

# 해양수산업 전망모형 개발 연구(Ⅲ)

A Study on the Development  
of Marine and Fisheries Outlook Model (3)

권장한 · 장정인 · 정수빈 · 김정우



한국해양수산개발원  
KOREA MARITIME INSTITUTE

<b>저자</b>	<b>권장한, 장정인, 정수빈, 김정우</b>
<b>내부연구진</b>	연구책임자 권장한 한국해양수산개발원 경제전망·데이터연구실 전문연구원 공동연구원 장정인 한국해양수산개발원 경제전망·데이터연구실 연구위원 공동연구원 정수빈 한국해양수산개발원 경제전망·데이터연구실 전문연구원
<b>외부연구진</b>	연구책임자 김정우 강릉원주대학교 경제학과 조교수 공동연구원 이지원 강릉원주대학교 산업경영공학 석사과정 공동연구원 김양희 숙명여자대학교 글로벌서비스 학사과정

<b>연구기간</b>	2021. 01. 01. ~ 2021. 12. 31.
-------------	-------------------------------

## **보고서 집필내역**

<b>연구책임자</b>	권장한 제1장 제2절 일부, 제2장 제1절 일부, 제2절, 제4장 제1절 일부, 제2절 일부, 제5장 제1절 일부
<b>내부연구진</b>	장정인 제1장 제1절, 제5장 제2절 정수빈 제2장 제1절 일부, 제3장 제1절
<b>외부연구진</b>	김정우 제1장 제2절 일부, 제3절 일부, 제3장 제2절 일부, 제4장 제1절 일부, 제2절 일부, 제5장 제1절 일부 이지원 제1장 제3절 일부, 제3장 제2절 일부, 제4장 제1절 일부, 제2절 일부 김양희 제1장 제3절 일부

<b>산·학·연·정 연구자문위원</b>	강인순 삼정회계법인 RCS 차장 정영호 강원연구원 책임연구원
---------------------------	--------------------------------------

※ 순서는 산·학·연·정 순임

---

# 목차

## 01

### 서론 \_p

제1절 연구배경 및 목적	1
1. 연구 배경 및 필요성	1
2. 연구 목적	2
제2절 연구 내용과 방법	3
1. 연구 내용	3
2. 연구 방법	3
제3절 선행연구 검토	5
1. 계량경제모형을 활용한 해양수산업 연구 사례	5
2. 머신러닝기법 선행연구	12
3. 시사점	18

## 02

### 전망모형 운영사례 \_19p

제1절 주요기관 전망모형 운영사례	19
1. 한국은행	20
2. 산업연구원	26
3. 에너지경제연구원	29
4. 대외경제정책연구원	34
제2절 소결	36

## 03

### 해양수산업 데이터 베이스 구축 \_39p

제1절 해양수산업 DB 구성	39
1. 해양수산업 DB 활용 통계	39
2. 해양수산업 DB 산업분류	43
3. 주요 해양수산업 DB 구성	46
2절 해양수산업 DB 자동화	48

1. 기존 데이터베이스 구축 방법	48
2. 개선된 데이터베이스 구축 방법	50

## 04 2021 해양수산업 전망모형 \_53p

제1절 다모형 접근법	53
1. 베이지안 VAR 모형	54
2. 머신러닝기법	57
제2절 2022 해양수산업 전망	76
1. 베이지안 VAR를 통한 전망	76
2. 머신러닝 기법에 의한 전망	85

## 05 결론\_111p

제1절 요약 및 결론	111
1. 요약	111
2. 결론	113
제2절 향후 추진 과제	114
1. 연구 기여	114
2. 향후 추진 과제	115

