다(Einav & Levin, 2013).

구글 검색 통계를 활용한 연구 분야를 보면 미국 소매판매, 자동차 판매 등의 예측(Choi & Varian, 2012), 소매판매 데이터와의 상관분석을 통한 영국의 소비패턴 분석(Chamberlin, 2010), 영국의 노동 시장 및 부동산 시장 분석(McLaren, 2011), 각국의 실업률 분석(Zimmermann, 2009; D'Amuri, 2009; UN Global Pulse), 이스라엘의 경기변동 분석(Suhoy, 2009), 실물지표와의 대응성 및 예측력 분석(김지은, 2013) 등 미시적 분석 뿐 아니라 거시경제 분석 등 다양한 분야에서 연구가 진행되었다.

검색엔진과 소셜네트워크 데이터의 강점은 데이터의 속보성과 방대한 양의 데이터, 그리고 다양한 형태로의 가공이 가능한 유연성이 큰 장점으로 작용하고 있어 향후 이 분야의 관심이 더 고조될 것으로 보인다.

[표 2-2] 빅데이터를 이용한 경제 분석 관련 선행연구

| 연구자(연도)                | 연구내용   |
|------------------------|--|
| Choi and Varian (2012) | · 구글 검색통계를 이용한 모형을 바탕으로 미국 소매판매, 자동차 판매 등 예측                                   |
| Chamberlin (2010)      | · 구글 검색통계와 소매판매 데이터 간 관계를 바탕으로 구글 검색통계가 영국 소비 패턴과 밀접하게 움직이는 것을 밝힘              |
| Zimmermann (2009)      | · 구글 검색통계가 독일 실업 데이터와 밀접하게 움직이고 있음을 밝힌 후 구글 검색통계를 이용하여 독일 실업 데이터 예측            |
| D'Amuri (2009)         | · 구글 검색통계를 이용하여 이탈리아 실업률 예측  |
| Suhoy (2009)           | · 구글 검색통계를 이용하여 이스라엘의 2008년 경기하강 예측  |
| McLaren (2011)         | · 구글 검색통계를 이용하여 영국 노동 및 부동산 시장 분석, 구글 검색통계의 유용성 평가                             |
| 김지은 (2013)             | · 구글 검색통계와 우리나라 실물지표 간 대응성 분석과 예측력 분석을 통해 구글 검색통계의 유용성 점검                      |
| UN 글로벌 펄스              | · 트위터 게시글의 감성분석 결과와 실업률 변화 비교, 트위터 게시글을 분석하여 글로벌 경제 위기 이후 미국과 인도네시아 사람들의 행태 분석 |

자료: 선행연구를 바탕으로 연구자 재작성

## 다. 뉴스 빅데이터의 경제 분석 활용

경제관련 뉴스기사는 개인, 기업, 정부 등 경제주체의 심리 변화를 촉발시킬 수 있고 이에 따라 경제주체의 경제활동에 관한 의사결정을 변화게 하고, 이로 인하여 정책당국의 정책도 수정·보완을 하도록 하는 일련의 촉매제 역할을 한다.

금융위기, 외환위기 등과 같은 부정적 뉴스는 경제주체의 심리 악화, 경제 불확실성 증대, 환율 및 주가 등의 변동 폭 상승, 소비와 투자의 악화 등을 초래할 수 있으며 정책당국은 경제의 불확실성 완화, 경제심리 안정을 위한 정책을 수립하게 되므로 경제 뉴스기사는 경제활동을 파악하고 예측하는 중요한 자료가 될 수 있다.

빅데이터를 활용한 경제 분석의 한 분야로 뉴스 빅데이터를 이용한 경기종합지수, 소비자대지수, 소비자평가지수 등 다양한 지수개발을 통한 경제 분석을 볼 수 있다. 빅데

이터의 범위로는 특정 언론사 중심으로 한 분석과 다양한 언론사를 중심으로 한 분석으로 데이터 범위가 나누어질 수 있으며, 이완수, 박양수(2016)의 경우에는 주요 신문 및 방송사인 조선일보와 중앙일보 1면, KBS와 SBS 저녁종합뉴스를 분석하여 지수를 작성하였고, 이금희 외(2017)는 빅카인즈 뉴스데이터를 분석하여 다양한 신문과 방송사의 데이터를 활용하였다.

경제 분석 분야로는 경기종합지수, 소비자기대지수, 소비자평가지수 등을 개발하였으며 경기지수 관련 분석은 주로 경기실사지수(BSI) 중심의 지수 개발이 진행되었다. 뉴스의 특성 및 데이터 공표주기가 통계청에서 작성하는 지수에 비해 선행성을 가진다는 특성을 활용하여 경기선행지수를 작성하는 연구가 진행되었다(이완수, 노성중, 2008; 이완수, 박양수, 2016).

[표 2-3] 뉴스 빅데이터를 이용한 경기지수 분석 관련 선행연구

| 연구자(연도)  | 연구내용  |
|--|---|
| 이완수, 심재철, 박양수 (2007)   | · 조선일보와 동아일보의 1면, KBS와 SBS의 저녁종합뉴스, 경제뉴스를 활용하여 경기종합지수, 소비자기대지수, 소매판매 등을 분석    |
| 이완수, 노성중 (2008)  | · 조선일보와 동아일보의 1면, KBS와 SBS의 저녁종합뉴스, 경제뉴스를 활용하여 경기선행지수, 소비자기대지수, 소비자평가지수 등을 분석 |
| 이완수, 박양수 (2016)  | · 조선일보와 동아일보의 1면, KBS와 SBS의 저녁종합뉴스 전체를 활용하여 경기선행지수, 소비자기대지수, 소비자평가지수 등을 분석    |
| Tobback et. al. (2016)   | · 5개 Flemish 신문과 온라인 신문 자료를 활용하여 금융 경기통계 분석                                   |
| 이금희 (2017)   | · 한국언론진흥재단에서 관리하는 빅카인즈 뉴스 데이터를 활용하여 뉴스 경기지수 개발                                |
| Chetty et al. (2011),<br>Finkelstein et al. (2012),<br>Einav et al. (2013) | · 공공 데이터 및 거래 데이터를 이용하여 경제 분석   |

자료: 선행연구를 바탕으로 연구자 재작성

경제관련 뉴스 빅데이터 분석을 통한 지역 경기현황 판단은 지역통계 작성이 중앙단위에 비해 열악한 현실을 감안한다면 경기판단을 위한 가장 좋은 대안이 될 수 있다. 하지만 개별 연구자가 소셜네트워크 텍스트로부터 주요 키워드를 수집하기 위해서는 서버 구축, 수집 기술 확보 등의 많은 비용이 소요되기 때문에 일부 업체에서는 소셜네트워크 상의 여러 종류 텍스트를 수집하여 데이터베이스로 구축하고 유료로 주요 키워드 시계열을 제공하기도 한다.



---

## 2. 뉴스 빅데이터의 특성 선행연구 검토

---

### 가. 국내 연구

이완수 외 2인(2007)은 국내 언론의 경제뉴스 보도가 국민 경제 주체인 소비자의 경제심리, 소비행위 간의 인과관계를 분석한 연구이다. 객관적인 경제상황 변수로 경기선행지수와 경기동행지수를, 주관적인 소비자심리 변수로 소비자 기대지수를, 소비활동 수준 변수로 소비재판매액지수를, 그리고 경제뉴스 변수로 1998년 12월부터 2005년 12월까지 조선일보와 동아일보의 1면, KBS와 SBS의 저녁종합뉴스의 경제뉴스 논조 및 부정적 뉴스 건수 등의 데이터를 분석변수로 활용하여 벡터자기회귀 모형(VAR)과 다중회귀 모형 추정을 통해 분석하였다.

분석결과 경제지표가 보여주는 실제 경제상황과 이에 대한 소비자 기대심리는 서로 예측관계를 갖는 것으로 나타났다. 또한 언론의 경제뉴스는 실질적인 경제 상황보다도 더 즉각적이며 강력하게 소비자의 경제심리에 영향을 주었으며, 이러한 소비자 경제심리는 다시 그들의 소비행위에 거시적 지표로서 영향을 주는 것을 확인할 수 있었다. 이는 언론이 경제이슈를 어떤 속성이나 관점으로 다루냐에 따라 경제현실에 대한 소비자 인식이 달라질 수 있기 때문에 언론은 경제뉴스를 보도함에 있어 보다 객관적이며 정확한 보도를 해야 한다는 시사점을 보여준다.

경제뉴스 속성의제가 소비자 의제로, 그리고 이러한 의제가 경기 국면에 따라 서로 다르게 소비자에게 전이된다는 사실을 시계열적으로 확인했다는 점에서 의미를 갖는다. 하지만 인터넷과 휴대폰 등 폭넓은 분석매체가 있지만 분석범위를 신문 1면과 지상파 방송의 저녁종합뉴스로 제한한 점과 국가부도 사태였던 IMF와 노무현 정부 전체시기를 분석 대상으로 하지 못한 점이 한계점으로 존재한다.

이완수 외 1인(2008)은 언론의 경제보도와 현실 경제지표, 국민의 경제인식 간의 예측관계가 상이한 경제구조, 정부와 언론의 관계, 국민 여론의 특징을 갖는 두 정부 하에서 언제 발생하고, 소멸되는지를 실증적으로 분석한 연구이다. 경제보도 변수로 1998년 12월부터 2005년 12월까지 조선일보와 동아일보의 1면, KBS와 SBS의 저녁종합뉴스의 경제뉴스 논조 및 부정적 뉴스 건수를, 경제상황 변수로는 경기선행지수와 경기동행지수를, 국민의 경제인식 변수로는 소비자평가지수와 소비자기대지수를 활용하여 언론의 경제보

---

도와 경제상황, 경제인식 간 상호영향의 시간차를 벡터자기회귀 모형(VAR)을 사용하여 시계열 분석하였다.

분석결과 언론의 경제보도가 현실 경제지표에 미치기까지 2개월 정도의 시간이 걸리는 것으로 나타났고, 국민의 경제인식도 2개월의 시간이 소요되는 것으로 분석되었다. 그리고 김대중과 노무현 정부 시기 모두에서 현실 경제지표가 언론의 경제보도에 반영되기까지는 상대적으로 짧은 1개월의 시간이 소요되었으며, 현실 경제-경제보도-경제인식 간 상호영향의 시간차는 정부 간에는 뚜렷한 차이가 발견되지 않는 것으로 나타났다.

거시경제지표를 활용하여 현실 경제-경제보도-경제인식 삼각 축에서의 상호 예측관계에서 발견할 수 있는 시간차를 다면적으로 살펴본 점에서 의의가 있다. 하지만 물가상승률 등과 같은 미시경제지표를 간과했고, 국민의 경제인식이 개인의 차원이나 자신이 속한 사회적 집단에 대한 고려가 함께 이루어지지 못한 점과 부정적 경제보도의 논조 건수만 사용되어 분석했다는 점에서 한계가 있다. 향후에는 좀 더 다양한 지표들을 사용하여 '언제 효과가 나타나고, 사라지는가'에 대한 커뮤니케이션 효과 연구가 필요하다고 하였다.

이완수 외 1인(2016)은 경제 주체(개인, 기업)가 언론의 경제 뉴스를 접했을 때 향후 경기전망을 어떻게 다르게 인식하여 반응하는지 경제 정보의 비대칭적 효과 개념을 분석한 연구이다. 1998년 12월부터 2014년 12월까지 조선일보와 동아일보 1면, KBS와 SBS의 저녁 종합 뉴스시간에 보도된 경제 뉴스 데이터와 경기동행지수 순환변동치를 활용하여 소비자동향지수(CSI), 기업경기실사지수(BSI) 등을 실증분석을 하였다.

분석결과 개인은 기업에 비해 경제 정보에 더욱 민감하게 반응하였으며, 특히 개인의 경우 남성과 자영업자가 경제 뉴스에 더 민감하게 반응하는 것을 파악할 수 있었다. 기업의 경우 대기업과 내수기업, 제조업 분야에서 상대적으로 민감하게 반응하였다. 그리고 경제 주체는 경제 정보에 대해 경기 호황기보다는 경기 불황기에 더 민감하게 반응하였고, 기업은 다른 정부에 비해 외환위기와 같은 극단적 상황에서 구조조정을 겪은 김대중 정부 시절에 상대적으로 더 민감한 반응을 보였다.

이와 같은 결과로 국내 경제 뉴스는 경제 상황과 관계없이 개인, 기업의 규모 및 특성에 따라 일관되게 영향을 미치지만 정보 해석 능력 차이에 따라 다른 강도를 보이는 '정보 처리 역량의 비대칭성' 현상을 볼 수 있었다. 따라서 정부가 실제 정책을 수립하고

---

의사결정 과정에서 개별적 특성을 고려하여 적절한 사후통제가 필요하다는 점을 제안하지만, 경제 주체의 민감도 차이가 경제 정보 처리 능력 때문인지 서로 다른 경제 활동의 분야 및 범위의 특성 때문인지 알 수 없다는 한계점을 보였다.

이금희(2017)는 한국언론진흥재단의 분석시스템인 빅카인즈에 축적된 뉴스 기사 자료에서 호황과 불황 관련 키워드들의 월별 뉴스 기사수를 추출하여 BSI 방식으로 빅카인즈 뉴스 경기지수를 개발하였다. 빅카인즈 뉴스 경기지수는 먼저, 경제 관련 뉴스기사 키워드 중 경기와 관련성이 높은 키워드를 발굴하여 빅카인즈의 뉴스 기사 자료에서 월별 시계열로 추출한 후, 추출된 키워드별 시계열자료로부터 경기호황 및 경기불황과 밀접한 키워드를 찾아 공표된 경제지표와의 관련성을 바탕으로 분류한 후 합산하여 호황지표와 불황지표를 작성하였다. 그리고 작성된 호황지표와 불황지표의 계절변동을 제거하여 계절조정된 호황지표와 불황지표를 이용하여 BSI 방식의 빅카인즈 뉴스 경기지수를 작성하였다.

작성된 빅카인즈 뉴스 경기지수의 유용성 검토를 위해 빅카인즈 뉴스 경기지수의 순환변동치를 작성하고 기존 공표된 경기통계와의 교차상관분석, 전환점 분석을 하였고, 로짓모형을 통해 빅카인즈 뉴스 경기지수 순환변동치를 경제심리지수 순환변동치, 경기선행지수 순환변동치와 비교하여 경기 수축 국면 예측력 분석을 하였다. 분석결과 빅카인즈 뉴스 경기지수 순환변동치는 경기선행지수 순환변동치와 경제심리지수 순환변동치 등과 밀접하게 움직이는 것으로 나타났고, 경제심리지수 순환변동치 또는 경기선행지수 순환변동치를 포함한 로짓모형보다 빅카인즈 뉴스 경기지수 순환변동치를 포함한 로짓모형이 경기 수축국면을 잘 예측하는 것으로 분석되었다.

빅카인즈 뉴스 경기지수는 경기 판단 시 속보성과 빈도성 측면에서 유용하게 사용될 수 있을 것이다. 더 나아가 모든 언론사의 뉴스를 보강하고, 언론사별 과거 뉴스 기사가 충분히 보완되면 빅카인즈 뉴스 경기지수는 보다 높은 대표성이 있는 지수가 될 것이다. 그리고 고용, 물가 등 다른 경제 분야에서도 키워드 군을 충분히 확보·검토하여 이와 관련된 뉴스지수를 작성하고, 로짓모형 외에 예측모형을 보다 다양하게 작성하여 빅카인즈 뉴스 경기지수의 경기 예측력을 높일 필요가 있으며, 경제성장률과 민간소비 증감률 등 거시경제변수를 예측하는 모형에 빅카인즈 뉴스 경기지수를 포함하여 예측력을 통해 그 유용성을 점검할 필요가 있다고 제시하였다.

---

## 나. 국외 연구

뉴스 빅데이터를 이용한 경제분석의 분야 중 하나는 텍스트 마이닝을 포함한 뉴스 빅데이터를 활용한 경제지수의 개발 및 개선 분야이다. Tobby et. al. (2016)은 경제정책의 불확실성을 측정하고자 하는 노력으로 자주 등장하는 EPUI(Economic Policy Uncertainty Index)를 개선하기 위해 텍스트 마이닝 기술을 활용하여 양식주식, 서포트 벡터 머신(SVM) 분류 방식을 원래 방식과 비교하여 벨기에의 EPUI를 만들었으며 이를 소비자 및 생산자 불확실성지수, 10년 OLO-Bund 스프레드와 비교함으로써 SVM 방식의 활용도가 높음을 증명하였으며 후속연구인 Tobby et. al. (2017)에서 텍스트 마이닝을 활용하여 벨기에의 경제정책불확실성지수(EPUI)를 개선할 수 있음을 보였다. 또한, Baker et. al. (2016)은 신문기사를 기반으로 새로운 경제정책불확실성 지수를 개발했다. 지수가 정책 관련 경제적 불확실성의 움직임을 대신한다는 것을 주장하고 그 증거로 박빙의 대통령선거, 1차 2차 걸프전, 9.11 공격, 리먼 브라더스 사태, 2011년 부채한도 분쟁과 재정 관련한 주요논쟁이 격화되는 시점에 지수가 민감하게 반응한다고 설명했다.

뉴스를 활용한 예측모형에서 자주 등장하는 부분이 뉴스검색어를 활용한 시장예측이다. Li et. al. (2016)은 온라인 뉴스 기사의 논조를 이용하여 원유가격을 예측했다. 또한 검색엔진인 구글 인덱스를 활용하여 원유가격을 예측하고(Yao & Zhang, 2017), 질병을 추적하고(Pelat et. al., 2009), 단기 경제지수를 예측(Choi & Varian, 2011) 하는 등 구글 검색엔진이 단기 지수예측에 많이 활용되고 있다.

한편, Gentzkow and Shapiro (2010)는 미디어 편향지수를 개발하여 독자들이 경제적으로 같은 생각을 가진 뉴스에 대해 상당한 선호도를 보이고 기업은 소비자의 선호에 20%를 설명할 정도로 강하게 반응하는 것을 발견했으나 언론사주의 정체성은 선호도에 영향을 덜 미치는 것으로 나타났다고 주장하는 등 빅데이터를 활용한 비정형데이터의 정형화로 다양한 분석이 시도되고 있다.



## 제 3장

### 뉴스 빅데이터 지역경기지수 작성

1. 경기관련 뉴스 키워드 데이터 수집
2. 뉴스 데이터 수집 및 통계 분석 결과
3. 뉴스 빅데이터 지역경기지수 작성





## 제 3 장 뉴스 빅데이터 지역경기지수 작성

### 1. 경기관련 뉴스 키워드 데이터 수집

#### 가. 뉴스 데이터 수집 대상 신문사

전북을 기반으로 하는 14개 일간지와 3개 방송사에서는 매일 엄청난 양의 뉴스를 생산하고 있다. 최근에 들어와서 전북에서도 많은 신문사들이 온라인 서비스를 개시하여 일반인이 관심 뉴스를 검색할 수 있도록 하고 있어 뉴스의 활용도가 높아지고 있다.

뉴스 제공 서비스 채널도 다양해져 개별 언론사를 통한 검색뿐만 아니라 다음과 네이버 포털을 통한 검색도 가능하게 서비스가 제공된다. 포털 서비스가 가지는 장점을 최대한 활용한다면 다양한 뉴스 서비스를 하나의 창구를 통해 취득할 수 있게 된다.

다양한 뉴스 일간지 중에서 본 연구의 뉴스기사 수집 대상 언론사는 네이버 뉴스에서 제공하는 ‘새전북신문’, ‘전북도민일보’, ‘전북일보’, ‘전북중앙신문’, ‘전주MBC’, ‘JTV전주방송’을 대상으로 하였다.

열독률이 가장 높은 ‘조선일보’, ‘중앙일보’, ‘동아일보’, ‘매일경제’, 그리고 ‘한겨레’와 같은 중앙지도 포함할 수 있겠으나<sup>2)</sup>, 본 연구에서는 중앙지를 데이터 수집 대상에 포함시키지 않았다. 본 연구가 초기연구이고 중앙의 경제상황이 지역의 특수성을 왜곡할 가능성과 중앙지보다는 지역지가 더 빠르고 풍부하게 경기 현황 정보를 제공할 것으로 판단하였기 때문이다.

#### 나. 경기 관련 검색어의 선정

본 연구에서 수집할 경제용어는 이공희(2017) 연구에서 사용된 바와 같이 불황 관련 용어 21개, 호황 관련 용어 12개, 일반 경제용어 14개의 총 47개 용어와 견조, 수출 증

2) 한국언론진흥재단의 ‘2021 신문잡지 이용조사’ 결과에서 열독률 상위 6개 신문사는 “조선일보”, “중앙일보”, “동아일보”, “매일경제”, “농민신문”, “한겨레” 순으로 나타났다.

가, 고용률 상승 등 6개의 단어를 추가하여 총 53개의 키워드로 데이터를 수집하려 한다. 구체적인 관련 경제용어는 아래와 같다.