참고문헌 \_119p

부록 \_129p

# 표 목차

〈표 1-1〉 계량모형을 활용한 해양수산업 예측 연구	10
〈표 2-1〉 KIET-DIMM 발전과정	28
〈표 2-2〉 에너지원별 전망 구조	32
〈표 3-1〉 한국은행 산업연관표 현황(1960~2019)	40
〈표 3-2〉 국민계정 활용 통계 및 시계열	42
〈표 3-3〉 기타 활용 통계 및 시계열	43
〈표 3-4〉 전망 DB 구축을 위한 산업연관표 대분류	44
〈표 3-5〉 순자본스톡 통합 코드	46
〈표 3-6〉 주요 해양수산업 DB의 구성	48
〈표 3-7〉 데이터베이스 구축 자동화 입력 파일	50
〈표 4-1〉 베이지안 VAR 모형 구성	76
〈표 4-2〉 모형별 경제변수	77
〈표 4-3〉 경제변수 전망치 출처	78
〈표 4-4〉 2022 해양수산 주요지표 전망	79
〈표 4-5〉 변수 설명 및 출처	80
〈표 4-6〉 기초통계량	81
〈표 4-7〉 모형별 최적 시차	82
〈표 4-8〉 해양수산업 부가가치 전망	83
〈표 4-9〉 해운업 부가가치 전망치	88
〈표 4-10〉 해운업 부가가치 전망치 신뢰구간	89
〈표 4-11〉 해운업 부가가치 예측성능 비교	90
〈표 4-12〉 항만업 부가가치 전망치	92
〈표 4-13〉 항만업 부가가치 전망치 신뢰구간	93
〈표 4-14〉 항만업 부가가치 예측성능 비교	94
〈표 4-15〉 조선업 부가가치 전망치	96
〈표 4-16〉 조선업 부가가치 전망치 신뢰구간	97
〈표 4-17〉 조선업 부가가치 예측성능 비교	98
〈표 4-18〉 수산업 부가가치 전망치	99
〈표 4-19〉 수산업 부가가치 전망치 신뢰구간	100
〈표 4-20〉 수산업 부가가치 예측성능 비교	101

---

〈표 4-21〉 해양관광업 부가가치 전망치 .....	103
〈표 4-22〉 해양관광업 부가가치 전망치 신뢰구간 .....	104
〈표 4-23〉 해양관광업 부가가치 예측성능 비교 .....	105
〈표 4-24〉 전 산업 부가가치 전망치 .....	106
〈표 4-25〉 전 산업 부가가치 전망치 신뢰구간 .....	107
〈표 4-26〉 전 산업 부가가치 예측성능 비교 .....	108
〈표 4-27〉 개별산업 부가가치 전망치(최저 절대비오차 기준) .....	109
〈표 4-28〉 전 산업 부가가치 전망치 및 절대비오차 .....	110

---

## 그림 목차

〈그림 1-1〉 연구 흐름 .....	5
〈그림 2-1〉 한국은행 거시경제모형 체계 .....	21
〈그림 2-2〉 한국은행 거시계량모형(BOK20) 체계 .....	24
〈그림 2-3〉 거시계량투입산출모형 흐름도 .....	25
〈그림 2-4〉 통합모형의 순환구조 .....	27
〈그림 2-5〉 KIET-DIMM 기본구조 .....	28
〈그림 2-6〉 단기 전망 모형 구조 .....	30
〈그림 2-7〉 KEEI-EGMS의 구조 .....	33
〈그림 3-1〉 해양수산업 중심 산업연관표 구축 엑셀 작업 .....	49
〈그림 3-2〉 데이터베이스 구축 자동화 결과 파일(예시) .....	51
〈그림 4-1〉 랜덤포레스트 변수 중요도 .....	60
〈그림 4-2〉 지도 주성분 분석 알고리즘 .....	61
〈그림 4-3〉 LASSO .....	64
〈그림 4-4〉 Elastic net .....	65
〈그림 4-5〉 매트릭스 $D$ (예시) .....	67
〈그림 4-6〉 $k$ -최근접 이웃법 .....	68
〈그림 4-7〉 $k$ -평균 군집화 .....	70
〈그림 4-8〉 주성분분석 .....	71
〈그림 4-9〉 서포트벡터회귀 .....	72
〈그림 4-10〉 Bagging .....	74
〈그림 4-11〉 베이지안 VAR 전망 절차 .....	78
〈그림 4-12〉 2022 해양수산업 전망 .....	85
〈그림 4-13〉 해운업 부가가치 전망치 비교 .....	87
〈그림 4-14〉 항만업 부가가치 전망치 비교 .....	91
〈그림 4-15〉 조선업 부가가치 전망치 비교 .....	95
〈그림 4-16〉 수산업 부가가치 전망치 비교 .....	99
〈그림 4-17〉 해양관광업 부가가치 전망치 비교 .....	102
〈그림 4-18〉 전 산업 부가가치 전망치 비교 .....	106
〈그림 5-1〉 해양수산업 전망모형 구축 단계(안) .....	116
〈그림 5-2〉 KMI 전망시스템(안) .....	117





# 요약

## 1. 연구 목적

- 본 연구는 해양수산업 전망모형 개발 연구의 연속과제로 해양수산 경제전망을 위한 데이터베이스 자동화와 해양수산의 특성과 부합하는 전망모형을 개발하는 것이 목적
- 해양수산업 데이터베이스 자동화 프로그램을 구축하여 경제적으로 해양수산업 데이터베이스를 구축하는 방안을 마련함
- 베이지안 VAR 모형과 머신러닝 기법을 활용하여 해양수산업 특성에 맞는 방법론을 적용하고 2022 해양수산업 부가가치를 전망함

## 2. 연구 방법 및 특징

### 1) 연구 방법

- 제1장 서론에서는 연구 배경과 목적, 연구 범위 및 내용을 서술
- 또한, 해양수산업 전망 관련 선행연구를 고찰하고, 시사점을 제시함
- 제2장은 우리나라 주요기관 전망모형 운영사례를 검토
- 우리나라 주요 기관의 전망시스템을 고찰하고, 각 전망시스템의 특징을 살펴봄

- 
- 제3장은 해양수산업 데이터베이스 자동화를 수행
    - 기존의 해양수산업 데이터베이스 구축방법은 유지하되 통계패키지를 활용하여 원자료를 입력하면 자동으로 데이터베이스가 구축되도록 프로그램을 작성함
  - 제4장에서는 2022 해양수산업 부가가치 전망모형을 구축
    - 베이지안 VAR 모형을 활용하여 시계열 부족문제를 극복하고, 해양수산업 주요 지표와 유기적인 관계를 고려한 전망치를 시산함
    - 머신러닝 기법을 적용함으로써 전망치의 강건성과 예측력을 제고함
  - 제5장 결론에서는 연구의 결과를 요약하고, 향후 과제에 대해 논의

## 2) 연구 특징

- 해양수산업 데이터베이스 자동화 프로그램을 구축함으로써 해양수산업 데이터베이스 작성에 소요되는 시간을 단축하고, 그로 인해 연구 효율성 증대를 기대
- 시계열 모형인 베이지안 VAR 모형에 의한 2022 해양수산업 전망치를 시산
  - 해양수산업 주요 부문의 여러 변수사이의 복잡한 동적관계를 전망에 반영하고, 현재 확보 가능한 시계열의 수가 부족하기 때문에 베이지안 VAR 모형을 통해 해양수산업 부가가치를 전망함
  - 해양수산업 대분류 중 유사한 성격을 지닌 산업을 중심으로 4개 부문 모형을 구축하고, 각 부문과 밀접한 관계가 있을 것으로 알려진 경제변수로 모형을 구성함

〈요약 표-1〉 베이지안 VAR 모형 구성

모형	해양수산업 대분류
수산업	수산물 생산업, 수산물 가공업, 수산물 유통업
해운·항만업	해운업, 항만업
조선·제조·건설업	선박 및 해양플랜트 건조 수리업, 해양수산 기자재 제조업, 해양자원 개발 및 건설업
기타 해양수산업	해양수산 레저관광업, 해양수산 관련 서비스업

- 경제변수는 ‘2022 해양수산 주요지표 전망’ 보고서에서 제시된 부문별 주요변수를 사용함
- 2022 전망은 경제변수에 대한 KMI 및 주요 기구의 전망치로 시나리오를 설정한 후, 부가가치를 전망함
- 머신러닝 기법은 독립변수 중 적절한 변수를 선별하기 위한 변수 선별 기법을 적용하고, 많은 독립변수로 인한 과적합 문제를 완화하기 위한 머신러닝 기법을 활용. 또한, 다양한 머신러닝 기법의 전망치를 종합하여 새로운 전망치를 도출하는 예측조합법을 적용
- 해양수산업 산업별 부가가치를 종속변수로 하고, 산업연관표 계정, 거시경제변수 등 60개 내외의 독립변수를 사용하여 전망치를 시산함
- 다양한 머신러닝 기법 및 예측조합법의 전망 결과의 정확도를 비교함

### 3. 연구 결과

- 베이지안 VAR에 의한 2022 해양수산업 전망은 국내외 경제의 회복세와 함께 해양수산분야도 양호한 흐름을 이어갈 것으로 예상되나 증가 폭은 전년에 비해 감소할 것으로 전망
- 2022년 해양수산업의 실질 부가가치는 약 3.3%의 성장세를 보일 것으로