[표 3-1] 경기 관련 수집대상 경제용어

| 불황 관련 용어   | 호황 관련 용어  | 경제 관련 용어   |
|--|---|--|
| 불경기, 불황, 경기침체, 경기불황, 경제불황, 위기, 금융위기, 경제위기, 외환위기, 적자, 부도, 파산, 폭락, 하한가, 신용불량, 정리하고, 해고, 부채, 금리인하, 디플레이션, IMF, 재고증가 | 호경기, 호황, 경기회복, 경기호황, 경제호황, 흑자, 채용, 폭등, 상한가, 거품, 금리인상, 인플레이션, 재고감소, 활황, 수출증가, 고용률 상승, 건조 | 환율, 주가, GDP, 실업률, 물가, 경제성장, 통화정책, 재정정책, 한국은행, 외환보유고, 경상수지, 신용등급, 유가, FRB |

자료: 이금희(2017). "뉴스 빅데이터를 이용한 경기판단: 빅인즈 뉴스 경기지수의 개발", 통계연구 22(2)

## 다. 뉴스 데이터 수집 방법

뉴스기사는 경제용어(불황, 호황, 경제관련)를 검색사이트에서 검색한 후, 검색된 뉴스 기사를 연도별 월별 일별로 스크랩하여 구축한다. 그러므로 가장 먼저 경제용어를 어디서 검색할 것인지 결정하고, 그 다음으로 검색된 결과를 어떠한 방식으로 수집할 것인지에 대한 결정이 필요하다.

뉴스기사는 네이버, 다음, 한국언론진흥재단과 같은 포털사이트 혹은 전북일보, 전북도민일보와 같은 개별 신문사 인터넷 사이트에서 검색할 수 있다. 모든 인터넷 사이트에서는 검색할 경제용어와 관련해서 검색 기간, 기사 유형과 같은 옵션을 제공하고 있으므로 특정 기간의 뉴스를 찾기에 적합하다. 하지만 개별 신문사 인터넷 사이트는 해당 신문사 기사만을 제공하고 있어 포털사이트에서 검색하는 것보다 시간과 비용이 많이 소요될 수 있다는 단점이 있다.

검색결과를 수집(scrap)하기 위해서는 별도의 프로그래밍을 작성해야 한다. 왜냐하면 본 연구에서 수집하려는 기사는 2007년 1월 1일부터 2021년 12월 31일까지 6개 지역 신문사와 53개 경제용어의 일별 자료를 검색하여 수집하는 것이므로 백만 번 이상의 수집 활동을 해야 하기 때문이다<sup>3)</sup>.

본 연구에서는 경제용어 검색과 검색결과 기사를 수집하기 위하여 웹 스크래핑 프로그



램을 작성하여 활용한다. 구체적으로 파이썬(python) 언어를 이용하여 네이버 뉴스기사를 스크래핑하고, 스크래핑 된 결과를 저장하도록 프로그램을 작성하는 것이다<sup>4)</sup>. 개략적인 프로그램의 구성은 아래와 같다.

```
# 날짜 지정 *****
date = str(datetime.now())
date = date[:date.rfind(':')].replace(' ', '_')
date = date.replace(':', '시') + '시'

sleep_sec = 0.5

##### 언론사별 본문 위치 링크 표시 함수 #####
print('본문 크롤링에 필요한 함수를 로딩하고 있습니다...\n' + '-' * 100)
def crawling_main_text(url):
    try:
        req = requests.get(url)
        req.encoding = None
        soup = BeautifulSoup(req.text, 'html.parser')

        # 5. JTV전주방송(호성)
        if 'jtv' in url:
            text = soup.find('div', {'id' : 'vContent'}).text

        # 6. 새경북신문(호성)
        if 'sjbnews' in url:
            text = soup.find('span', {'class' : 'news_text cl6 p-b-25'}).text

        # 7. 전북도민일보(호성)
        if 'domin' in url:
            text = soup.find('div', {'id' : 'article-view-content-div'}).text

        # 8. 전북일보(호성)
        if 'jjan' in url:
            text = soup.find('div', {'class' : 'article_txt_container'}).text

        # 9. 전북중앙(호성)
        if 'jcn' in url:
            text = soup.find('div', {'id' : 'article-view-content-div'}).text

        # 10. 전주MBC(호성)
        if 'jmbc' in url:
            text = soup.find('div', {'class' : 'contents_wrap'}).text

        #elif 'URL코드' in url:
        #text = soup.find('div', {'class' : '본문 class'}).text
        #elif 'URL코드' in url:
        #text = soup.find('div', {'class' : '본문 class'}).text

    # 그 외
    else:
        text == None
    except:
        pass
    else:
        return text.replace('\n', '').replace('\r', '').replace('<br>', '').replace('\t', '')

press_nm = '전북도민일보'
print('경성할 언론사 : {}'.format(press_nm))
```

[그림 3-1] Jupyter notebook을 이용한 웹 스크래핑 프로그램

- 3) 경제용어별 신문기사가 1일에 1개씩 생산되었다고 가정하면, 본 연구에서 스크랩을 위한 활동은 2007년 1월 1일부터 2021년 12월 31일까지 총 5,479일×6개 지역신문사×53개 경제용어이므로 최대 총 1,742,322회가 된다.
- 4) 프로그램은 파이썬 3.9 버전과 주피터 노트북(jupyter notebook)을 활용하였다.

## 2. 뉴스 데이터 수집 및 통계 분석 결과

### 가. 뉴스 데이터 수집 결과

본 연구에서는 앞에서 기술한 바와 같이 불황 관련 용어 22개, 호황 관련 용어 17개, 일반 경제용어 14개로 총 53개 용어에 대하여 수집하였다. 6개 지역신문사의 경제용어 스크래핑 결과를 종합하면 아래와 같다. 불황 관련 용어는 70,963개, 호황 관련 용어는 17,859개, 경제 관련 용어는 11,395개로 총 100,217개의 뉴스기사가 검색되었다. 이러한 결과를 신문사별 제시하면 다음과 같다.

[표 3-2] 6개 지역신문의 경기변동 관련 경제용어 웹 스크래핑 결과

| 구분        | 불황 관련 용어      | 호황 관련 용어      | 경제 관련 용어      |
|-----------|---------------|---------------|---------------|
| 새전북신문     | 19,865        | 4,847         | 4,670         |
| 전북도민일보    | 24,281        | 5,876         | 70            |
| JTV전주방송   | 229           | 85            | 77            |
| 전북일보      | 18,790        | 5,134         | 4,832         |
| 전북중앙      | 7,133         | 1,855         | 1,648         |
| 전주MBC     | 665           | 62            | 98            |
| <b>합계</b> | <b>70,963</b> | <b>17,859</b> | <b>11,395</b> |

주1: 저자 직접 작성

주2: 스크래핑은 총 2회를 실시하였고, 첫 번째는 2022년 3월 12일 ~ 4월 5일까지 총 25일 동안 수행하였으며, 두 번째는 2022년 11월 1일 ~ 4일까지 총 4일 추가 수행

#### 1) 새전북신문 경제 관련 용어 스크래핑 결과

새전북신문의 불황 관련 경제용어를 스크래핑한 결과 2011년 12월부터 뉴스기사가 검색되었으며, 2021년 12월 31일까지 총 19,865개의 뉴스기사가 검색된 것으로 나타났다.

[표 3-3] 새전북신문 불황 관련 용어 스크래핑 결과

| 구분        | 1            | 2            | 3            | 4            | 5            | 6            | 7            | 8            | 9            | 10           | 11           | 12           | 합계            |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|
| 2011      |              |              |              |              |              |              |              |              |              |              |              | 139          | 139           |
| 2012      | 223          | 150          | 130          | 119          | 172          | 147          | 202          | 110          | 226          | 170          | 202          | 202          | 2,053         |
| 2013      | 205          | 147          | 150          | 121          | 114          | 167          | 132          | 93           | 96           | 151          | 132          | 132          | 1,640         |
| 2014      | 156          | 142          | 143          | 107          | 117          | 103          | 149          | 119          | 130          | 132          | 120          | 160          | 1,578         |
| 2015      | 139          | 100          | 128          | 83           | 148          | 142          | 112          | 101          | 120          | 107          | 117          | 123          | 1,420         |
| 2016      | 129          | 107          | 117          | 101          | 134          | 99           | 98           | 109          | 97           | 117          | 121          | 165          | 1,394         |
| 2017      | 199          | 113          | 105          | 127          | 137          | 115          | 125          | 124          | 93           | 91           | 129          | 114          | 1,472         |
| 2018      | 112          | 133          | 198          | 184          | 174          | 119          | 164          | 177          | 157          | 148          | 160          | 171          | 1,897         |
| 2019      | 207          | 125          | 129          | 196          | 176          | 161          | 205          | 171          | 186          | 161          | 133          | 187          | 2,037         |
| 2020      | 155          | 268          | 509          | 410          | 321          | 321          | 302          | 257          | 228          | 218          | 228          | 284          | 3,501         |
| 2021      | 277          | 210          | 247          | 228          | 213          | 245          | 213          | 193          | 210          | 192          | 238          | 268          | 2,734         |
| <b>합계</b> | <b>1,802</b> | <b>1,495</b> | <b>1,856</b> | <b>1,676</b> | <b>1,706</b> | <b>1,619</b> | <b>1,702</b> | <b>1,454</b> | <b>1,543</b> | <b>1,487</b> | <b>1,580</b> | <b>1,945</b> | <b>19,865</b> |

호황 관련 경제용어는 2011년 12월부터 뉴스기사가 검색되었으며, 2021년 12월 31일까지 총 4,847개의 뉴스기사가 검색되었다.

[표 3-4] 새전북신문 호황 관련 용어 스크래핑 결과

| 구분        | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         | 합계           |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|
| 2011      |            |            |            |            |            |            |            |            |            |            |            | 46         | 46           |
| 2012      | 52         | 54         | 44         | 46         | 64         | 38         | 40         | 32         | 56         | 44         | 41         | 42         | 553          |
| 2013      | 41         | 31         | 33         | 35         | 32         | 37         | 25         | 21         | 20         | 40         | 39         | 24         | 378          |
| 2014      | 48         | 39         | 33         | 25         | 31         | 16         | 35         | 20         | 23         | 34         | 38         | 50         | 392          |
| 2015      | 48         | 29         | 34         | 22         | 26         | 33         | 35         | 27         | 42         | 31         | 43         | 33         | 403          |
| 2016      | 41         | 30         | 33         | 28         | 30         | 45         | 42         | 38         | 27         | 34         | 24         | 27         | 399          |
| 2017      | 48         | 46         | 29         | 36         | 47         | 44         | 42         | 28         | 40         | 37         | 51         | 37         | 485          |
| 2018      | 44         | 40         | 50         | 45         | 37         | 35         | 30         | 52         | 62         | 65         | 62         | 55         | 577          |
| 2019      | 34         | 39         | 43         | 44         | 34         | 53         | 33         | 26         | 37         | 53         | 45         | 36         | 477          |
| 2020      | 51         | 46         | 33         | 25         | 28         | 64         | 47         | 47         | 42         | 35         | 49         | 49         | 516          |
| 2021      | 42         | 43         | 54         | 56         | 55         | 53         | 52         | 38         | 47         | 63         | 64         | 54         | 621          |
| <b>합계</b> | <b>449</b> | <b>397</b> | <b>386</b> | <b>362</b> | <b>384</b> | <b>418</b> | <b>381</b> | <b>329</b> | <b>396</b> | <b>436</b> | <b>456</b> | <b>453</b> | <b>4,847</b> |

마지막으로 경제 관련 용어는 2021년 12월 31일까지 총 4,670개의 뉴스기사가 검색된 것으로 나타났다.

[표 3-5] 새전북신문 경제 관련 용어 스크래핑 결과

| 구분        | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         | 합계           |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|
| 2011      |            |            |            |            |            |            |            |            |            |            |            | 64         | 64           |
| 2012      | 75         | 68         | 56         | 56         | 46         | 37         | 84         | 34         | 78         | 60         | 50         | 55         | 699          |
| 2013      | 73         | 73         | 42         | 37         | 35         | 40         | 32         | 25         | 33         | 40         | 36         | 34         | 500          |
| 2014      | 67         | 42         | 28         | 24         | 21         | 28         | 32         | 21         | 16         | 32         | 43         | 59         | 413          |
| 2015      | 39         | 23         | 34         | 35         | 31         | 29         | 27         | 39         | 32         | 25         | 33         | 37         | 384          |
| 2016      | 35         | 26         | 41         | 27         | 31         | 30         | 35         | 48         | 28         | 29         | 28         | 33         | 391          |
| 2017      | 48         | 33         | 29         | 29         | 22         | 32         | 46         | 27         | 35         | 21         | 38         | 27         | 387          |
| 2018      | 32         | 25         | 47         | 26         | 36         | 33         | 53         | 55         | 39         | 48         | 49         | 38         | 481          |
| 2019      | 56         | 31         | 34         | 35         | 51         | 28         | 37         | 25         | 48         | 42         | 25         | 44         | 456          |
| 2020      | 44         | 35         | 50         | 31         | 25         | 41         | 42         | 43         | 30         | 25         | 34         | 31         | 431          |
| 2021      | 31         | 51         | 31         | 33         | 26         | 31         | 48         | 33         | 60         | 50         | 34         | 36         | 464          |
| <b>합계</b> | <b>500</b> | <b>407</b> | <b>392</b> | <b>333</b> | <b>324</b> | <b>329</b> | <b>436</b> | <b>350</b> | <b>399</b> | <b>372</b> | <b>370</b> | <b>458</b> | <b>4,670</b> |

## 2) 전북도민일보 경제 관련 용어 스크래핑 결과

전북도민일보의 불황 관련 경제용어를 스크래핑한 결과 2013년 10월부터 뉴스기사가 검색되었으며, 2021년 12월 31일까지 총 24,281개의 뉴스기사가 검색된 것으로 나타났다. 이 결과는 새전북신문의 불황 관련 뉴스기사 검색결과보다 많다.

[표 3-6] 전북도민일보 불황 관련 용어 스크래핑 결과

| 구분        | 1            | 2            | 3            | 4            | 5            | 6            | 7            | 8            | 9            | 10           | 11           | 12           | 합계            |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|
| 2013      |              |              |              |              |              |              |              |              |              | 88           | 163          | 236          | 487           |
| 2014      | 140          | 205          | 233          | 199          | 153          | 168          | 218          | 116          | 175          | 198          | 184          | 220          | 2,209         |
| 2015      | 201          | 177          | 261          | 180          | 159          | 83           | 211          | 222          | 175          | 194          | 186          | 189          | 2,238         |
| 2016      | 261          | 191          | 221          | 175          | 153          | 105          | 228          | 203          | 127          | 213          | 256          | 275          | 2,408         |
| 2017      | 306          | 235          | 248          | 241          | 239          | 208          | 202          | 255          | 148          | 144          | 210          | 178          | 2,614         |
| 2018      | 175          | 272          | 298          | 256          | 125          | 203          | 102          | 246          | 216          | 177          | 189          | 219          | 2,478         |
| 2019      | 274          | 239          | 248          | 274          | 250          | 291          | 320          | 282          | 312          | 288          | 228          | 259          | 3,265         |
| 2020      | 256          | 497          | 756          | 212          | 528          | 473          | 454          | 333          | 382          | 345          | 371          | 402          | 5,009         |
| 2021      | 399          | 287          | 309          | 324          | 249          | 279          | 269          | 288          | 267          | 255          | 309          | 338          | 3,573         |
| <b>합계</b> | <b>2,012</b> | <b>2,103</b> | <b>2,574</b> | <b>1,861</b> | <b>1,856</b> | <b>1,810</b> | <b>2,004</b> | <b>1,945</b> | <b>1,802</b> | <b>1,902</b> | <b>2,096</b> | <b>2,316</b> | <b>24,281</b> |

호황 관련 경제용어는 2013년 10월부터 검색되었으며, 2021년 12월 31일까지 총 5,876개의 뉴스기사가 검색되었다.

[표 3-7] 전북도민일보 호황 관련 용어 스크래핑 결과

| 구분        | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         | 합계           |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|
| 2013      |            |            |            |            |            |            |            |            |            | 23         | 18         | 72         | 113          |
| 2014      | 63         | 62         | 11         | 68         | 53         | 43         | 38         | 53         | 56         | 60         | 59         | 59         | 625          |
| 2015      | 68         | 68         | 93         | 67         | 52         | 53         | 49         | 63         | 54         | 17         | 54         | 14         | 652          |
| 2016      | 17         | 67         | 51         | 49         | 45         | 63         | 58         | 47         | 49         | 56         | 53         | 49         | 604          |
| 2017      | 51         | 70         | 72         | 58         | 54         | 72         | 75         | 55         | 64         | 56         | 61         | 54         | 742          |
| 2018      | 85         | 63         | 78         | 78         | 62         | 49         | 53         | 56         | 65         | 81         | 84         | 57         | 811          |
| 2019      | 75         | 55         | 80         | 60         | 59         | 67         | 43         | 48         | 55         | 71         | 67         | 56         | 736          |
| 2020      | 56         | 56         | 63         | 51         | 67         | 101        | 61         | 56         | 68         | 66         | 87         | 57         | 789          |
| 2021      | 71         | 67         | 61         | 75         | 73         | 86         | 57         | 54         | 58         | 71         | 77         | 54         | 804          |
| <b>합계</b> | <b>486</b> | <b>508</b> | <b>509</b> | <b>506</b> | <b>465</b> | <b>534</b> | <b>434</b> | <b>432</b> | <b>469</b> | <b>501</b> | <b>560</b> | <b>472</b> | <b>5,876</b> |

경제 관련 용어도 2013년 10월부터 검색되었으며, 2021년 12월 31일까지 총 70개의 뉴스기사가 검색되었다.

[표 3-8] 전북도민일보 경제 관련 용어 스크래핑 결과

| 구분        | 1        | 2        | 3         | 4        | 5        | 6        | 7        | 8        | 9        | 10       | 11       | 12       | 합계        |
|-----------|----------|----------|-----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| 2013      |          |          |           |          |          |          |          |          |          |          | 4        | 1        | 5         |
| 2014      | 3        | 5        | 1         | 2        |          | 2        | 1        | 1        | 3        |          |          |          | 18        |
| 2015      | 2        | 1        |           | 1        |          |          | 2        | 1        |          | 1        |          |          | 8         |
| 2016      | 2        |          |           | 1        |          | 1        |          | 1        |          |          |          | 1        | 6         |
| 2017      | 1        |          |           | 1        |          | 1        |          |          |          |          |          | 2        | 5         |
| 2018      | 1        |          | 2         |          |          |          | 1        | 1        |          |          |          |          | 5         |
| 2019      |          |          | 5         | 1        | 2        |          |          | 2        | 1        |          |          |          | 11        |
| 2020      |          |          | 2         |          |          |          |          | 1        |          | 1        | 1        |          | 5         |
| 2021      |          | 1        | 1         |          | 1        |          | 3        |          |          |          | 1        |          | 7         |
| <b>합계</b> | <b>9</b> | <b>7</b> | <b>11</b> | <b>6</b> | <b>3</b> | <b>4</b> | <b>7</b> | <b>7</b> | <b>4</b> | <b>2</b> | <b>6</b> | <b>4</b> | <b>70</b> |

### 3) JTV전주방송 경제 관련 용어 스크래핑 결과

JTV전주방송의 불황 관련 용어를 스크래핑한 결과 2021년 3월부터 검색되어 그 이전 기사는 찾을 수 없는 것으로 나타났다. 2021년 3월부터 2021년 12월까지의 10개월 간 불황 관련 경제용어 검색결과는 총 229개이다.

[표 3-9] JTV전주방송 불황 관련 용어 스크래핑 결과

| 구분   | 1 | 2 | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 합계  |
|------|---|---|----|----|----|----|----|----|----|----|----|----|-----|
| 2021 |   |   | 16 | 17 | 31 | 28 | 22 | 22 | 27 | 23 | 31 | 12 | 229 |
| 합계   |   |   | 16 | 17 | 31 | 28 | 22 | 22 | 27 | 23 | 31 | 12 | 229 |

호황 관련 용어도 불황 관련 용어와 마찬가지로 검색결과가 2021년 3월부터 검색되었으며 10개월 간 호황 관련 용어는 총 85개이다.

[표 3-10] JTV전주방송 호황 관련 용어 스크래핑 결과

| 구분   | 1 | 2 | 3 | 4 | 5  | 6  | 7 | 8 | 9 | 10 | 11 | 12 | 합계 |
|------|---|---|---|---|----|----|---|---|---|----|----|----|----|
| 2021 |   |   | 6 | 9 | 13 | 14 | - | 9 | 9 | 11 | 2  | 12 | 85 |
| 합계   |   |   | 6 | 9 | 13 | 14 | - | 9 | 9 | 11 | 2  | 12 | 85 |

경제 관련 용어도 위의 두 경제 관련 용어와 마찬가지로 2021년 3월부터 검색되었으며 경제 관련 용어 검색결과는 총 77개이다.