---

예상되며, 우리나라 GDP에 대한 직접 기여도는 지난해에 이어 약 2.5% 수준을 유지할 것으로 전망함

- 머신러닝 기법을 활용한 전망 결과는 변수 선별 방식, 머신러닝 기법 및 예측조합법의 종류에 따라 상이했으나, 머신러닝 기법 중에서는 BAGG 및 LASSO, 그리고 예측조합법 중에서는 Fore1 및 Fore2가 가장 높은 예측성능을 보이는 경우가 많이 관찰

# 01

## 서론

### 제1절 연구배경 및 목적

---

#### 1. 연구 배경 및 필요성

2021년 세계경제는 회복세를 보였다. 2021년 세계 교역량은 전년 대비 9.7% 증가했으며, 세계 경제성장률은 2020년의 -3.2%에서 2021년에는 5.9%로 회복세를 보였다. 다만 수요증대 대비 공급 부족으로 국제 원자재 가격 등 전반적인 물가는 상승세를 보였다. 팬데믹은 디지털 기술 기반 경제 구조 전환의 시기를 예상보다 앞당기고 글로벌 가치사슬에도 큰 변화를 가져다주었다. 세계 GVC 참여율은 2018년 59.3%에서 2020년 52%까지 하락했고, 중국 중심의 생산 네트워크도 신남방 국가로 확장되고 있다. 이러한 세계 경제의 대전환기에, 해양수산분야는 글로벌 가치사슬의 재편성과 기후환경규제에 대비하고 디지털 해양수산으로의 전환을 준비해야 하는 시기를 맞이했다. 해양수산업의 성장전략을 마련하기 위해서는 객관적인 근거에 기반 한 해양수산업의 경제 내 기여도에 대한 진단과 전망이 필요하다.

해양수산업은 생산, 부가가치, 고용 측면에서 국민경제에 기여해왔다.

---

해양수산업은 국내 32개 산업과 비교했을 때, 총산출 측면에서 14위(2019년 투입산출표 기준)를 차지했다. 이는 한국 자동차 산업의 71.5% 수준이며, 농림업의 약 2.5배에 해당한다. 또한 해양수산업은 1차 산업부터 4차 산업까지 다양한 산업을 포함하고 있는 거대 산업군이다. 해운, 항만, 조선업은 글로벌 경기에 따라 수요가 파생되는 산업이며, 수산업과 해양레저관광업은 최종소비재로 주로 국내경기요인, 기후환경요인, 국내외 규제요인 등의 영향을 받는다. 따라서 해양수산업의 전망을 위해서는 팬데믹 이후 변화된 국내외 거시경제 요인과 산업별 요인들을 고려해야 한다.

이러한 복합적인 해양수산업을 예측하고 전망하기 위해 KMI는 해양수산업 전망모형 개발연구를 수행하고자 한다. 수산, 해운, 항만, 조선, 기차재, 해양레저 등 총 10개 해양수산업의 산출, 부가가치, 고용에 대한 전망모형을 개발하여 현 경제상황에 부합하는 전망을 통해 대전환기의 해양수산업 성장전략 마련에 기여하고자 한다.

## 2. 연구 목적

본 연구는 2019년부터 매년 수행되어온 ‘해양수산업 전망모형 개발 연구’의 연속과제로 해양수산 경제전망을 위한 데이터베이스 자동화와 해양수산의 특성과 부합하는 전망모형을 개발하는 것을 목적으로 한다.

이를 위해 첫째, 해양수산업 데이터베이스 자동화 프로그램을 구축하여 경제적으로 해양수산업 데이터베이스를 구축하는 방안을 마련한다. 둘째, 베이지안 VAR 모형과 머신러닝 기법을 활용하여 해양수산업 특성에 맞는 방법론을 적용하고 2022 해양수산업 부가가치를 전망한다.

## 제2절 연구 내용과 방법

### 1. 연구 내용

본 연구는 해양수산업 전망모형 개발 연구의 연속과제로 해양수산 경제 전망을 위한 데이터베이스 자동화와 해양수산의 특성과 부합하는 전망모형을 개발하는 것을 목적으로 한다.