[표 3-11] JTV전주방송 경제 관련 용어 스크래핑 결과

| 구분   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 합계 |
|------|---|---|---|---|---|---|---|---|---|----|----|----|----|
| 2021 |   |   | 1 | 3 | 8 | 6 | 4 | 9 | 9 | 11 | 12 | 14 | 77 |
| 합계   |   |   | 1 | 3 | 8 | 6 | 4 | 9 | 9 | 11 | 12 | 14 | 77 |

#### 4) 전북일보 경제 관련 용어 스크래핑 결과

전북일보의 불황 관련 용어를 스크래핑한 결과 2014년 3월부터 뉴스기사가 검색되었으며, 2021년 12월 31일까지의 불황 관련 경제용어 검색결과는 총 18,790개이다.

[표 3-12] 전북일보 불황 관련 용어 스크래핑 결과

| 구분   | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    | 합계     |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| 2014 |       |       | 57    | 168   | 140   | 145   | 185   | 140   | 155   | 156   | 191   | 218   | 1,555  |
| 2015 | 191   | 105   | 223   | 135   | 148   | 166   | 164   | 139   | 150   | 129   | 176   | 162   | 1,888  |
| 2016 | 223   | 168   | 155   | 130   | 147   | 149   | 170   | 156   | 136   | 127   | 158   | 158   | 1,877  |
| 2017 | 214   | 181   | 147   | 150   | 164   | 185   | 150   | 173   | 137   | 118   | 110   | 136   | 1,865  |
| 2018 | 125   | 197   | 241   | 196   | 221   | 149   | 152   | 199   | 175   | 149   | 208   | 240   | 2,252  |
| 2019 | 322   | 198   | 250   | 291   | 275   | 278   | 350   | 322   | 314   | 304   | 256   | 249   | 3,409  |
| 2020 | 134   | 278   | 438   | 316   | 294   | 265   | 317   | 218   | 281   | 204   | 229   | 318   | 3,292  |
| 2021 | 291   | 229   | 197   | 223   | 189   | 203   | 194   | 223   | 211   | 199   | 223   | 270   | 2,652  |
| 합계   | 1,500 | 1,356 | 1,708 | 1,609 | 1,578 | 1,540 | 1,682 | 1,570 | 1,559 | 1,386 | 1,551 | 1,751 | 18,790 |

호황 관련 용어도 2014년 3월부터 검색되었으며, 2021년 12월 31일까지 총 5,134개의 뉴스기사가 검색되었다.

[표 3-13] 전북일보 호황 관련 용어 스크래핑 결과

| 구분        | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         | 합계           |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|
| 2014      |            |            | 23         | 58         | 54         | 35         | 47         | 41         | 55         | 53         | 52         | 64         | 482          |
| 2015      | 57         | 50         | 97         | 52         | 51         | 36         | 52         | 50         | 40         | 47         | 54         | 59         | 645          |
| 2016      | 52         | 57         | 53         | 51         | 62         | 53         | 64         | 61         | 47         | 49         | 39         | 46         | 634          |
| 2017      | 61         | 89         | 58         | 55         | 57         | 65         | 77         | 50         | 44         | 29         | 45         | 43         | 673          |
| 2018      | 60         | 47         | 52         | 50         | 47         | 33         | 51         | 48         | 47         | 59         | 71         | 70         | 635          |
| 2019      | 66         | 48         | 73         | 74         | 59         | 70         | 61         | 47         | 64         | 64         | 65         | 73         | 764          |
| 2020      | 39         | 48         | 37         | 46         | 41         | 65         | 42         | 47         | 57         | 42         | 53         | 58         | 575          |
| 2021      | 66         | 53         | 54         | 70         | 61         | 69         | 46         | 49         | 48         | 69         | 72         | 69         | 726          |
| <b>합계</b> | <b>401</b> | <b>392</b> | <b>447</b> | <b>456</b> | <b>432</b> | <b>426</b> | <b>440</b> | <b>393</b> | <b>402</b> | <b>412</b> | <b>451</b> | <b>482</b> | <b>5,134</b> |

경제 관련 용어 또한 2014년 3월부터 검색되었으며, 2021년 12월 31일까지 총 4,832개의 뉴스기사가 검색되었다.

[표 3-14] 전북일보 경제 관련 용어 스크래핑 결과

| 구분        | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         | 합계           |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|
| 2014      |            |            | 21         | 52         | 46         | 50         | 41         | 60         | 59         | 43         | 60         | 53         | 485          |
| 2015      | 73         | 63         | 58         | 51         | 51         | 46         | 51         | 58         | 41         | 51         | 54         | 52         | 649          |
| 2016      | 94         | 52         | 58         | 50         | 63         | 54         | 51         | 55         | 41         | 45         | 54         | 58         | 675          |
| 2017      | 79         | 69         | 48         | 35         | 44         | 55         | 44         | 38         | 44         | 42         | 35         | 44         | 577          |
| 2018      | 54         | 45         | 31         | 37         | 44         | 37         | 55         | 65         | 56         | 42         | 56         | 74         | 596          |
| 2019      | 88         | 46         | 51         | 74         | 72         | 79         | 56         | 92         | 63         | 62         | 42         | 45         | 770          |
| 2020      | 39         | 42         | 54         | 48         | 47         | 34         | 44         | 42         | 38         | 32         | 48         | 46         | 514          |
| 2021      | 48         | 43         | 36         | 35         | 33         | 45         | 57         | 55         | 62         | 58         | 51         | 43         | 566          |
| <b>합계</b> | <b>475</b> | <b>360</b> | <b>357</b> | <b>382</b> | <b>400</b> | <b>400</b> | <b>399</b> | <b>465</b> | <b>404</b> | <b>375</b> | <b>400</b> | <b>415</b> | <b>4,832</b> |

## 5) 전북중앙신문 경제 관련 용어 스크래핑 결과

전북중앙신문의 불황 관련 용어를 스크래핑한 결과 2018년 4월부터 뉴스기사가 검색되었으며, 2021년 12월 31일까지 총 7,133개의 뉴스기사가 검색된 것으로 나타났다.

[표 3-15] 전북중앙신문 불황 관련 용어 스크래핑 결과

| 구분        | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         | 합계           |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|
| 2018      |            |            |            | 12         | 118        | 119        | 180        | 162        | 175        | 168        | 149        | 142        | 1,225        |
| 2019      | 130        | 58         | 45         | 77         | 40         | 58         | 86         | 46         | 71         | 87         | 63         | 150        | 911          |
| 2020      | 132        | 258        | 418        | 348        | 311        | 268        | 256        | 167        | 222        | 167        | 147        | 222        | 2,916        |
| 2021      | 274        | 212        | 179        | 161        | 143        | 148        | 176        | 156        | 166        | 149        | 149        | 168        | 2,081        |
| <b>합계</b> | <b>536</b> | <b>528</b> | <b>642</b> | <b>598</b> | <b>612</b> | <b>593</b> | <b>698</b> | <b>531</b> | <b>634</b> | <b>571</b> | <b>508</b> | <b>682</b> | <b>7,133</b> |

호황 관련 용어도 2018년 4월부터 검색되었으며, 2021년 12월 31일까지 총 1,855개의 뉴스기사가 검색되었다.

[표 3-16] 전북중앙신문 호황 관련 용어 스크래핑 결과

| 구분        | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         | 합계           |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|
| 2018      |            |            |            | 6          | 36         | 44         | 43         | 44         | 60         | 68         | 70         | 51         | 422          |
| 2019      | 59         | 40         | 34         | 32         | 42         | 39         | 32         | 21         | 24         | 48         | 42         | 54         | 467          |
| 2020      | 37         | 37         | 28         | 29         | 35         | 44         | 39         | 28         | 41         | 37         | 34         | 43         | 432          |
| 2021      | 45         | 40         | 44         | 52         | 54         | 51         | 41         | 39         | 40         | 41         | 45         | 42         | 534          |
| <b>합계</b> | <b>141</b> | <b>117</b> | <b>106</b> | <b>119</b> | <b>167</b> | <b>178</b> | <b>155</b> | <b>132</b> | <b>165</b> | <b>194</b> | <b>191</b> | <b>190</b> | <b>1,855</b> |

경제 관련 용어도 2018년 4월부터 2021년 12월 31일까지 총 1,648개의 뉴스기사가 검색되었다.

[표 3-17] 전북중앙신문 경제 관련 용어 스크래핑 결과

| 구분        | 1          | 2          | 3          | 4         | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         | 합계           |
|-----------|------------|------------|------------|-----------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|
| 2018      |            |            |            | 3         | 47         | 35         | 56         | 48         | 45         | 49         | 54         | 49         | 386          |
| 2019      | 57         | 28         | 25         | 30        | 42         | 31         | 44         | 32         | 24         | 33         | 41         | 43         | 430          |
| 2020      | 45         | 40         | 44         | 27        | 43         | 36         | 33         | 29         | 34         | 29         | 31         | 32         | 423          |
| 2021      | 33         | 38         | 34         | 38        | 25         | 28         | 29         | 40         | 54         | 41         | 26         | 23         | 409          |
| <b>합계</b> | <b>135</b> | <b>106</b> | <b>103</b> | <b>98</b> | <b>157</b> | <b>130</b> | <b>162</b> | <b>149</b> | <b>157</b> | <b>152</b> | <b>152</b> | <b>147</b> | <b>1,648</b> |



## 6) 전주MBC 경제 관련 용어 스크래핑 결과

전주MBC의 불황 관련 용어를 스크래핑한 결과 2019년 10월부터 검색되었으며, 2021년 12월 31일까지의 불황 관련 경제용어 검색결과는 총 665개이다.

[표 3-18] 전주MBC 불황 관련 용어 스크래핑 결과

| 구분        | 1         | 2         | 3         | 4         | 5         | 6         | 7         | 8         | 9         | 10        | 11        | 12        | 합계         |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| 2019      |           |           |           |           |           |           |           |           |           | 2         | 12        | 26        | 40         |
| 2020      | 31        | 34        | 59        | 34        | 23        | 41        | 38        | 18        | 24        | 12        | 5         | 23        | 342        |
| 2021      | 26        | 19        | 19        | 21        | 23        | 27        | 33        | 18        | 21        | 26        | 24        | 26        | 283        |
| <b>합계</b> | <b>57</b> | <b>53</b> | <b>78</b> | <b>55</b> | <b>46</b> | <b>68</b> | <b>71</b> | <b>36</b> | <b>45</b> | <b>40</b> | <b>41</b> | <b>75</b> | <b>665</b> |

호황 관련 용어도 2019년 10월부터 검색되었으며, 2021년 12월 31일까지 총 62개의 뉴스기사가 검색되었다.

[표 3-19] 전주MBC 호황 관련 용어 스크래핑 결과

| 구분        | 1        | 2        | 3        | 4        | 5        | 6        | 7        | 8        | 9        | 10       | 11        | 12       | 합계        |
|-----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|----------|-----------|
| 2019      |          |          |          |          |          |          |          |          |          | 1        | 4         | 1        | 6         |
| 2020      | 0        | 1        | 0        | 1        | 5        | 2        | 3        | 1        | 1        | 0        | 2         | 2        | 18        |
| 2021      | 2        | 3        | 0        | 0        | 4        | 7        | 4        | 4        | 2        | 3        | 6         | 3        | 38        |
| <b>합계</b> | <b>2</b> | <b>4</b> | <b>0</b> | <b>1</b> | <b>9</b> | <b>9</b> | <b>7</b> | <b>5</b> | <b>3</b> | <b>4</b> | <b>12</b> | <b>6</b> | <b>62</b> |

경제 관련 용어 또한 2019년 10월부터 검색되었으며, 2021년 12월 31일까지 총 98개의 뉴스기사가 검색되었다.

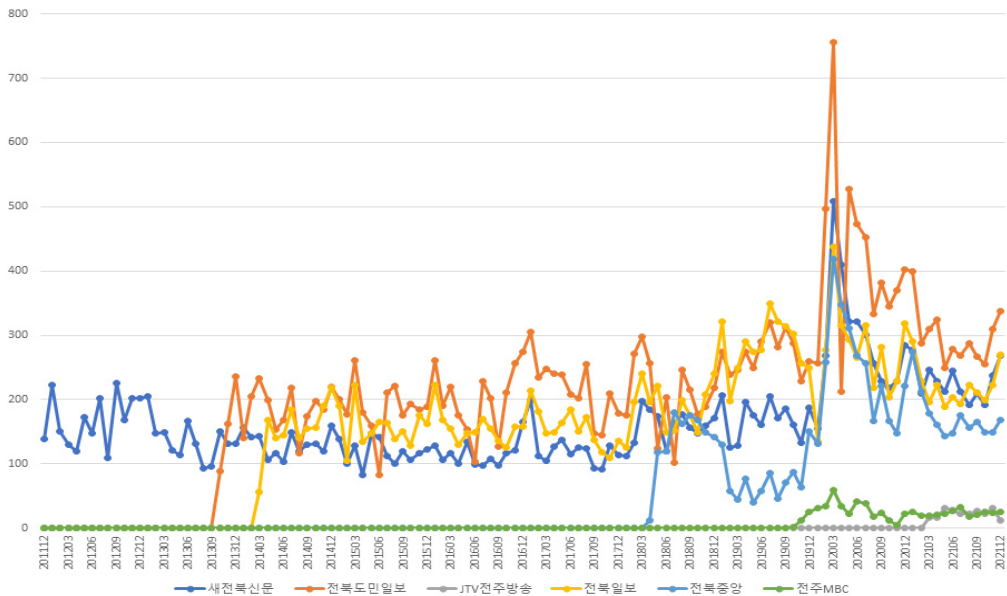
[표 3-20] 전주MBC 경제 관련 용어 스크래핑 결과

| 구분        | 1        | 2        | 3        | 4        | 5         | 6        | 7         | 8        | 9        | 10       | 11        | 12        | 합계        |
|-----------|----------|----------|----------|----------|-----------|----------|-----------|----------|----------|----------|-----------|-----------|-----------|
| 2019      |          |          |          |          |           |          |           |          |          | 1        | 4         | 4         | 9         |
| 2020      | 2        | 7        | 5        | 4        | 7         | 6        | 6         | 3        | 0        | 0        | 3         | 4         | 47        |
| 2021      | 1        | 1        | 0        | 2        | 6         | 3        | 4         | 4        | 3        | 7        | 8         | 3         | 42        |
| <b>합계</b> | <b>3</b> | <b>8</b> | <b>5</b> | <b>6</b> | <b>13</b> | <b>9</b> | <b>10</b> | <b>7</b> | <b>3</b> | <b>8</b> | <b>15</b> | <b>11</b> | <b>98</b> |

## 나. 신문사별 뉴스기사 시계열 추이

### 1) 불황 관련 용어 기사 수 추이 및 상관분석

6개 신문사의 불황 관련 뉴스기사 추이를 살펴보면, 6개 신문사의 기사 수 양 차이는 있으나, 그 변화 양상은 유사한 것으로 보인다. 특히 전북도민일보와 전북일보의 기사 변동폭이 거의 유사한 것을 알 수 있다. JTV전주방송과 전주MBC도 비슷한 추이를 보이나 기사 수가 다른 신문사와 차이가 큰 것을 알 수 있다. 그러나 전북중앙신문은 초기 발생시점에서 2019년 11월까지의 데이터 추이 양상이 다른 신문들과 반대로 나왔으며, 그 이후에는 유사한 형태를 보이는 것을 확인할 수 있다.



[그림 3-2] 6개 신문사 불황 관련 기사 수 추이

신문사간의 상관관계를 정량적으로 확인하기 위하여 상관관계 분석을 실시한 결과, 먼저 새전북신문은 전북도민일보, 전북일보 그리고 전북중앙신문과의 상관관계가 0.8 수준의 높은 상관관계를 보였다. 그중에서도 전북중앙신문과는 0.84의 높은 상관관계를 나타냈다. 전북도민일보는 전북일보와 전북중앙신문과 평균 0.68 정도 수준으로 나타났으며, 이중에서도 전북일보와는 0.73으로 가장 높게 나타났다. 마지막으로 전북일보와 전

북중앙신문의 상관관계는 0.31로 가장 낮게 나타났다. 상관계수 결과를 정리하면 다음과 같다.

[표 3-21] 불황 관련 용어 상관관계 분석 결과

| 구분     | 새전북신문 | 전북도민일보 | 전북일보  | 전북중앙신문 |
|--------|-------|--------|-------|--------|
| 새전북신문  | 1.000 | 0.801  | 0.759 | 0.845  |
| 전북도민일보 | 0.801 | 1.000  | 0.730 | 0.627  |
| 전북일보   | 0.759 | 0.730  | 1.000 | 0.307  |
| 전북중앙신문 | 0.845 | 0.627  | 0.307 | 1.000  |

주: 신문사별 스크래핑 기간이 다르므로 시계열을 조정하여 상관관계를 분석하였음

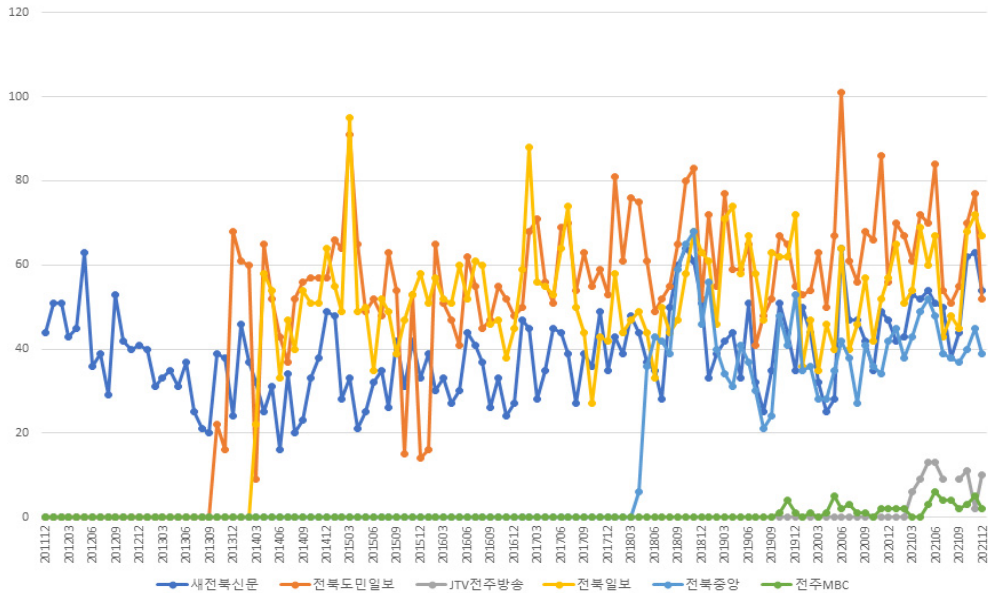
상관관계가 가장 높은 신문사끼리 묶어보면, 새전북신문-전북중앙신문과 전북도민일보-전북일보로 묶을 수 있다. 다만 새전북신문의 경우는 전북도민일보 및 전북중앙신문과 0.8 이상의 높은 상관관계를 보이는 것을 확인할 수 있다.

[표 3-22] 불황 관련 경제용어 상관성이 가장 높은 신문사 묶음

| 구분     | 새전북신문 | 전북도민일보 | 전북일보 | 전북중앙신문 |
|--------|-------|--------|------|--------|
| 새전북신문  |       |        |      |        |
| 전북도민일보 |       |        |      |        |

## 2) 호황 관련 용어 기사 수 추이 및 상관분석

6개 신문사의 호황 관련 뉴스기사 추이를 살펴보면, 6개 신문사는 기사 수의 차이는 있으나 전체적으로 비슷한 변동 양상을 보이는 것을 알 수 있다. 특히 2018년 3월 이후부터 전북중앙신문은 새전북신문 기사의 수와 변동이 매우 유사한 것을 볼 수 있다. 또한 불황 관련 단어의 경우와 마찬가지로 전주MBC와 JTV전주방송의 경우는 기사의 수가 확연히 작지만, 대체로 다른 신문사들과 유사한 추이를 보이는 것을 확인할 수 있다.



[그림 3-3] 6개 신문사 호황 관련 기사 수 추이

호황 관련 용어의 상관관계 분석 결과, 호황 관련 상관계수는 불황의 경우와는 다르게 전체적으로 0.5 이하로 낮게 나타났다. 그러나 신문사별 상관도는 불황 관련 경제용어의 결과와 같은 것을 알 수 있다. 구체적으로 새전북신문과 전북중앙신문이 상관성이 높고, 전북도민일보와 전북일보가 상관성이 높게 나타났으며, 전북일보와 전북중앙신문이 다른 신문사에 비해 상관성이 낮은 것으로 나타났다.