첫째, 해양수산업 데이터베이스 자동화는 기존의 데이터베이스 구축방법을 유지하되 통계패키지를 활용하여 데이터베이스를 자동화한다. 즉 한국은행의 산업연관표를 기초로 하여 해양수산업 특수분류의 경제지표를 추출하는 방식을 활용한다. 자동화는 한국은행의 경제통계시스템 자료, 한국은행의 비공개 자료 등 기초자료를 입력하면 해양수산업 계정이 자동으로 산출되는 방식을 따르게 될 것이다.

둘째, 해양수산업 전망모형 개발은 시계열모형이 베이지안 VAR(Vector Autoregressive, 벡터자기회귀) 모형과 머신러닝기법을 활용하여 모형을 개발하고, 2022년 해양수산업 부가가치 전망을 수행한다.

### 2. 연구 방법

본 연구는 모두 5장으로 구성되었다. 제1장 서론에서는 연구 배경과 목적, 연구 내용 및 방법을 서술한다. 또한 해양수산업 분야에 계량경제모형과 머신러닝 기법을 적용하여 전망을 한 선행연구를 고찰하고, 시사점을 도출한다.

제2장에서는 주요기관의 전망모형 운영사례를 검토한다. 우리나라의 주

---

요 기관인 한국은행, 산업연구원, 에너지경제연구원, 대외경제정책연구원의 전망시스템을 고찰하고, 분석 대상 및 목적에 따른 각 기관 전망모형의 특징을 살펴본다.

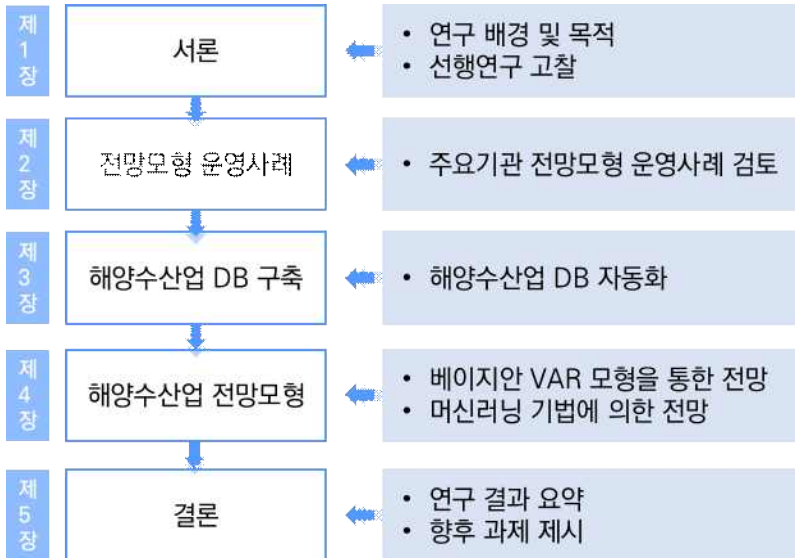
제3장에서는 해양수산업 데이터베이스 자동화를 수행한다. 통계패키지 R을 활용하여 해양수산업 데이터베이스를 자동화한다. 기존의 엑셀을 활용한 작업과정을 프로그래밍 언어로 변환하여 원자료를 입력하면 데이터베이스가 자동으로 산출되는 프로그램을 개발한다.

제4장은 2022년 해양수산업 전망을 위한 해양수산업 전망모형을 구축한다. 해양수산업 주요 부문의 경제지표와 부가가치 간의 유기적인 상관관계를 고려한 베이지안 VAR 모형을 활용한 전망치를 제시한다. 또한 전망치의 예측력을 제고하기 위해 머신러닝 기법을 활용한다. 독립변수 중 적절한 변수를 선별하기 위하여 변수 선별 기법을 적용하고, 많은 독립변수 개수를 고려하여 독립변수가 종속변수에 주는 영향력을 조절하는 일련의 머신러닝 기법을 사용한다. 마지막으로, 다양한 머신러닝 기법의 전망치를 종합하여 새로운 전망치를 시산하는 예측조합법을 적용한다.

제5장 결론에서는 연구의 결과를 요약하고, 향후 과제에 대해 논의한다.



〈그림 1-1〉 연구 흐름



자료: 저자 작성

## 제3절 선행연구 검토

### 1. 계량경제모형을 활용한 해양수산업 연구 사례

#### 1) 해운업

박성진 등(2012)은 내항상선 선원의 수요 전망을 예측하기 위해 내항상선 선원의 수급실태와 선박척수의 변화를 회귀분석으로 분석하였다. 연구 결과, 선박척수의 증가율 0.4%를 기준으로 할 경우 향후 10년간 승선인원의 수요는 7,890~8,025명으로 예측되며, 0.5%를 기준으로 할 경우 향후 10년간 승선인원의 수요는 7,894~8,063명으로 예측되었다.