[표 3-23] 호황 관련 용어 상관관계 분석 결과

| 구분     | 새전북신문 | 전북도민일보 | 전북일보  | 전북중앙신문 |
|--------|-------|--------|-------|--------|
| 새전북신문  | 1.000 | 0.405  | 0.356 | 0.479  |
| 전북도민일보 | 0.405 | 1.000  | 0.443 | 0.331  |
| 전북일보   | 0.356 | 0.443  | 1.000 | 0.345  |
| 전북중앙신문 | 0.479 | 0.331  | 0.345 | 1.000  |

주: 신문사별 스크래핑 기간이 다르므로 시계열을 조정하여 상관관계를 분석하였음

상관관계가 가장 높은 신문사끼리 묶어보면, 새전북신문-전북중앙신문과 전북도민일보-전북일보로 묶을 수 있다. 그러나 불황 관련 경제용어와는 다르게 새전북신문-전북도민

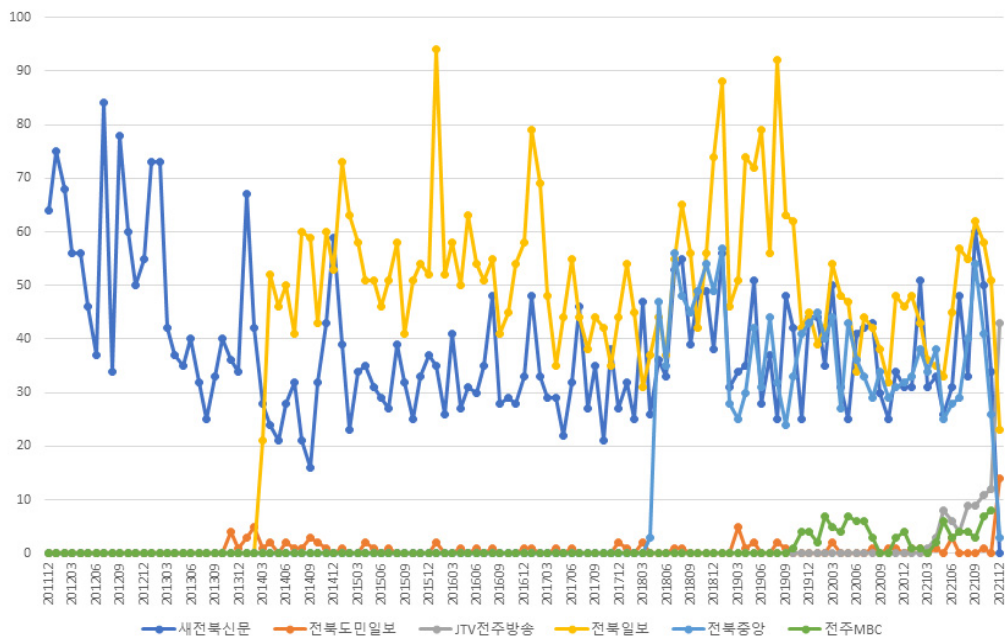
일보, 새전북신문-전북중앙신문, 전북도민일보-전북일보 모두 0.4 이상의 상관관계를 보이는 것을 확인할 수 있다.

[표 3-24] 호황 관련 경제용어 상관성이 가장 높은 신문사 묶음

| 구분     | 새전북신문 | 전북도민일보 | 전북일보 | 전북중앙신문 |
|--------|-------|--------|------|--------|
| 새전북신문  |       |        |      |        |
| 전북도민일보 |       |        |      |        |

### 3) 경제 관련 용어 기사 수 추이 및 상관분석

6개 신문사의 경제 관련 뉴스기사 추이를 살펴보면, 6개 신문사는 초기 신문사마다 변화의 형상이 서로 다르게 나타나는 것을 볼 수 있다. 이후 2017년 후반부터는 모든 신문사가 대체로 유사한 형태의 기사 수 변화를 보여주고 있다. 불황 관련 단어의 경우와 마찬가지로 전주MBC와 JTV전주방송의 경우는 기사 수의 차이는 보이나 유사한 추이를 보이는 것을 확인할 수 있다.



[그림 3-4] 6개 신문사 경제 관련 기사 수 추이

경제 관련 용어의 상관관계 분석 결과, 경제 관련 상관계수는 호황의 경우처럼 계수값이 전체적으로 0.5 이하로 낮게 나타났다. 그러나 신문사별 상관도는 호황 관련 용어의 결과와 같은 것을 알 수 있다. 구체적으로 새전북신문과 전북중앙신문이 상관성이 높고, 전북도민일보와 전북일보가 상관성이 높게 나타났으며, 전북일보와 전북중앙신문이 다른 신문사에 비해 상관성이 낮은 것으로 나타났다.

[표 3-25] 경제 관련 용어 상관관계 분석 결과

| 구분     | 새전북신문 | 전북도민일보 | 전북일보  | 전북중앙신문 |
|--------|-------|--------|-------|--------|
| 새전북신문  | 1.000 | 0.405  | 0.356 | 0.479  |
| 전북도민일보 | 0.405 | 1.000  | 0.443 | 0.331  |
| 전북일보   | 0.356 | 0.443  | 1.000 | 0.340  |
| 전북중앙신문 | 0.479 | 0.331  | 0.340 | 1.000  |

주: 신문사별 스크래핑 기간이 다르므로 시계열을 조정하여 상관관계를 분석하였음

상관관계가 가장 높은 신문사끼리 묶어보면, 새전북신문-전북중앙신문과 전북도민일보-전북일보로 묶을 수 있다. 그러나 호황 관련 용어의 경우에서처럼 새전북신문-전북도민일보, 새전북신문-전북중앙신문, 전북도민일보-전북일보 모두 0.4 이상의 상관관계를 보이는 것을 확인할 수 있다.

[표 3-26] 경제 관련 경제용어 상관성이 가장 높은 신문사 묶음

| 구분     | 새전북신문 | 전북도민일보 | 전북일보 | 전북중앙신문 |
|--------|-------|--------|------|--------|
| 새전북신문  |       |        |      |        |
| 전북도민일보 |       |        |      |        |

---

## 다. 뉴스 키워드 데이터 수집 및 결과 종합

본 장에서는 새전북신문, 전북도민일보, 전북일보, 전북중앙신문, JTV전주방송, 전주MBC의 기사 데이터를 구축하고, 구축된 자료를 바탕으로 기사 수 추이 및 신문사별 상관관계를 분석하였다. 기사는 파이썬을 활용한 뉴스기사 스크래핑 프로그램 자체적으로 만들어 수집하였고, 신문사별 상관관계 분석은 구축된 데이터가 부족한 JTV전주방송과 전주MBC를 제외한 새전북신문, 전북도민일보, 전북일보, 전북중앙신문만을 이용하였다.

분석 결과, 불황 관련 용어에서는 0.8 이상의 높은 상관관계를 보였으나, 호황과 경제용어 기사의 상관관계는 0.4 이하의 낮은 상관관계를 보였다. 특히 호황과 경제용어 관련 기사의 미디어별 상관계수가 매우 유사하게 나타났다. 이러한 결과를 볼 때, 전라북도의 호황 및 경제용어 관련 기사에 대한 신문사의 반응은 유사하다고 할 수 있다.

신문사별 상관관계를 보면 신문사별 높은 상관성을 보이는 그룹을 볼 수 있었다. 구체적으로 새전북신문과 전북중앙신문이 불황, 호황 그리고 경제용어 관련 분석에서 높은 상관도를 보였다. 마찬가지로 전북도민일보와 전북일보가 불황, 호황 그리고 경제용어 관련 분석에서 높은 상관도를 보였다. 이러한 결과를 볼 때, 새전북신문과 전북중앙신문이 각각의 용어에 대한 기사 수의 변동이 유사하고 전북도민일보와 전북일보의 기사 수의 변동이 유사함을 알 수 있었다.

---

### 3. 뉴스 빅데이터 지역경기지수 작성

---

#### 가. 개발 전제

##### 1) 경기상황 반영 데이터 추출

전북뉴스 경기종합동행지수 개발을 위해서는 수집된 뉴스데이터가 전북경제를 잘 반영하는지가 가장 중요하다. 다시 말해, 전라북도 경기종합동행지수, 전라북도 경기실사지수 및 소비자동향지수, 경제심리지수와 상관성이 있어야 한다. 하지만 앞서 수집된 데이터는 경기상황과 상관없이 수집된 데이터이므로, 경기상황을 잘 반영할 수 있는 데이터를 추출하여야 한다.

##### 2) 데이터 활용 시점

스크롤링된 뉴스데이터는 가장 긴 데이터가 2011년 12월부터 시작되며, 가장 짧은 데이터는 2019년 2월부터 시작된다. 이렇듯 신문사 및 단어별 수집시점이 다르기 때문에, 분석을 위해서는 데이터의 시점을 통일할 필요가 있다. 그러나 시점을 맞추는 경우, 활용할 수 있는 데이터의 시계열이 짧아지고 그 양도 적어진다. 따라서 “전북뉴스 지역경기지수”를 개발하기에는 부족할 수 있다. 따라서 본 연구에서는 시점을 통일하지 않고 수집된 모든 데이터를 전부 활용하도록 한다.

##### 3) 전북뉴스 지역경기지수 산정 방법

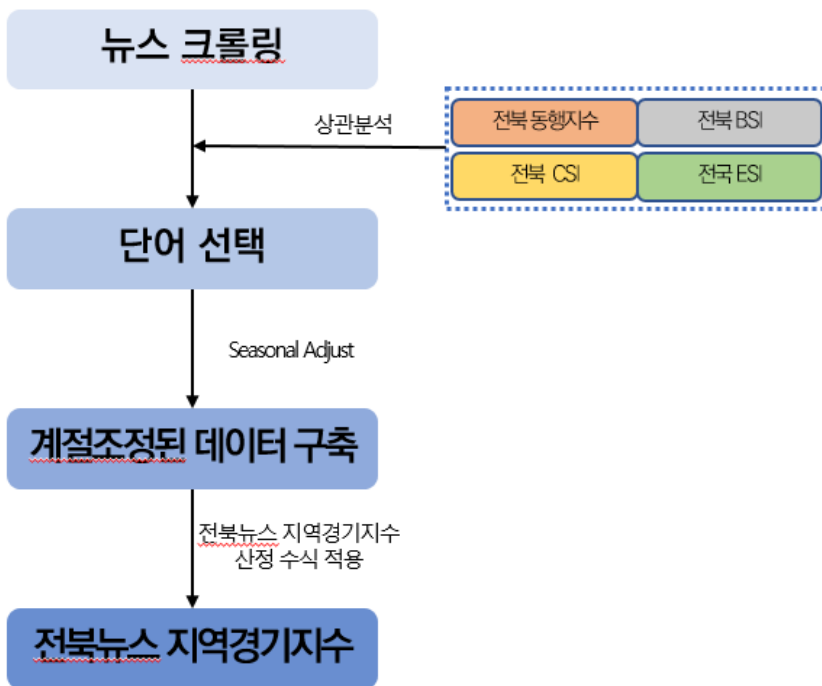
전북뉴스 지역경기지수는 뉴스데이터의 불황, 호황에 대한 기사를 활용하는 것이므로, 일반적인 지수산정방식을 사용하기 어렵다. 따라서 본 연구서는 이공희(2017)의 연구에서 사용한 방법을 따른다.



## 나. 개발 과정

전북뉴스 경기지수의 개발 과정은 총 세 단계로 구성된다. 첫 번째는 수집된 53개 단어 중에서 전북지역 경기를 잘 반영하는 단어를 선택하는 단계이다. 다시 말해, 53개 단어별 총 데이터와 전북 경기종합동행지수 순환변동치, 전북 기업실사지수(BSI, 업황), 전북 소비자동향지수(BSI, 생활형편), 전국 경제심리지수(ESI)와 상관성이 있는 데이터를 선택한다.

둘째는 시계열 조정이다. 최종 선택된 데이터는 시계열 데이터이므로 데이터의 안정성을 위해서는 시계열 조정이 필요하다. 세 번째는 계절조정된 데이터를 사용하여 전북뉴스 지역경기지수를 개발하는 것이다. 이러한 과정을 그림으로 표현하면 아래와 같다.

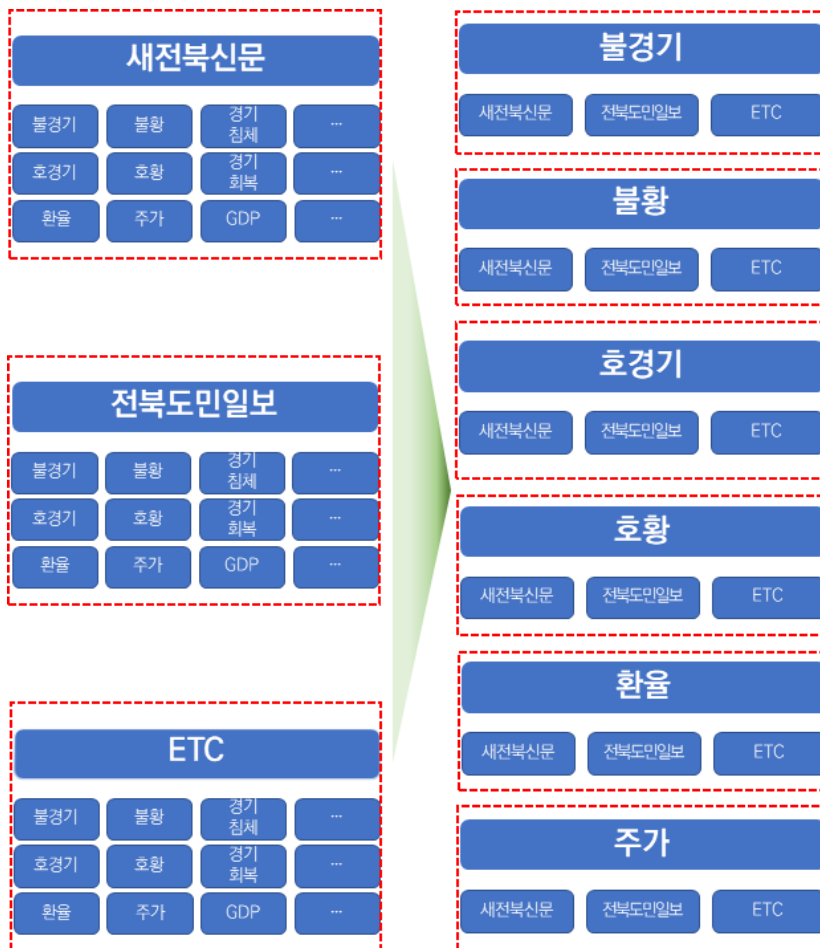


[그림 3-5] 전북뉴스경기종합지수 개발 FrameWork

## 다. 뉴스 데이터와 경기지수 상관분석

### 1) 데이터 구축

데이터 현황분석에서는 불황, 호황 및 경제단어에 대한 총계 데이터만을 제시하여 시간적 흐름을 살펴보았다. 하지만 본 장에서는 경기지수와 상관분석을 위하여 단어별 데이터로 전환한다. 구체적으로 신문사별 수집된 단어의 시계열 데이터를 단어별로 정리하는 것이다. 이를 그림으로 표현하면 아래와 같다.



[그림 3-6] 상관분석을 위한 데이터 구축 방식(일부 표현)

## 2) 상관분석

상관분석은 53개 단어의 시계열 데이터들과 전북 경기종합동행지수 순환변동치, 전북 기업실사지수, 전북 소비자동향지수 그리고 전북 경제심리지수에 대해 시행하였다. 결과를 요약하여 나타내면 아래의 표와 같다<sup>5)</sup>.

상관분석 결과를 살펴보면, 동행지수는 최대 0.48 최소 0.0042이며, 중앙값은 0.11, 평균값은 0.19 그리고 표준편차는 0.15로 나타났다. 전북 기업경기실사지수는 최대 0.44, 최소 0.025, 중앙값 0.20, 평균 0.16 그리고 표준편차는 0.14로 나타났다. 전북 소비자동향지수는 최대 0.45, 최소 0.003, 중앙값 0.11, 평균 0.16 그리고 표준편차는 0.14로이다. 마지막으로 전국 경제심리지수는 최대 0.62 최소 0.015, 중앙값 0.17, 평균 0.23 그리고 표준편차 0.17로 나타났다.

이러한 결과를 볼 때, 뉴스데이터와 상관성이 가장 높은 지수는 전국 경제심리지수라는 것을 알 수 있다. 그 다음으로는 전북 동행지수 순환변동치이다. 그리고 표준편차가 가장 작은 지수는 전북 기업경기실사지수와 전북 소비자동향지수로 나타났다.

[표 3-27] 상관분석 결과 요약

| 구분   | 전북 동행지수<br>순환변동치 | 전북<br>기업경기실사지수 | 전북<br>소비자동향지수 | 전국<br>경제심리지수 |
|------|------------------|----------------|---------------|--------------|
| 최소   | 0.0042           | 0.0247         | 0.0032        | 0.0146       |
| 최대   | 0.4823           | 0.4351         | 0.4461        | 0.6215       |
| 중앙값  | 0.1143           | 0.2002         | 0.1078        | 0.1723       |
| 평균   | 0.1927           | 0.2104         | 0.1594        | 0.2257       |
| 표준편차 | 0.1514           | 0.1086         | 0.1377        | 0.1717       |

5) 47개 단어와 경기지수변수들과의 전체 상관분석결과는 부록 표를 참조하기 바란다.

### 3) 단어 선택

#### 가) 선택의 기준지수 선정

앞에서 제시되었듯이, 뉴스데이터와 상관성이 높게 나타난 경기지수는 전북 동행지수 순환변동치와 전국 경제심리지수이다. 그리고 표준편차가 가장 작은 지수는 전북 기업경기실사지수와 전북 소비자동향지수이다. 이렇듯 상관분석 결과의 대표 통계량만을 가지고 어떤 지수를 이용해야 할지 확정하기 어려운 상황이다. 따라서 본 연구에서는 상관계수 값이 특정 조건 이상인 값에서 가장 많은 단어가 포함된 지수로 결정한다.

본 연구에서 상관계수 값의 특정 조건은 상관계수의 절댓값이 0.4 이상인 경우로 한정한다. 이러한 이유는 4개의 지수 모두 최댓값이 0.4 이상으로 나타나 모든 지수를 다 포함하여 조건을 적용할 수 있기 때문이다<sup>6)</sup>. 분석 결과, 동행지수는 7개의 단어, 기업경기실사지수는 2개의 단어, 소비자동향지수는 4개, 경제심리지수는 9개로 나타났다. 따라서 본 연구에서는 전북 동행지수 순환변동치와 전국 경제심리지수를 선택기준 지수로 한다.

[표 3-28] 상관계수 0.4 이상인 경기지수별 단어 개수와 단어별 상관계수

| 전북 동행지수<br>순환변동치 |      | 전북<br>기업경기실사지수 |      | 전북<br>소비자동향지수 |      | 전국<br>경제심리지수 |      |
|------------------|------|----------------|------|---------------|------|--------------|------|
| 개수               | 7    | 개수             | 2    | 개수            | 4    | 개수           | 9    |
| 경기침체             | 0.48 | 경기침체           | 0.44 | 경기침체          | 0.41 | 경기침체         | 0.59 |
| 금융위기             | 0.43 | 금리인하           | 0.42 | 위기            | 0.40 | 위기           | 0.49 |
| 경제위기             | 0.41 |                |      | 경제위기          | 0.45 | 금융위기         | 0.42 |
| 해고               | 0.48 |                |      | 한국은행          | 0.41 | 경제위기         | 0.62 |
| 금리인하             | 0.45 |                |      |               |      | 해고           | 0.53 |
| 거품               | 0.42 |                |      |               |      | 거품           | 0.42 |
| 주가               | 0.43 |                |      |               |      | 주가           | 0.52 |
|                  |      |                |      |               |      | 한국은행         | 0.47 |
|                  |      |                |      |               |      | 신용등급         | 0.49 |

6) 조건을 0.5이상으로 조정하였을 경우, 전국 경제심리지수에서만 만족하는 결과가 나타나서 조건을 0.4로 낮추었다.

나) 선택의 기준지수 적용

최적 단어의 선택은 두 기준지수가 특정 조건을 동시에 만족하는 경우로 한정한다. 구체적으로 전국 경제심리지수가 0.4 이상이고 전북 동행지수 순환변동치가 0.4 이상인 경우로 한정한다는 것이다. 이러한 이유는 하나의 지수를 만족하지만 다른 지수에서는 상관성이 너무 낮아 분석의 효율성을 저해할 수 있기 때문이다<sup>7)</sup>.

위 기준을 적용하여 분석하면, 아래와 같이 총 9개의 단어가 조건을 충족한다. 구체적으로 불황 단어는 경기침체, 위기, 금융위기, 경제위기, 해고 5개이고, 호황 단어는 거품 1개이며, 마지막으로 경제단어는 주가, 한국은행, 신용등급 3개로 나타났다. 각 단어별 전북동행지수 순환변동치와 전국 경제심리지수 상관계수를 정리하면 아래의 표와 같다.

[표 3-29] 조건에 부합하는 단어와 상관계수

| 구분   | 전북 동행지수 순환변동치 | 전국 경제심리지수 |
|------|---------------|-----------|
| 경기침체 | 0.482         | 0.587     |
| 금융위기 | 0.434         | 0.421     |
| 경제위기 | 0.409         | 0.622     |
| 해고   | 0.481         | 0.525     |
| 거품   | 0.421         | 0.421     |
| 주가   | 0.434         | 0.517     |

최종적으로 선택하는 단어는 불황 단어를 중심으로 하되, 호황 단어와 연계하여 선택한다. 왜냐하면 뉴스 기사는 호황보다는 불황, 위기에 주목하기 때문이다. 이에 따라 불황 단어는 “경기침체”와 “해고”를 선택하고 이에 대응하는 호황 단어는 “경기회복”과 “채용” 2개의 단어를 우선 선택한다<sup>8)</sup>.

[표 3-30] 불황 중심 단어쌍

| 구분     | 키워드 |          |
|--------|-----|----------|
| 2개 키워드 | 불황  | 경기침체, 해고 |
|        | 호황  | 경기회복, 채용 |

7) 경제심리지수를 0.5, 전북 동행지수 순환변동치를 0.4로 기준을 정할 경우, 이를 충족하는 단어는 “경기침체”, “경제위기”, “해고”, “주가” 4개로 나타났다. 또한 기준을 두 지수 모두 0.4로 할 경우, 기준을 충족하는 단어는 “경기침체”, “금융위기”, “경제위기”, “해고”, “거품”, “주가”로 나타났다.

8) 단어를 2쌍만 선택한 것은 불황 단어와 호황 단어 쌍이 이루어질 수 있는지를 고려하였기 때문이다.

불황 중심으로 선택된 2개의 키워드로는 지역 경기를 파악하기에는 다소 부족할 것으로 판단하여 불황 단어와 호황 단어의 상관계수 값의 차이가 큰 단어를 추가로 선택한다. 차이가 큰 단어는 “금리 인하”와 “금리 인상”, “디플레이션”과 “인플레이션”으로 나타났다.

[표 3-31] 불황 단어와 호황 단어 상관계수 값 차이

| 구분    | 동행지수  | BSI   | CSI   | ESI_national |
|-------|-------|-------|-------|--------------|
| 불경기   | -0.07 | -0.02 | -0.06 | -0.02        |
| 불황    | -0.15 | 0.21  | -0.29 | -0.08        |
| 경기불황  | 0.03  | 0.06  | -0.07 | 0.12         |
| 경제불황  | 0.12  | 0.12  | 0.09  | 0.09         |
| 적자    | 0.31  | 0.36  | -0.15 | 0.25         |
| 폭락    | 0.09  | 0.42  | -0.17 | 0.20         |
| 하한가   | -0.03 | 0.06  | -0.10 | 0.01         |
| 정리해고  | 0.04  | -0.11 | 0.09  | -0.03        |
| 금리인하  | 0.78  | 0.54  | 0.26  | 0.72         |
| 디플레이션 | 0.53  | 0.57  | 0.13  | 0.39         |

최종 선정 단어는 아래의 표와 같이 “경기침체”, “해고”, “금리인하”, “디플레이션” 4개의 단어쌍으로 구성된다.

[표 3-32] 최종 선정 키워드 집합

| 구분     | 키워드 |                       |
|--------|-----|-----------------------|
| 4개 키워드 | 불황  | 경기침체, 해고, 금리인상, 디플레이션 |
|        | 호황  | 경기회복, 채용, 금리인하, 인플레이션 |

## 라. 시계열 조정

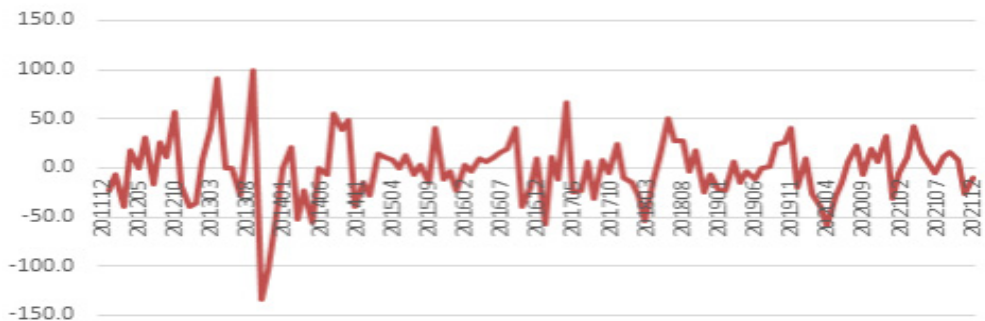
일반적으로 시계열 데이터는 추세, 순환, 불규칙 변동 등 불안정적 요인들을 내포하고 있다. 이러한 요인들은 실제 경기동향을 왜곡할 수 있으므로 계절변동을 제거해야 정확한 전복뉴스 경기동행지수를 생산할 수 있다.

본 연구에서 계절조정방법은 X-13-ARIMA-SEATS의 X11필터를 이용하였다. 계절조정 결과 계절성을 얻을 수 있었던 시계열은 경기침체와 채용 두 시계열이었으며 나머지 계열에서는 계절조정 계열을 얻을 수 없었다. 이는 계열의 불안정성이 큰 요인으로 작용한 것으로 보이며 이는 count data의 특성과 시계열의 짧음으로 인한 것으로 판단된다.

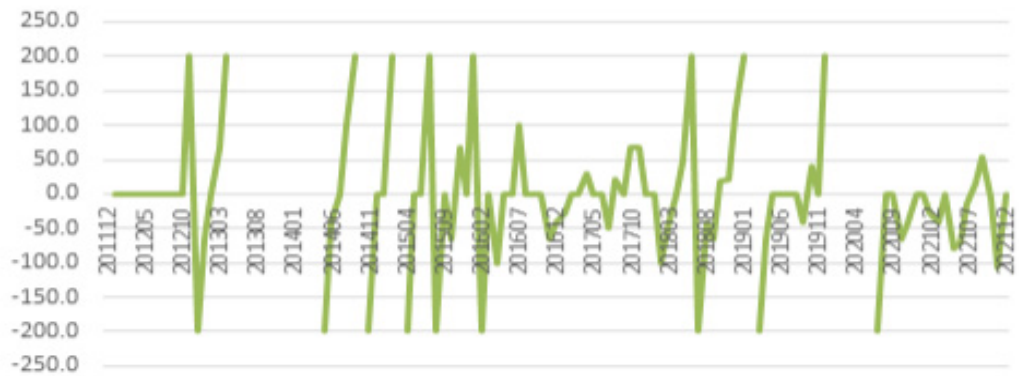
아래 그림은 각 계열의 변화율을 표시한 그림으로 일부 시계열에서는 단절된 현상을 볼 수 있었으며 이로 인해 적절한 계절조정을 얻을 수 없어 원 계열을 활용했다.



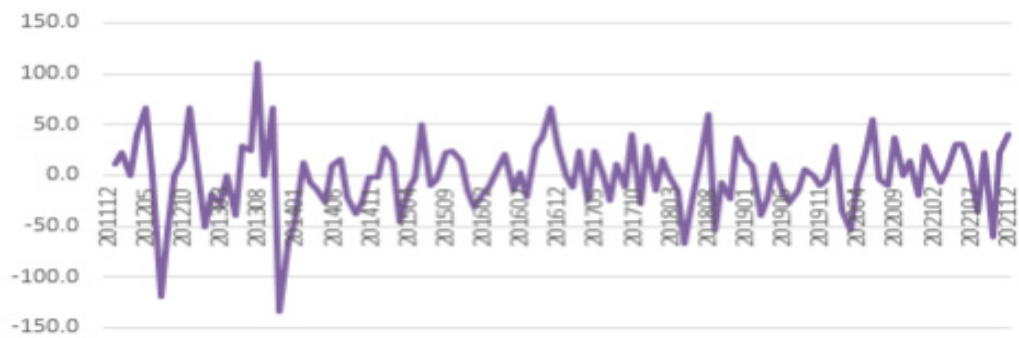
[그림 3-7] 경기침체 변화율



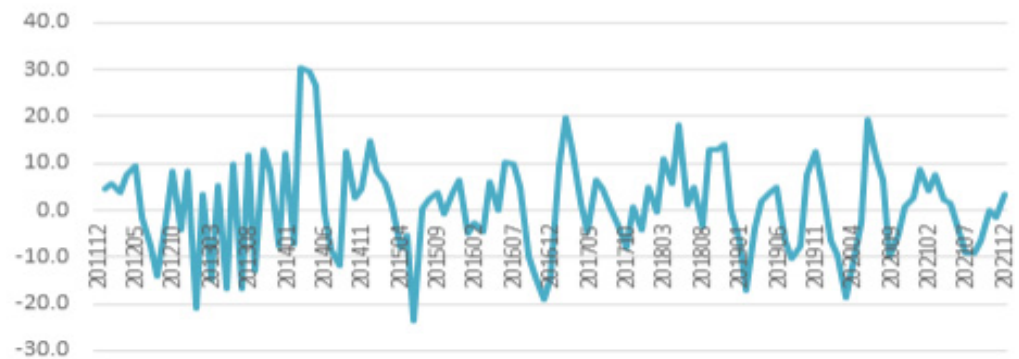
[그림 3-8] 해고 변화율



[그림 3-9] 인플레이션 변화율

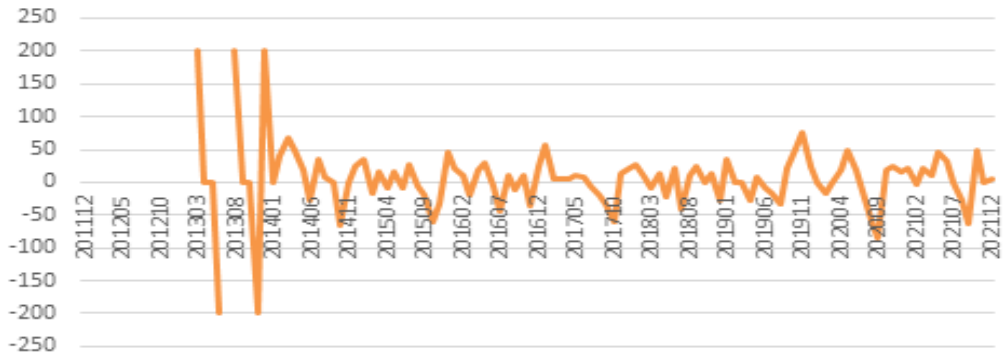


[그림 3-10] 금리인하 변화율



[그림 3-11] 채용 변화율

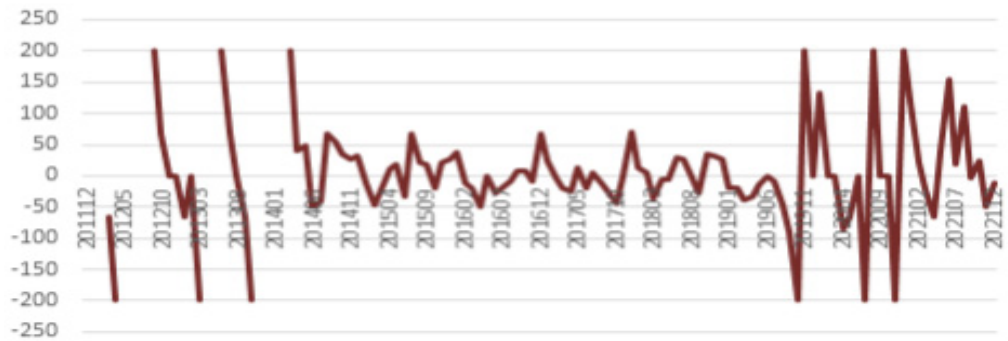




[그림 3-12] 경기회복 변화율



[그림 3-13] 디스플레이션 변화율



[그림 3-14] 금리인상 변화율

## 마. 전북뉴스 지역경기(종합)지수 작성

### 1) 전북뉴스 지역경기지수(BSI 방식) 작성

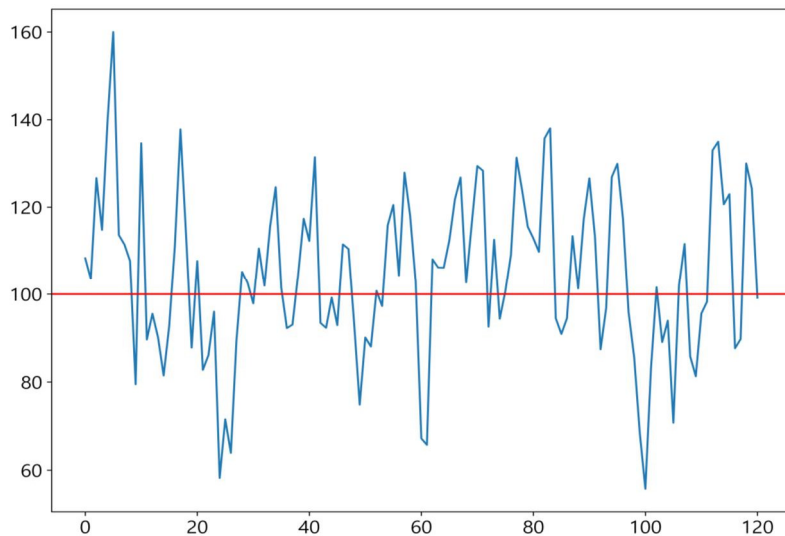
전북뉴스 지역경기종합지수(BSI) 작성방법은 이금희(2017)의 연구에서 사용된 것을 적용한다. 즉, 호황지표와 불황지표의 차를 호황지표와 불황지표의 합으로 나누어 BSI 방식을 적용하였다. 구체적인 식은 아래의 식과 같다.

$$JNI = \frac{\text{호황지표} - \text{불황지표}}{\text{호황지표} + \text{불황지표}} \times 100 + 100$$

위 식을 통해 산출된 전북뉴스 지역경기지수(BSI)와 그래프는 다음과 같다.

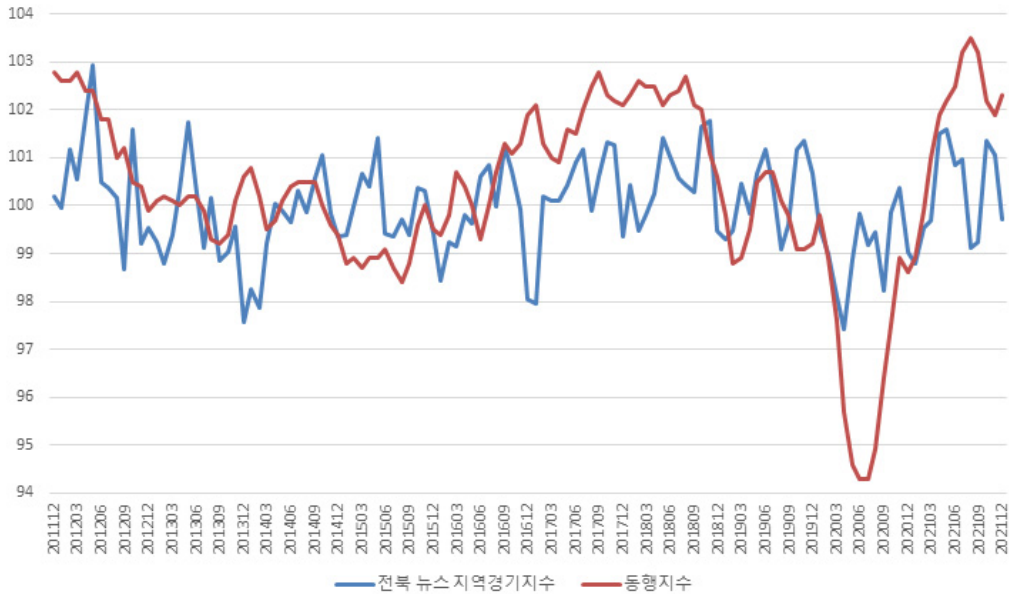
[표 3-33] 전북뉴스 지역경기지수(BSI) 산출 결과

| 구분                | 2011.12 | 2012.1 | 2012.2 | 2012.3 | ... | 2021.9 | 2021.10 | 2021.11 | 2021.12 |
|-------------------|---------|--------|--------|--------|-----|--------|---------|---------|---------|
| 전북뉴스 지역경기 지수(BSI) | 108.33  | 103.57 | 126.66 | 114.81 | ... | 89.75  | 130.00  | 124.27  | 99.137  |



[그림 3-15] 전북뉴스 지역경기지수(BSI)

작성된 “전북뉴스 지역경기지수(BSI)”와 전라북도 경기종합동행지수의 방향성을 확인하기 위하여 그래프로 나타내면 다음 표와 같다. 비교 결과, 전북뉴스 지역경기지수(BSI)와 전라북도 경기종합 동행지수는 유사한 사이클을 갖는 것을 확인할 수 있다<sup>9)</sup>.



[그림 3-16] 전북뉴스 지역경기지수(BSI)와 동행지수 순환변동치

전북뉴스 지역경기지수와 전라북도 경기종합동행지수, 전라북도 기업경기실사지수(BSI), 전라북도 소비심리지수(CSI), 그리고 전국 경제심리지수(ESI)에 대한 상관성을 분석하면 전라북도 경기종합동행지수와 상관성이 높게 나타났다. 그러나 상관계수 값은 0.38로 낮은 상관성을 보였다.

[표 3-34] 전북뉴스 지역경기지수(BSI)와 각 지수와의 상관관계

| 구분               | 동행지수 | BSI  | CSI  | ESI  |
|------------------|------|------|------|------|
| 전북뉴스 지역경기지수(BSI) | 0.38 | 0.31 | 0.31 | 0.36 |

9) 전라북도 경기종합동행지수와 진폭을 맞추기 위하여 “전북뉴스 지역경기지수(BSI)”를 정규화한 데이터를 이용하였다.

---

## 2) 전북뉴스 지역경기종합지수(Composite Index 방식) 작성

### 가) 전북뉴스 지역경기종합지수 작성 개요

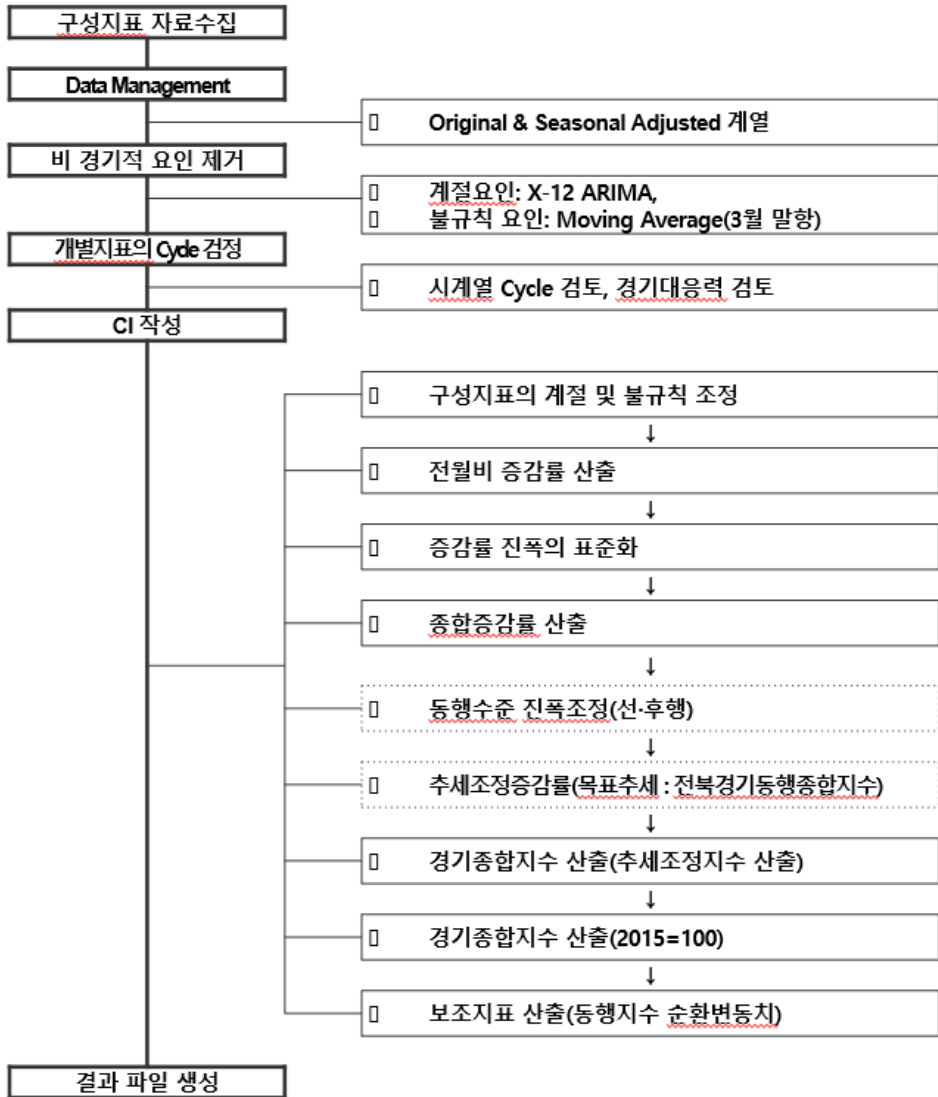
BSI 형식의 경기종합지수는 긍정과 부정의 차를 이용한 간편성이 있는 반면에 데이터 간의 비중에 차이가 있는 것을 포착하지 못하는 한계가 있으므로 통계청에서 사용하는 경기동행종합지수 작성방법을 활용하여 비교할 필요가 있다.