---

기준을 0.1~1.0%폭의 수준으로 상정할 경우 7,879~8,063명으로 예측되어 연구 시점 당시 기준 추가 승선인원의 필요성을 시사했다.

Lu *et al.*(2015)은 에너지 효율을 높이는 최적 경로 탐색을 위해 기존의 배의 속도를 측정하는 수리적 모델을 기반으로 새로운 예측 모델을 개발하였다. 이 방법은 선박 운항 데이터와 해상 시험 데이터를 이용하여, 해상과 파도 방향에 따른 추가 저항을 예측하는 기존의 방법을 기반으로, 파도 및 기상 조건에 대한 선박의 성능을 예측하였다. 이후 수에즈막스(Suezmax) 유조선과 아프라막스(Aframax) 유조선의 두 가지 사례 연구를 통해 제안된 모델을 적용한 결과, 선택한 최적 경로의 연료 소비량은 기록된 경로보다 10% 감소됨을 확인하여 운항 성능 예측에 신뢰할 수 있음을 확인하였다.

## 2) 항만업

오진호(2016)는 제주도의 무역항인 제주항과 서귀포항의 미래 물동량을 알아보기 위해 2005년에서 2015년의 물동량 데이터를 활용하였다. 이때 각 무역항별 세부 품목의 물동량 예측을 위하여 비중분석을 통하여 도출한 주요 품목 세 가지에 관한 미래 물동량 예측에 선형회귀분석 기법을 이용하였다. 분석 결과, 동식물성 생산품들은 지속적으로 물동량이 상승하고, 모래 및 유류 품목은 초반에는 하락세를 보이지만 2025년까지는 지속적으로 물동량이 상승할 것이라 예측되어 항만인프라 개발을 시사했다.

Jugović *et al.*(2011)은 항만 서비스의 교통 수요 예측을 위해 시장 변화, 항만 내륙의 변화, 상품 이동 경향의 변화 등을 이동한 회귀분석 모델을 개발하여 크로아티아의 최대 항만인 리예카 항에 적용하여 모델

의 예측성능을 테스트하였다. 모델 적용 결과, 이 항만에서 시장 수요와 컨테이너의 환적은 당시의 수용량을 넘어설 것으로 예측되어 항만 인프라 확충을 시사했다.

Zhang *et al.*(2013)은 항만화물 처리량 예측 모델 정확도를 높이기 위해 두 가지 예측모델(grey-forecast model and Logistic-growth-curve model)을 혼합한 예측 모델을 개발하였다. 상기 모델을 중국 남부 항만에 적용한 결과, 2015년 항만 화물 처리량을 6억 4400만 톤으로 예측하며 항만 개발에 대한 적절한 가이드라인을 제시하였다.

### 3) 조선업

백천현 외(2014)는 조선업의 온실가스 감축 잠재량을 분석하기 위해 상향식 최적화 모형을 사용하였다. 이때, 조선업의 최종수요, RES, 공정 압출력 요소 등을 정의하고 국내 통계데이터를 분석하여 필요한 데이터를 토출하였다. 그리고 상향식 최적화 모형 중 하나인 MESSAGE 분석 모형을 사용하여 연도별 건조량, 에너지 사용량, 온실가스 배출량을 구하고 3가지 감축기술을 도입할 시의 감축잠재량을 분석하였다. 분석 결과, 당시 조선업에서 도입할 수 있는 감축 기술로 절감할 수 있는 온실가스 배출량은 극히 미미한 것으로 드러나 조선업에서의 에너지 사용량 감소를 시사했다.

배남규 외(2020)는 국가 경제에 막대한 영향을 미치는 조선업에 대해 회계정보를 바탕으로 한 도산예측모형을 개발하고자 하였다. 비도산기업과 도산기업의 회계자료를 비교하기 위하여 다변량분석모형을 사용하였으며 동일업종, 유사규모 기업들의 비교를 통해 도산하지 않고 생존하기 위한 재무지표 예측항목들을 제시하였다.