뉴스 경기종합지수로 활용된 데이터는 위의 BSI와 동일한 8개의 series로 구성하여 Composite index를 작성하였다. 일반적으로 경기지수 작성시 각 지표별로 데이터의 시계열 사이클의 검토와 경기대응력을 판단하여 최종지표를 선정하는 과정을 거치고 있으나 뉴스데이터를 활용하는 선행연구에서 수행한 방법인 BSI와의 비교검토를 위해서 경기종합지수 산정에 BSI와 동일한 지표를 활용하였다.

Composite Index 작성방법은 통계청에서 제공하는 작성매뉴얼에 따라 작성하였으며 KOSTAT-CIS 프로그램을 활용하여 지수산정, 순환변동치, 계절조정 등을 수행하였다.

### 나) 전북뉴스 지역경기종합지수 작성 과정

전북뉴스 지역경기종합지수 작성과정은 통계청의 지역경기종합지수 실무가이드를 따라서 작성하고 작성과정은 아래 표 3-9와 같다. 지수작성을 위해서 구성지표의 자료를 수집하고 데이터의 원계열로 활용할 것인지와 계절조절이 필요한지에 대한 판단을 한다. 비경기적 요인을 제거하기 계절요인은 X-12 ARIMA를 활용하고 불규칙 요인은 이동평균법을 활용한다. 이후, 개별지표의 Cycle을 검정하여 시계열 사이클과 경기대응력을 검토한다. CI 작성은 구성지표의 계절 및 불규칙조정 → 전월비 증감률 산출 → 증감률 진폭의 표준화 → 종합증감률 산출 → 추세조정지수 산출 → 경기종합지수 산출 → 동행지수 순환변동치 산출 등의 과정을 통해 결과를 도출한다.



[그림 3-17] 전북뉴스 지역경기종합지수 작성 흐름도

## ■ 구성지표 자료 수집

전북뉴스 지역경기종합지수의 산정은 구성지표 자료 수집을 함으로써 시작된다. 경기 관련 기사 키워드를 선정하고 그에 대한 DB를 구축하는 것은 많은 시간과 노력이 필요

하다. DB구축 관련해서 빅카인즈와 같은 유료서비스를 활용하면 상대적으로 용이할 수 있으나 지수를 유지관리 하는 차원에서는 지속가능하지 않아 web crawling을 통해 DB를 구축하기로 하였다.

호황, 불황, 경제관련 용어의 데이터 계열은 부록에 수록하였다. 데이터 계열에 대한 분석을 통해 활용가능성을 검토하고 각 계열별 경기대응력을 판단하여 지수산정을 위해 활용할 계열을 선정하는 것이 일반적인 방법이나 전북뉴스 지역경기지수(BSI) 작성에서 활용한 방법과 동일한 8개의 지표를 선정하였다.

### ■ 구성지표의 계절 및 불규칙 조정

데이터 계열을 전북뉴스 지역경기종합지수에 활용하기 위해 개별 구성지표의 비경기적 요인을 제거하였다. 개별 경제지표의 원계열(X)을 분해하면 경기적 변동요인 + 비경기적 변동요인으로 분해 할 수 있다. 전자는 추세변동과 순환변동요인으로, 후자는 계절변동요인과 불규칙 변동요인(천재지변 등) 등으로 구성되어 있다.

$$\begin{aligned} \text{원계열}(X) &= \text{Trend}(T) \times \text{Cycle}(C) \times \text{Seasonal}(S) \times \text{Irregular}(I) \\ &\quad \text{【경기적 변동요인】} \qquad \qquad \qquad \text{【비경기적 변동요인】} \end{aligned}$$

비경기적 변동요인 중 계절변동요인은 X-13-ARIMA 방식을 이용하여 계절요인을 제거한다. 계절변동요인에는 명절효과(설, 추석)와 조업일수 효과를 반영하여 추계한다.

한편, 불규칙변동요인은 이동평균법을 활용하며 기준점에 따라 First Point, Centered, End Point 방법을 활용하며 이동평균 기간을 3개월 또는 6개월 Term을 선택한다. 필터<sup>10)</sup>의 활용은 Seasonal, Henderson, H-P 필터를 활용할 수 있으며 3 Centered Weight 방법<sup>11)</sup>도 활용가능하다. OECD는 불규칙 조정을 H-P 필터를 활용하나 우리나라에서는 3월 말항 이동평균을 일반적으로 활용하고 본 연구에서도 3월 말항을 활용하였다.

10) Seasonal 필터와 Henderson 필터는 주어진 기간의 이동평균방법이고 Hodrick-Prescott 방법은 정해진 기간보다 짧은 순환요소를 제거하는데 활용된다.

11) 3Centered\_weight는 Centered 방식을 보완한 것으로 3개월 중심항 이동평균 계산 시 양끝 결항보정을 위해 가중치를 임의로 부여하여 계산하는 방법이다.

## ■ 전월비 증감률 산출

전월대비 증감률(symmetric percent changes:  $Y_{i,t}$ ) 산출은 구성지표의 증가와 감소를 대칭적으로 처리하기 위해 전월치와 금월치의 평균을 분모로 하여 증감의 변화를 동일하게 산정하고자 하는 것이다.

Level Indicators는 아래와 같은 식으로 구하고,

$$Y_{i,t} = \frac{X_{i,t} - X_{i,t-1}}{X_{i,t} + X_{i,t-1}} \times 200$$

Ratio Indicators, 0 또는 음수는 다음과 같은 식을 활용한다.

$$Y_{i,t} = X_{i,t} - X_{i,t-1}$$

단,  $i = 1, 2, \dots, N$ (각 구성 지표),  $t = 2, 3, \dots, T$ (시점)

## ■ 증감률 진폭의 표준화 및 평균증감률 산출

개별지표별로 순환진폭(cyclical amplitude)의 특성이 다르므로 이를 지수화 하는데 고려할 사항으로는 개별지표가 지수를 크게 좌우하는 현상이 발생하지 않아야 한다. 개별지표의 순환진폭을 평균값 그대로 활용하면 진폭이 큰 지표가 지수를 dominate 하는 현상이 발생할 수 있다. 이러한 현상을 사전에 막기 위해서는 구성지표의 표준화 과정을 거쳐야 하며 이를 통해 순환진폭이 평균적으로 같아지게 된다.

표준화증감률 및 평균증감률은 아래와 같은 방식으로 산출한다.

표준화증감률 = 각 구성지표의 대칭변화율( $Y_{i,t}$ ) / 절대치평균인 표준화인자( $S_i$ )

$$Z_{i,t} = \frac{Y_{i,t}}{S_i}$$

단,  $S_i = \frac{\sum_{t=2}^T |Y_{i,t}|}{T-1}$ ,  $i = 1, 2, \dots, N$ (각 구성 지표),  $t = 2, 3, \dots, T$ (시점)

평균증감률( $AC_{i,t}$ ) = 표준화증감률( $Z_{i,t}$ )의 합/구성지표수

$$AC_t = \frac{\sum_{i=1}^N Z_{i,t}}{N}$$

단,  $i = 1, 2, \dots, N$ (각 구성지표),  $t = 2, 3, \dots, T$ (시점)

## ■ 지수화

경기지수의 원지수( $RCI_t$ )는 매월 평균증감률( $AC_t$ )의 누적값을 이용하여 산출하며 원지수( $RCI_t$ )를 기준년도(2015년)로 조정하여 최종 경기지수( $CI_t$ )를 산출한다.

$$RCI_t = RCI_{t-1} \times \frac{200 + AC_t}{200 - AC_t} \quad \text{where, } RCI_1 = 100$$

$$CI_t = \frac{RCI_t}{BASE} \times 100$$

단,  $BASE$ 는 기준년도의  $RCI_t$ 의 평균

### 다) 전북뉴스 지역경기종합지수 작성 결과

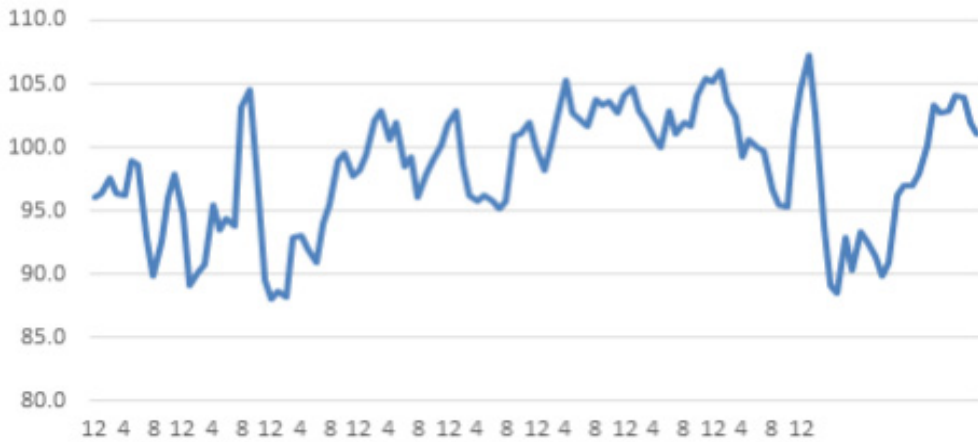
Composite Index 방식을 통해 산출된 “전북뉴스 지역경기종합지수”와 그래프는 다음과 같다. 전북뉴스 지역경기종합지수는 2021년 12월 101.1로 전월비 0.9pt 하락한 것을 볼 수 있다. 전북 경기동행종합지수와 비교해 볼 때 지수의 변화가 민감한 것을 볼 수 있으며 이는 데이터의 특성과 시계열의 부족에 기인한 것으로 보인다.

한편, BSI 방식과 Composite Index를 비교하면 후자가 더 안정적인 계열임을 볼 수 있다. BSI 방식의 경우 마지막 4개월의 편차 뿐 아니라 최초의 4개월 또한 편차가 상대적으로 큰 것을 볼 수 있었다.

[표 3-35] 전북뉴스 지역경기종합지수 산출 결과

| 구분                 | 2011.12 | 2012.1 | 2012.2 | 2012.3 | ... | 2021.9 | 2021.10 | 2021.11 | 2021.12 |
|--------------------|---------|--------|--------|--------|-----|--------|---------|---------|---------|
| 전북<br>뉴스경기종합지수(CI) | 96.0    | 96.4   | 97.6   | 96.3   | ... | 104.1  | 104.0   | 102.0   | 101.1   |





[그림 3-18] 전북뉴스 지역경기종합지수

BSI 방식에 비해 안정적이거나 그림에서 보는 바와 같이 뉴스경기종합지수의 순환변동치가 크게 움직이고 있어 데이터의 안정성 면에서는 경기동행종합지수에 비해 떨어지는 것을 볼 수 있다. 진폭의 차이가 있으나 경기동행종합지수를 기준으로 순환하는 모습을 보여 데이터의 안정성이 확보될 시 유용성이 있을 것으로 보인다.





## 제4장

### 뉴스 빅데이터 지역경기지수 유용성 검증

1. 경기지수의 순환변동치 작성 및 유용성 점검
2. 주요 경기국면에서 전환점 분석





## 제4장 뉴스 빅데이터 지역경기지수 유용성 검토

### 1. 경기지수의 순환변동치 작성 및 유용성 점검

#### 가. 경기순환 판단을 위한 보조지표로서의 순환변동치 작성

##### 1) 순환변동치 작성 방법

경기의 상승, 하강 변화를 판단하기 위해서 활용되는 지표가 순환변동치이며 경기지수 만큼 중요하게 판단되는 보조지표이다. 경기지수는 경제의 성장에 따라 증가하는 경향을 보이며 이는 경기의 상승과 하강을 동시에 포함하고 있어 경기국면 판단에 한계가 있다. 이러한 한계를 극복하기 위해 보조적 판단지표가 필요하여 순환변동치를 작성하여 경기 판단의 자료로 활용하고 있다.

순환변동치는 다음과 같은 방법으로 구하며,  $T_t$ 는 H-P 필터 방법으로 산출한다.

$$C_t = \frac{CI_t}{T_t} \times 100$$

여기서,  $T_t$ 는 경기지수 t 기간의 중심항 이동평균치

##### 2) 순환변동치 작성 결과

Composite Index 방식을 통해 산출된 “전북뉴스 지역경기종합지수” 순환변동치는 2021년 12월 102.1로 전월비 1.0pt 하락한 것을 볼 수 있으며 이 시점을 기준으로 증가 후 감소세를 보이고 있다.

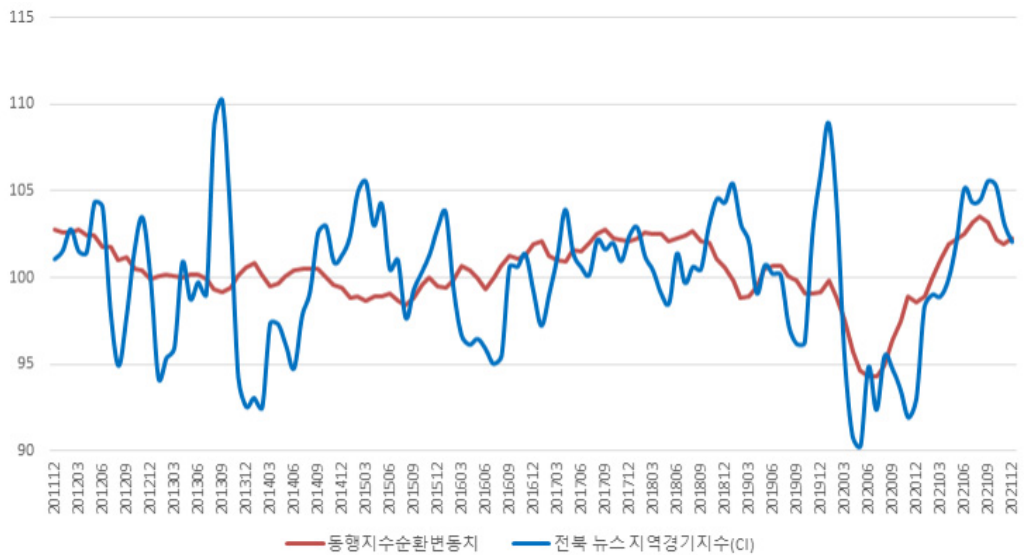
[표 4-1] 전북뉴스 지역경기종합지수 순환변동치 산출 결과

| 구분                  | 2011.12 | 2012.1 | 2012.2 | 2012.3 | ... | 2021.9 | 2021.10 | 2021.11 | 2021.12 |
|---------------------|---------|--------|--------|--------|-----|--------|---------|---------|---------|
| 전북뉴스 지역경기 종합지수 (CI) | 101.1   | 101.6  | 102.8  | 101.5  | ... | 105.6  | 105.3   | 103.1   | 102.1   |

## 나. 유용성 검증

전북 경기동행종합지수 순환변동치와 비교해 볼 때 양 지수가 근접하게 움직이는 것을 볼 수 있다. 다만, 2013년 7월에서 2014년 1월까지의 이상치와 유사한 형태의 진폭, 2019년 10월에서 2020년 5월은 설명하기 어려울 만큼 변동폭이 큰 것을 볼 수 있다. 다른 한편으로는 코로나가 발생하기 시작한 2020년 3월은 뉴스경기종합지수가 더 현실감 있게 반영된 것을 볼 수 있으며 회복하는 과정에서도 진폭이 크다는 특징이 있으나 전북 경기동행종합지수와 유사한 패턴으로 회복되는 것을 볼 수 있어 제한적으로는 설명력을 가지는 것으로 보인다.

전반적으로 지수의 변화가 민감한 것을 볼 수 있으며 BSI로 작성한 경기지수의 변화폭이 큰 이유와 마찬가지로 데이터의 특성과 시계열의 부족에 기인한 것으로 보인다.



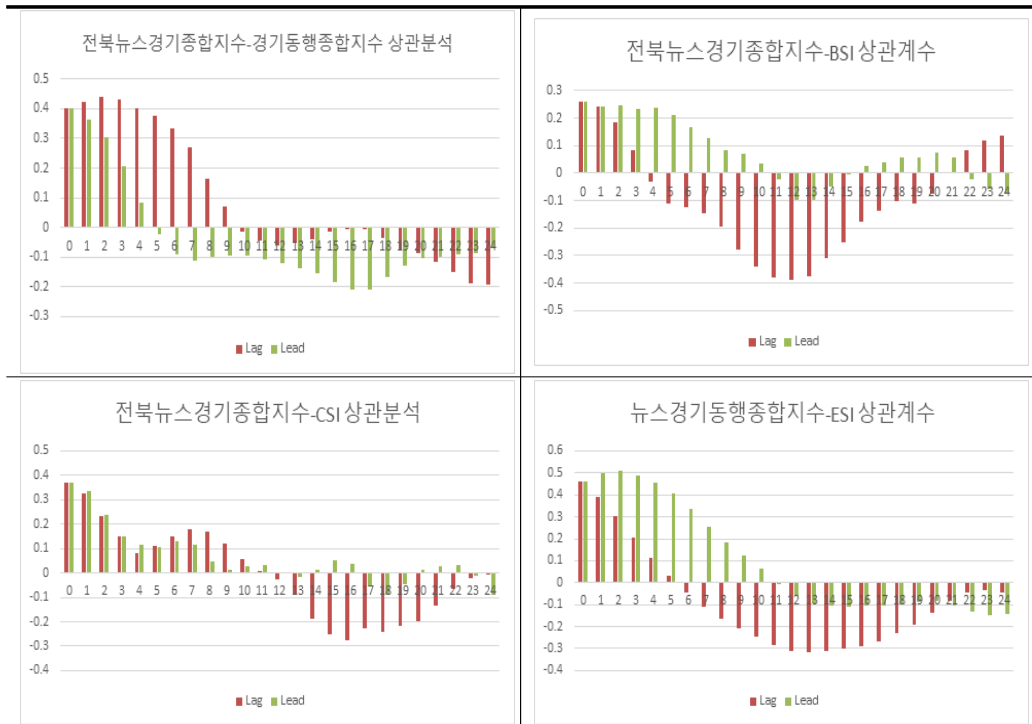
[그림 4-1] 전북뉴스 지역경기종합지수와 동행지수 순환변동치

전북뉴스 지역경기종합지수(CI)와 전라북도 경기종합동행지수, 전라북도 기업경기실사지수(BSI), 전라북도 소비심리지수(CSI) 그리고 전국 경제심리지수(ESI)에 대한 상관성을 분석하면, 전라북도 경기종합동행지수와 상관성이 ESI 다음으로 높게 나타났다. 그러나 상관계수 값은 0.40으로 낮은 상관성을 보였다

[표 4-2] 전북뉴스 지역경기종합지수(CI)와 각 지수와의 상관관계

| 구분         | 동행지수 | BSI  | CSI   | ESI  |
|------------|------|------|-------|------|
| 뉴스         | 0.40 | 0.26 | 0.37  | 0.46 |
| 지역경기지수(CI) | (-2) | (0)  | (-12) | (+2) |

한편 전북뉴스 지역경기종합지수 순환변동치는 경기동행지수 순환변동치와도 밀접하게 움직이며 2개월 선행하는 것으로 나타났다. 다음 그림은 상관관계의 시차분석을 표시한 것이다. 그림에서 보는 바와 같이 전북뉴스 지역경기종합지수와 전북 경기동행종합지수와는 2개월 lag에서 가장 높은 상관관계를 보이고 있어 2개월의 선행성을 가지는 것으로 나타났다. 반면 ESI와는 2개월 후행하는 것으로 나타났고 BSI와는 동행성을 가지는 것으로 나타났다. CSI는 12개월의 선행성을 가지는 것으로 나타났으나 BSI와 역관계를 가지는 것으로 볼 수 없으므로 해석은 동행성을 가지는 것으로 판단하는 것이 적절할 것으로 보인다.



[그림 4-2] 뉴스경기종합지수와의 상관계수 분석

---

순환변동치의 작성과 상관관계 분석을 통해서 본 결과 뉴스경기종합지수 순환변동치와 경기동행종합지수 순환변동치가 유사한 형태로 움직이고 있으나 변동성이 큰 관계로 상관도가 낮게 나타났다. 향후에 변동성 조정에 대한 추가 연구가 필요할 것으로 보인다.

뉴스경기종합지수가 경기동행종합지수에 비해 선행성을 가지는 것은 향후 데이터가 정제되고 자료가 충분히 확보된다면 선행지표로서의 가능성을 가지게 되어 활용가치가 높을 것으로 보인다.



## 2. 주요 경기국면에서 전환점 분석

### 가. 전환점 분석 방법

경기순환에서 전환점 분석은 국면을 파악하고 예측하는데 중요한 관심사이다. 전북뉴스 지역경기종합지수의 전환점 분석을 위해 Harding and Pagan (2002)의 근사치를 Bry-Boschan(B.B.) 알고리즘(Bry & Boschan, 1971)에 적용하여 분석하였다.

BB 알고리즘은 NBER에서 전통적으로 사용하는 방법으로 순환주기를 자동화하는 방법이다. 자동화 하는 방법은 정점과 저점을 분류하기 위해 계열의 움직임에 대한 규칙을 정하고 이에 대한 판단으로 정점과 저점을 찾아내는 것이다. 경기후퇴는 최고점에서 최저점으로 발생하고 경기확장은 최저점에서 최고점으로 발생한다는 원칙하에 먼저 local max와 local min을 범위 설정(parameter k)하고 다음으로 경기순환주기의 최소기간을 선정한다(parameter p). 이후 순환사이클의 최소 지속기간을(parameter c) 선정하는 방식을 활용하여 전환점 분석을 한다. 일반적으로 분기는 5분기, 월 데이터는 15개월의 지속기간을 선택한다.

### 나. 전환점 분석 결과

전환점 분석 결과 전북 경기동행종합지수의 전환점이 가장 적게 나타났고, BSI는 경제 순환주기가 가장 짧았다. 뉴스 경기지수가 2013년 9월을 저점으로 2014년 2월 정점, 2015년 3월 저점, 2016년 7월 정점, 2019년 1월 저점, 2020년 5월 정점을 보였다. 전북 경기동행종합지수와 비교할 때 2014년 2월의 정점은 7개월 선행, 2015년 5월의 저점은 5개월 선행하는 것을 볼 수 있었으나 이후 순환주기에서는 설명력이 약한 것을 보였다.

[표 4-3] NBER Bry-Bochan 방법에 의한 Turning Point 산출 결과

| 구분             | Trough | Peak   | Trough | Peak   | Trough | Peak   | Trough |
|----------------|--------|--------|--------|--------|--------|--------|--------|
| BSI            | 2013.3 | 2014.4 | 2016.3 | 2017.6 | 2018.8 | 2019.4 | 2020.7 |
| 뉴스경기지수         | 2013.9 | 2014.2 | 2015.3 | 2016.7 | 2019.1 | 2020.5 |        |
| 전북경기동행<br>종합지수 | 2013.9 | 2014.9 | 2015.8 | 2018.2 | 2020.7 |        |        |

---

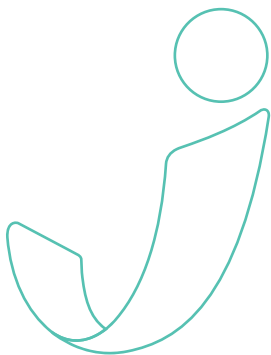
전환점의 설명력 제고를 위해서 다양한 분석을 수행하였으나 설명력이 개선되지는 않았다. 결국 데이터의 한계를 극복하는 것만이 설명력을 제고할 수 있을 것으로 보인다.



# 제 5 장

## 결론 및 향후과제

1. 결론
2. 향후과제





## 제 5 장 결론 및 향후과제

### 1. 결론

전북뉴스 지역경기종합지수 작성을 위해 새전북신문, 전북도민일보, 전북일보, 전북중앙신문, JTV전주방송, 전주MBC의 기사 데이터를 구축하였다. 기사는 파이선을 활용한 뉴스기사 스크래핑 프로그램을 자체적으로 만들어 수집했다.

경기종합지수의 작성은 BSI 방식과 Composite Index 방식의 2가지를 사용하였다. 전북뉴스 지역경기지수 BSI 작성방법은 이근희(2017)의 연구에서 사용된 것을 적용한다. 즉, 호황지표와 불황지표의 차를 호황지표와 불황지표의 합으로 나누어 BSI 방식을 적용했다.

수집된 뉴스데이터의 활용가능성을 검토하기 위해 전북 동행지수 순환변동치와 전국 경제심리지수, 전북 기업경기실사지수와 전북 소비자동향지수와의 상관도를 각각 분석했다. 분석을 통해 상관계수의 절대값이 0.4 이상인 데이터를 1차적으로 선정하여 추가분석을 했다. 동행지수는 7개의 단어, 기업경기실사지수는 2개의 단어, 소비자동향지수는 4개, 경제심리지수는 9개를 활용할 수 있었다. 구체적으로 불황 단어는 경기침체, 위기, 금융위기, 경제위기, 해고 5개이고, 호황 단어는 거품 1개이며, 마지막으로 경제 단어는 주가, 한국은행, 신용등급 3개로 나타났다. 최적 단어의 선택은 전국 경제심리지수가 0.4 이상이고 전북 동행지수 순환변동치가 0.4 이상인 경우로 한정했다. 최종적으로 선택하는 단어는 불황 단어를 중심으로 하되, 호황 단어와 연계하여 선택한다. 기존 지표와 상관도가 높고 경기변화를 잘 반영하는 불황과 호황의 상관도 차이가 큰 지표를 선정했다. 최종 선정 단어는 “경기침체”, “해고”, “금리인상”, “디플레이션” 4개의 단어쌍으로 총 8개로 구성했다.

Composite Index 방식의 전북뉴스 지역경기종합지수 작성과정은 통계청의 지역경기종합지수 실무가이드를 따라서 작성했다. 지수작성을 위해서 구성지표의 자료를 수집하고 데이터의 원계열로 활용할 것인지와 계절조절이 필요한 지에 대한 판단을 한다. 비경기적 요인을 제거하기 위한 계절요인은 X-12 ARIMA를 활용하고 불규칙 요인은 이동평

---

균법을 활용한다. 이후 개별지표의 Cycle을 검정하여 시계열 사이클과 경기대응력을 검토한다. CI 작성은 구성지표의 계절 및 불규칙조정 → 전월비 증감률 산출 → 증감률 진폭의 표준화 → 종합증감률 산출 → 추세조정지수 산출 → 경기종합지수 산출 → 동행지수 순환변동치 산출 등의 과정을 통해 결과를 도출했다.

전북의 뉴스경기종합지수는 2021년 12월 101.1로 전월비 0.9pt 하락한 것을 볼 수 있다. 전북 경기동행종합지수와 비교해 볼 때 지수의 변화가 민감한 것을 볼 수 있으며 이는 데이터의 특성과 시계열의 부족에 기인한 것으로 보인다. 한편, BSI 방식과 Composite Index를 비교하면 후자가 더 안정적인 계열임을 볼 수 있다. BSI 방식의 경우 마지막 4개월의 편차 뿐 아니라 최초의 4개월 또한 편차가 상대적으로 큰 것을 볼 수 있었다.

뉴스경기종합지수의 유용성을 판단하기 위해 순환변동치를 구하고 전환점 분석을 하였다. 전북 경기동행종합지수 순환변동치와 비교해 볼 때 양 지수가 근접하게 움직이는 것을 볼 수 있다. 다만, 2013년 7월에서 2014년 1월까지의 이상치와 유사한 형태의 진폭, 2019년 10월에서 2020년 5월은 설명하기 어려울 만큼 변동폭이 큰 것을 볼 수 있다. 다른 한편으로는 코로나가 발생하기 시작한 2020년 3월은 뉴스경기종합지수가 더 현실감 있게 반영된 것으로 볼 수 있으며 회복하는 과정에서도 진폭이 크다는 특징이 있으나 전북 경기동행종합지수와 유사한 패턴으로 회복되는 것을 볼 수 있어 제한적으로는 설명력을 가지는 것으로 보인다. 전환점 분석결과 전북 경기동행종합지수의 전환점이 가장 적게 나타났고, BSI는 경제순환주기가 가장 짧았다. 뉴스 경기지수가 2013년 9월을 저점으로 2014년 2월 정점, 2015년 3월 저점, 2016년 7월 정점, 2019년 1월 저점, 2020년 5월 정점을 보였다. 전북 경기동행종합지수와 비교할 때 2014년 2월의 정점은 7개월 선행, 2015년 5월의 저점은 5개월 선행하는 것을 볼 수 있었으나 이후 순환주기에서는 설명력이 약한 것으로 나타났다.

BSI 모델과 Composite Index 모델을 통해서 알아본 뉴스빅데이터를 활용한 지역경기지수 개발은 소기의 성과를 거두었다. 순환변동치의 국면전환의 유사성을 찾을 수 있었으며 뉴스경기종합지수의 선행성은 향후 지역경기동향 파악의 중요한 자료로 활용될 수 있을 것이다. 통상적으로 경기지수 개발 후 6개월에서 1년 정도는 모델의 안정성과 경기대응력을 판단하기 위한 모니터링 기간을 갖게 되므로 지속적인 업데이트를 통해 시계열 확보와 데이터 정제를 통한 모델의 고도화가 필요하다.

---

## 2. 향후과제

### 가. 뉴스 데이터 구축 체계화

뉴스 기사 데이터는 연구 수행에 있어서 가장 핵심이 되는데, 데이터 구축에 있어서 두 가지의 어려움이 존재하였다. 첫째는 데이터 구축을 위한 스크랩 프로그래밍과 장기 시계열 데이터 구축이며, 둘째는 목적과 적합한 내용의 기사를 스크랩하는 것이다. 이를 구체적으로 설명하면 다음과 같다.

첫째, 본 연구에서 사용한 데이터는 네이버 뉴스에서 제공하는 지역 신문 기사를 스크랩하여 구축하였다. 네이버를 이용한 데이터 구축 시점은 2011년 12월부터인데 ‘전북도 민일보’, ‘새전북신문’ 등 지역신문 대다수는 2000년부터 기사를 제공하고 있다. 장기 시계열 데이터 확보 측면에서 지역신문을 스크랩하는 것이 맞겠지만, 신문사마다 홈페이지 구성이 상이하여 기사 스크랩을 위한 프로그램을 새롭게 구축해야 하는 어려움과 신문사마다 키워드별 스크랩을 시도하는 것이 기초연구 단계에서 시간상 어려움이 존재하였다<sup>12)</sup>.

둘째, 구축 데이터의 적합성이다. 연구수행을 위해서 네이버에서 검색된 기사와 지역신문 홈페이지를 방문하여 검색된 기사를 비교하였다. 검색 키워드로 도출된 기사 내용이 네이버와 지역신문이 미세하게 상이한 부분이 확인되었는데, 이 부분은 광고 노출 빈도의 차이인 것으로 확인하였다<sup>13)</sup>. 광고 뉴스를 제외하는 것은 기술적으로는 가능하지만, 완벽하게 제외한다는 것은 높은 기술력과 시간이 필요하다. 이에 대한 추가적인 연구가 필요하다<sup>14)</sup>.

- 
- 12) 연구자들은 스크랩코드를 작성하기 위하여 시도를 하였지만, 일부 홈페이지에서는 검색기간 및 검색 키워드 기사뿐만 아니라 최근기사까지도 같이 노출되는 경우가 발생하여 초기단계에서의 프로그램 코드 구현이 어려웠다.
  - 13) 네이버 뉴스기사에서는 지역신문에서 검색되는 광고가 일정부분 배제되는 것을 확인할 수 있었다. 그렇다고 네이버에서 제공하는 기사가 광고가 없다는 것은 아니며, 과제에 투입되는 시간을 효율적으로 운영하기 위하여 광고가 많이 배제된 네이버를 선택하였다는 것을 의미한다.
  - 14) 스크랩된 뉴스기사가 경제뉴스인지 아닌지 판별하는 것은 사람이 체크할 수 있지만, 인공지능 프로그램을 활용하는 것도 하나의 방법일 수 있다.

---

## 나. 경기지표 개발 및 빅데이터 활용도 제고

뉴스경기지수 개발에서 선행지수의 작성 필요성이 더 절실한 상황이다. 모델의 설명력 제고 및 유용성을 측정하기 위해서 경기동행종합지수 하나로는 reference가 되는 지수가 부족하기 때문에 선행종합지수의 작성과 관리를 통해 보조지표의 활용성이 제고될 수 있을 것으로 판단된다.

또한 뉴스경기동행종합지수의 발간주기를 주 단위로 조절하여 속보성이 있는 서비스를 제공함으로써 지역 경제주체 및 정책입안자들이 지역경제 경쟁력 향상을 위한 효율적 의사결정에 도움이 될 수 있도록 할 필요가 있다.

지역의 빅데이터를 정책에 활용할 수 있도록 다양한 방안이 모색되어야 할 것으로 보인다. 정부에서 민간과 공공 빅데이터를 관리하고 있으나 활용도는 여전히 떨어지고 있어 이를 적극적으로 활용하기 위한 노력이 필요하다.



## 참 고 문 헌

### REFERENCE

- 김유신, 김남규 & 정승렬. (2012). 뉴스와 주가 : 빅데이터 감성분석을 통한 지능형 투자의사결정 모형. *지능정보연구*, 18(2), 143-156.
- 김지은. (2013). 경기지표로서 인터넷 검색지표의 유용성 분석. *한국은행*.
- 나현주 & 최정재. (2016). 경제 뉴스에 따른 금융시장 반응 분석 - 주식 및 채권선물 시장을 중심으로. *한국은행*.
- 박대민. (2016). 장기 시계열 내용 분석을 위한 뉴스 빅데이터 분석의 활용 가능성 : 100만 건 기사의 정보원과 주제로 본 신문 26년. *한국언론학보*, 60(5), 353-407.
- 송치영. (2002). 뉴스가 금융시장에 미치는 영향에 관한 연구. *국제경제연구*, 8(3), 1-34.
- 안희준, 최종범 & 전승표. (2010). 남북관계 관련 뉴스가 주식시장에 미치는 영향. *한국경제의 분석패널*, 16(2), 199-238.
- 이근영. (2006). 북한 핵관련 뉴스가 국내주식 및 외환시장에 미치는 영향. *동북아경제연구*, 18(1), 61-90.
- 이금희. (2017). 뉴스 빅데이터를 이용한 경기판단: 빅카인즈 뉴스 경기지수의 개발. *통계연구*, 22(2), 67-94.
- 이금희 & 황상필. (2014). 빅데이터를 이용한 경기판단지표 개발: 네이버 검색 경기지수 작성과 유용성 검토. *한국은행 경제연구원 「經濟分析」*, 20(4).
- 이완수, 심재철 & 박양수. (2007). 경제뉴스, 경제상황, 소비자 기대심리 그리고 소비행위의 상호 속성 의제설정 관계에 대한 시계열 분석. *한국언론학보*, 51(4), 280-307.
- 이완수 & 노성중. (2008). '무엇'에서 '언제'로 : 벡터자기회귀모형을 통한 경제현실, 경제보도, 경제인식 간 상호영향의 시간차 탐구. *한국언론학보*, 52(5), 320-345.
- 이완수 & 박양수. (2016). 경제 정보에 대한 비대칭적 반응. *한국언론학보*, 60(1), 165-201.
- 이완수. (2007). 한국 경제뉴스의 속성(attributes) 프레임 효과 연구. *언론과 사회*, 15(1), 86-112.
- 이완수 & 배정근. (2013). 국내 경제저널리즘의 현황과 품질제고 방안 연구 : 경제뉴스 전문매체의 형식과 내용구성 체계. *한국언론진흥재단*.
- 최윤희. (2016). 진화하는 빅카인즈, 데이터저널리즘 플랫폼으로 기여. *한국조사기자협회, 조사연구 제28권*, 89-97.

- Bry, G., & Boschan, C. (1971). Programmed selection of cyclical turning points. In *Cyclical analysis of time series: Selected procedures and computer programs* (pp. 7-63). NBER.
- Castle, J. L., Fawcett, N. W., & Hendry, D. F. (2009). Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review*, 210, 71-89.
- Chamberlin, G. (2010). Googling the present. *Economic & Labour Market Review*, 4(12), 59-95.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (No. w17699). National Bureau of Economic Research.
- Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic record*, 88, 2-9.
- Einav, L., & Levin, J. (2013). The Data Revolution and Economic Analysis. Technical Report. NBER Innovation Policy and the Economy Conference. NBER Working Paper 19035.
- Einav, L., Knoepfle, D., Levin, J., & Sundaresan, N. (2013). Sales Taxes and Internet Commerce. NBER Working Paper 18018.
- Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11), 87-92.
- Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., ... & Oregon Health Study Group. (2012). The Oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, 127(3), 1057-1106.
- Francesco, D. A. (2009). Predicting unemployment in short samples with internet job search query data. Bank of Italy Research Department.
- Harding, D., & Pagan, A. (2002). Dissecting the cycle: a methodological investigation. *Journal of monetary economics*, 49(2), 365-381.
- Li, J., Xu, Z., Yu, L., & Tang, L. (2016). Forecasting oil price trends with sentiment of online news articles. *Procedia Computer Science*, 91, 1081-1087.
- McLaren. (2011). Using internet Search Data as Economic Indicators. Bank of England Quarterly Bulletin, 51(2), 134-140.
- Pelat, C., Turbelin, C., Bar-Hen, A., Flahault, A., & Valleron, A. J. (2009). More

diseases tracked by using Google Trends. *Emerging infectious diseases*, 15(8), 1327.

Suhoy. (2009). Query Indices and a 2008 downturn: Israeli Data. Bank of Israel Discussion Paper.

Tobback, E., Naudts, H., Daelemans, W., de Fortuny, E. J., & Martens, D. (2018). Belgian economic policy uncertainty index: Improvement through text mining. *International journal of forecasting*, 34(2), 355-365.

Yao, T., & Zhang, Y. J. (2017). Forecasting crude oil prices with the Google index. *Energy Procedia*, 105, 3772-3776.

Zimmermann. (2009). Google Econometrics and Unemployment Forecasting. *Applied economic Quarterly*, 55(2), 107-120.

## SUMMARY

---

### A Study on Developing Jeonbuk Economic Indicators by Utilizing News Big Data

Kangjin Lee · Sunghwan Choi · Minkyung Kim

Recently, with the growing interest in big data, media data produced daily are comprehensively written in various fields of society, and the amount is very large and the data cycle is very fast, so it can be used to understand the current state of the economy.

The function of internet portal was strengthened so that the data of each newspaper and broadcasting company could be accessed on the Internet platform, making it easy to provide data service and building a database for analysis, making it possible to develop a regional economic index using news big data.

In Jeonbuk, it is also required to create an economic index to be used as a supplement to improve the utilization of the Jeonbuk Coincidental Index.

This study area is an unexplored field in the region and is the first attempt by the local government level. The scope of the study was selected through a review of domestic and foreign research cases, and the necessary data were obtained by directly creating a web crawling program using Python and accessing local media outlets. To create the News Business Composite Index, the BSI method and the Composite Index method using the KOSTAT-CIS

program provided by the National Statistical Office were used to compare the two results. To verify usefulness of the index, various quantitative analyzes such as correlation analysis and turning point analysis are performed.

To create the Jeonbuk News Gyeonggi Comprehensive Index, article data from the New Jeonbuk Newspaper, Jeonbuk Domin Ilbo, Jeonbuk Ilbo, Jeonbuk Joongang Newspaper, JTV Jeonju Broadcasting, and Jeonju MBC were collected. Articles were collected by making a news article scraping program using Python.

Both the BSI method and the Composite Index method were used to compare and test the usability.

The Jeonbuk News Coincidence Index (BSI method), as in the study of Lee Geung-hee (2017), was applied by dividing the difference between the boom and recession indicators by the sum of the boom and recession indicators. The process of creating the Composite Index-based Jeonbuk News Economic Composite Index was prepared in accordance with the practical guide of the Regional Economic Composite Index of the National Statistical Office.

CI creation is derived through the process of seasonal and irregular adjustment of component indicators → calculation of change rate compared to the previous month → standardization of amplitude of change rate → calculation of overall change rate → calculation of trend adjustment index → calculation of composite economic index → calculation of cyclical fluctuation of coincident index.

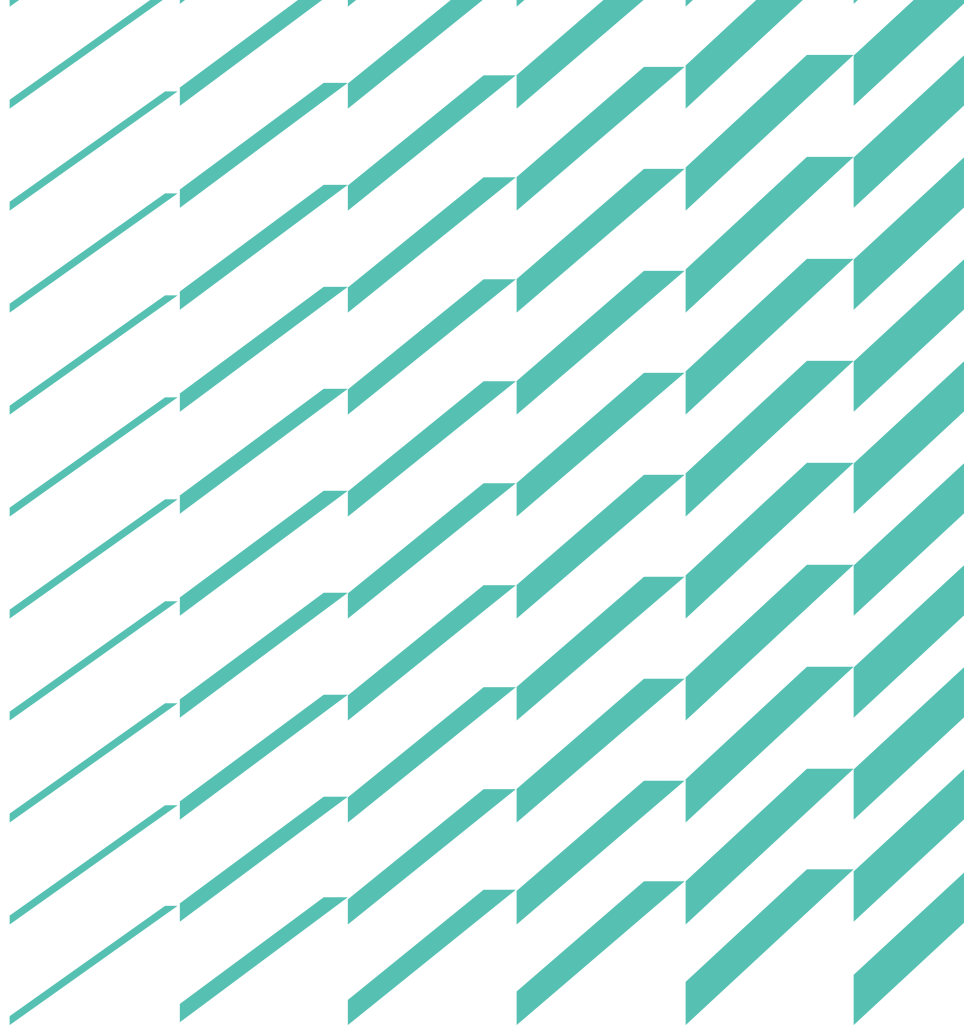
In order to determine the usefulness of the News Business Composite Index, the cyclical fluctuations were obtained and turning point analysis was conducted. When compared with the cyclical fluctuation of the Jeollabuk-do coincidental Index, it can be seen that both indices move closer. However,

from July 2013 to January 2014, the amplitude similar to the outlier, and from October 2019 to May 2010, it can be seen that the fluctuation range is so large that it is difficult to explain. As a result of the turning point analysis, the Jeonbuk Coincidental Index showed the fewest turning points, and the BSI had the shortest economic cycle. Compared to the Jeonbuk Coincidental Index, the peak in February 2014 was 7 months ahead and the trough in May 2015 was 5 months ahead.

The development of the regional agricultural index using news big data, which was investigated through the BSI model and the Composite Index model, achieved the expected results. The similarity of the phase transition of the cyclical variables was found, and the precedence of the news business composite index could be used as important data for identifying regional economic trends in the future. Typically, 6 months to 1 year after the development of the economic index is a testing and monitoring period to determine the stability and responsiveness of the model, so it is necessary to secure time series through continuous updates and seek ways to stabilize the data to advance the model.

**Key Words**

Big data, News big data, BSI, Coincidental index, Python, Turning point, Cyclical fluctuation



## 부록

### 데이터 계열

- 전라북도 경제동행종합지수 순환변동치
- 전라북도 뉴스경기종합지수 순환변동치
- 호황, 불황, 경제관련 용어







## 부 록

### APPENDIX

#### ■ 데이터 계열

##### ○ 전라북도 경기동행종합지수 순환변동치

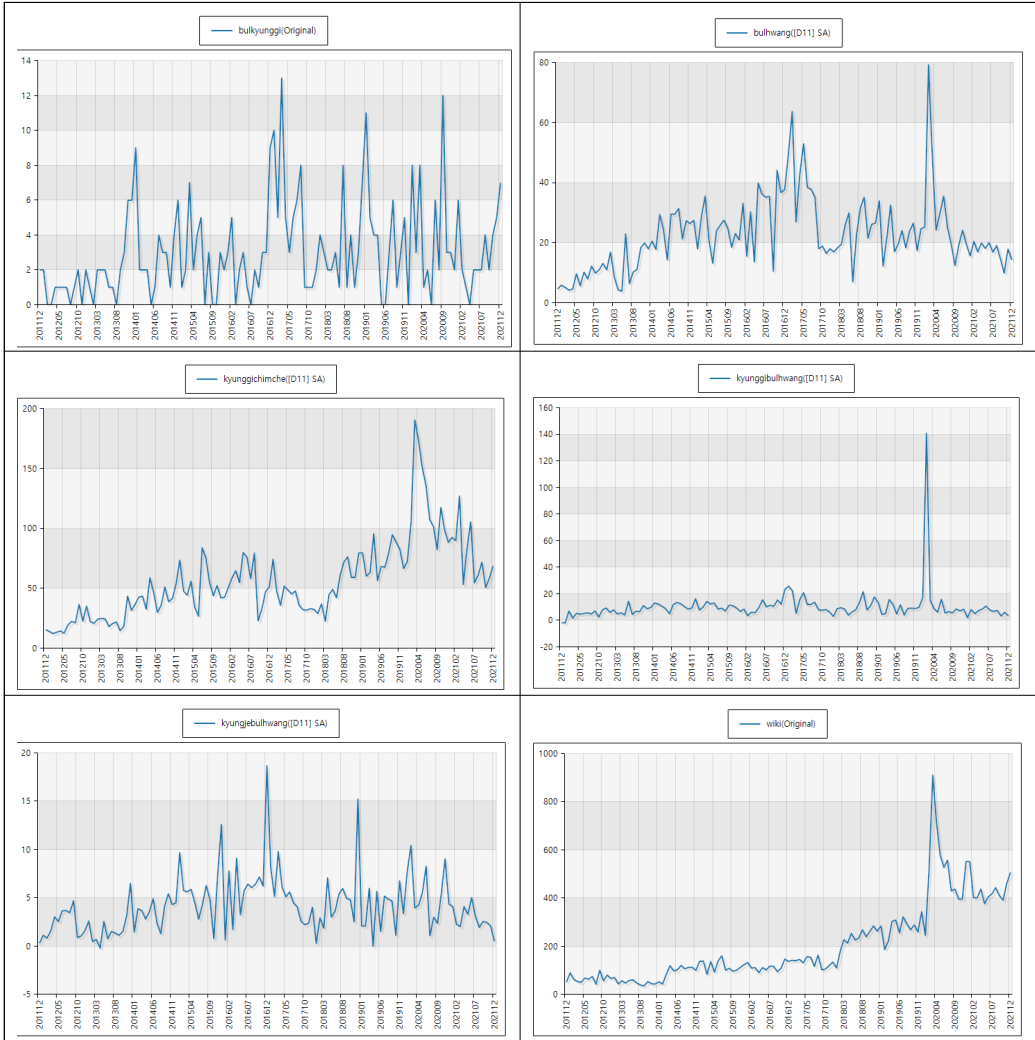
| 구분   | 1월    | 2월    | 3월    | 4월    | 5월    | 6월    | 7월    | 8월    | 9월    | 10월   | 11월   | 12월   |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2011 |       |       |       |       |       |       |       |       |       |       |       | 102.8 |
| 2012 | 102.6 | 102.6 | 102.8 | 102.4 | 102.4 | 101.8 | 101.8 | 101.0 | 101.2 | 100.5 | 100.4 | 99.9  |
| 2013 | 100.1 | 100.2 | 100.1 | 100.0 | 100.2 | 100.2 | 99.9  | 99.3  | 99.2  | 99.4  | 100.1 | 100.6 |
| 2014 | 100.8 | 100.2 | 99.5  | 99.7  | 100.1 | 100.4 | 100.5 | 100.5 | 100.5 | 100.0 | 99.6  | 99.4  |
| 2015 | 98.8  | 98.9  | 98.7  | 98.9  | 98.9  | 99.1  | 98.7  | 98.4  | 98.8  | 99.6  | 100.0 | 99.5  |
| 2016 | 99.4  | 99.8  | 100.7 | 100.4 | 100.0 | 99.3  | 100.0 | 100.7 | 101.3 | 101.1 | 101.3 | 101.9 |
| 2017 | 102.1 | 101.3 | 101.0 | 100.9 | 101.6 | 101.5 | 102.0 | 102.5 | 102.8 | 102.3 | 102.2 | 102.1 |
| 2018 | 102.3 | 102.6 | 102.5 | 102.5 | 102.1 | 102.3 | 102.4 | 102.7 | 102.1 | 102.0 | 101.1 | 100.6 |
| 2019 | 99.8  | 98.8  | 98.9  | 99.5  | 100.5 | 100.7 | 100.7 | 100.1 | 99.8  | 99.1  | 99.1  | 99.2  |
| 2020 | 99.8  | 98.9  | 97.6  | 95.7  | 94.6  | 94.3  | 94.3  | 94.9  | 96.4  | 97.5  | 98.9  | 98.6  |
| 2021 | 98.9  | 99.9  | 101.0 | 101.9 | 102.2 | 102.5 | 103.2 | 103.5 | 103.2 | 102.2 | 101.9 | 102.3 |

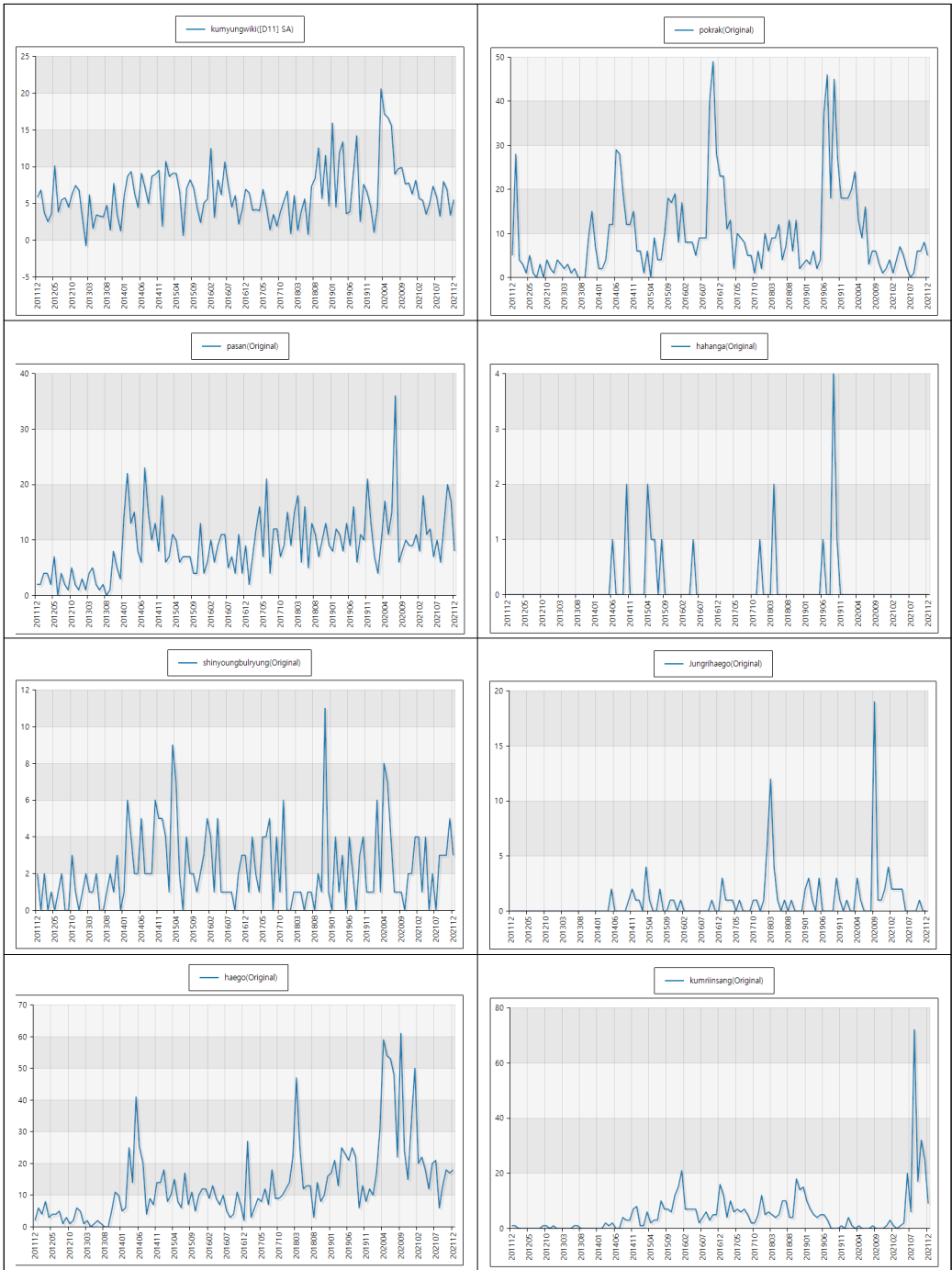
자료: 전라북도, 전라북도 경기종합지수

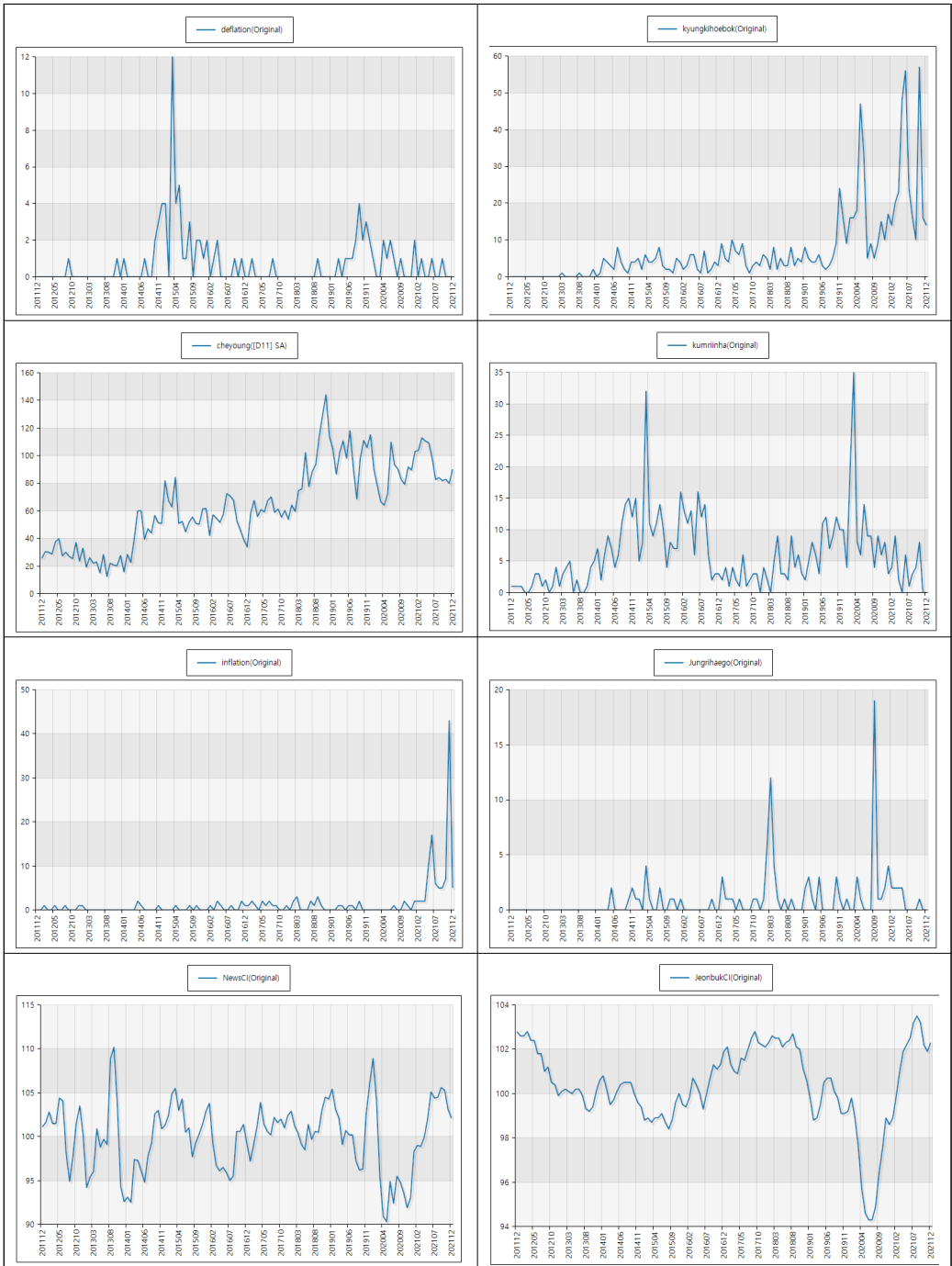
##### ○ 전라북도 뉴스경기종합지수 순환변동치

| 구분   | 1월    | 2월    | 3월    | 4월    | 5월    | 6월    | 7월    | 8월    | 9월    | 10월   | 11월   | 12월   |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2011 |       |       |       |       |       |       |       |       |       |       |       | 101.1 |
| 2012 | 101.6 | 102.8 | 101.5 | 101.5 | 104.4 | 104.1 | 98.1  | 94.9  | 97.8  | 101.6 | 103.5 | 100.0 |
| 2013 | 94.2  | 95.4  | 96.0  | 100.9 | 98.8  | 99.7  | 99.1  | 108.9 | 110.2 | 103.7 | 94.3  | 92.6  |
| 2014 | 93.1  | 92.5  | 97.4  | 97.3  | 96.1  | 94.8  | 97.8  | 99.2  | 102.6 | 103.0 | 100.9 | 101.3 |
| 2015 | 102.4 | 104.9 | 105.5 | 103.0 | 104.3 | 100.5 | 101.0 | 97.7  | 99.3  | 100.3 | 101.4 | 102.9 |
| 2016 | 103.8 | 99.3  | 96.7  | 96.1  | 96.5  | 95.9  | 95.0  | 95.5  | 100.6 | 100.6 | 101.4 | 99.2  |
| 2017 | 97.2  | 99.1  | 101.2 | 103.9 | 101.4 | 100.6 | 100.2 | 102.2 | 101.6 | 102.0 | 101.0 | 102.4 |
| 2018 | 102.9 | 101.2 | 100.4 | 99.1  | 98.5  | 101.4 | 99.7  | 100.6 | 100.5 | 103.0 | 104.5 | 104.3 |
| 2019 | 105.4 | 103.1 | 102.1 | 99.1  | 100.7 | 100.2 | 100.2 | 97.2  | 96.2  | 96.3  | 102.6 | 106.0 |
| 2020 | 108.9 | 104.2 | 95.5  | 90.9  | 90.3  | 94.9  | 92.4  | 95.5  | 94.8  | 93.6  | 91.9  | 93.0  |
| 2021 | 98.3  | 99.0  | 98.9  | 99.9  | 102.0 | 105.1 | 104.4 | 104.5 | 105.6 | 105.3 | 103.1 | 102.1 |

## ■ 데이터 계열







기본연구 2022-09

**뉴스 빅데이터를 활용한 전북지역경기지수 개발 연구**

---

발행인 | 권혁남

발행일 | 2022년 12월 31일

발행처 | 전북연구원

55068 전북 전주시 완산구 콩쥐팥쥐로 1696

전화: (063)280-7100 팩스: (063)286-9206

---

ISBN 978-89-6612-457-2 93520 (PDF)

본 출판물의 판권은 전북연구원에 속합니다.



## 2022년도 주요 연구과제

### 기본연구

전라북도 신산업 생태계 및 산업기술지도 분석  
뉴스 빅데이터를 활용한 전북지역경기지수개발 연구  
지속가능한 농촌 발전을 위한 중간지원조직의 역할과 과제  
전라북도 광역생태축 구축을 위한 정책방향 연구  
도시 회복력 관점의 전라북도 주거지 특성 분석 연구  
2022 전북형 행복지표 실태조사  
전북형 젠더거버넌스 기본 구상 및 활성화 방안  
전북 농산어촌 활성화를 위한 크리에이터 육성 방향  
전라북도 스마트관광 환경 조성을 위한 기초연구

### 기획연구

생태관광 활성화를 위한 장거리 생태·문화탐방로 조성방안  
전북의 대중국 초광역협력 방안 연구  
전라북도 주력산업 경쟁력 진단과 중장기 발전방향 : 신재생에너지산업  
새만금 대규모 테마파크 유치 구상 연구  
전라북도의 대기업 유치 전략 수립을 위한 탐색 연구

### 정책연구

전라북도 항만물류서비스 특성화 방안 연구  
인구감소지역 대응전략 및 지원방안 마련  
전라북도 노후 공동주택 관리 방안  
전라북도 여건에 맞는 수계기금 지원 운영 개선방안 연구  
전라북도 도시계획 심의 가이드라인 연구  
전북 서남권 해상풍력 지역기업 참여방안 연구  
전라북도 인삼산업 육성정책 기본방안  
전라북도 바이오화학산업 활성화 방안  
전라북도 산업단지 진흥 및 구조도화 계획 수립  
전라북도 관광특구 활성화 방안 연구  
전라북도 전기차충전소 보급 및 관리방안 연구  
전라북도 자연환경 훼손지역 녹색복원 추진방안 연구  
전라북도 지역산업맞춤형 일자리창출 지원사업 개선방안  
전라북도 마을공동체 자립 역량지표 개발 및 활용방안  
지방도매시장 농산물 유통혁신 전략화 방안 연구  
전북여성인재 육성 및 활용방안 연구 : 과학기술분야를 중심으로  
전라북도 문화예술교육 지원방안 연구  
전라북도 먹거리 종합계획 추진전략 연구  
전라북도 중소벤처 스마트혁신지구 조성 방안  
전북형 지역사회 통합돌봄 지원체계 구축 방안 연구

 **전북연구원**

55068 전라북도 전주시 완산구 공취말쭈로 1696

Tel 063. 280. 7100

Fax 063. 286. 9206

[www.jthink.kr](http://www.jthink.kr)