---

Li(2010)는 중국 조선 산업을 예측하기 위해 회귀 분석을 이용하였다. 이때, 금리, 선박 가격, 조선소 규모, 철강 가격, 정부 정책, 노동 생산성, 국제 무역, 출하 수익 및 시장 기대치 등 다양한 변수를 이용하였다. 세계 경제, 정부 지원 등의 변수를 고려하여 중국의 조선 산업은 가장 높게는 평균 12.6% 성장하지만, 가장 낮게는 12% 성장하는 것으로 예측하였다.

#### 4) 수산업

남종오·노승국(2012)은 대한민국 천해양식 굴의 월별 생산량을 예측하기 위해 다중회귀분석모형, SARIMA 모형, VECM 모형 등을 이용하였다. 분석 결과, 다중회귀분석모형이 다른 두 모형에 비해 상대적으로 실제값과 예측값 사이의 오차가 적은 것으로 드러났으나, 예측 적합도 측면에서는 VECM 모형이 다중회귀분석모형보다 더 적합한 것으로 나타났다.

박해훈·윤갑동(1996)은 대한민국 멸치 어업에서의 1971년~1992년 동안의 22년간 월별 어획량 자료를 바탕으로 시계열 분석 기법을 이용하여 향후의 어획량을 분석하고 예측하였다. 이때, 자료의 비정상성을 줄이기 위해 어획량 자료를 로그(log) 변환과 Box-Cox 변환을 시켜 ARIMA 모형에 적용하여 두 결과를 비교 분석하였다. 분석 결과 Box-Cox 변환시킨 것이 로그 변환시킨 것보다 예측오차가 적었다. 그리고 이 모형을 이용하여 1991~1992년 사이의 월별 예측 어획량과 실제 어획량과의 월별 오차범위는 1.0~63.2%로 나타났으나 전체 2년 동안의 총어획량에 관한 오차는 12.3%의 수준으로 나타났다.

이광남·정진호(2012)는 어업용 면세유류 사용량의 예측을 위해

Cochrane Orcutt 반복절차 모형을 사용하였다. 독립변수로는 어선세력, 어업용 면세유류 가격 및 고효율 기술의 개발 등과 같은 기타 기술요인을 고려하였다. 이를 기반으로 어업용 면세유류 사용량을 분석한 결과, 2015년에는 4,508~4,840 드림, 2020년에는 4,236~4,580 드림 정도가 사용될 것으로 추정되어, 2010년 대비 2015년에는 약 86% 수준, 2020년은 81% 수준으로 감소할 것으로 예측되었다.

Velarde *et al.*(2004)은 걸프만에 있는 태평양 정어리의 총 어획량과 단위 작업당 어획량을 예측하기 위해 선형모형을 사용하였다. 총 어획량의 경우, 남부 진동 지수, 히르만 갈매기의 클러치 크기, 엘레강스 갈매기의 식단에서 정어리 비율을 사용한 모델에 의해 54%로 예측되었다. 단위 작업당 어획량의 경우, 엘레강스 갈매기의 식단에서 정어리의 비율, 히르만 갈매기의 번식 성공, 걸프 지역의 봄철 해수면 온도 이상 등을 변수로 사용한 모델에 의해 73%의 정확도로 예측되어 갈매기 등의 바다새 번식 생태계는 전세계 및 지역의 해양의 조건들과 밀접한 것으로 나타났다.

Gomara *et al.*(2021)은 계절별 열대 태평양의 해양 생태계 자원의 예측을 위해 지구 해양 생태계를 포함하는 대기 재분석 데이터, 역사적 어획 기록 및 FishMIP 시뮬레이션을 활용하였다. 이때, 물리적 생물학적 지표 및 통계는 시계열 분석의 기법을 이용하여 분석하였으며, SST 기반 통계 계절 예측 모형(S4CAST)을 활용하기도 하였다. 분석 결과, 적도 부근의 해양온도는 열대 태평양 지역 어업의 중요한 예측 요소가 되며, 총 3년치의 예측이 가능한 것으로 나타났다.

## 5) 해양관광업

김길훈 외(2008)는 제주도 해양관광 발전계획을 수립하는 데 있어서 제주도의 주요 권역별 해양관광 수요를 예측하였다. 이때, 제주도 남동부와 남서부 해양관광 수요 예측을 위해 회귀분석 및 시장접근법을 이용하였다. 분석 결과, 해양관광 수요 예측치는 지역별로 다소 차이가 나타났으나 시간에 따라 점진적으로 증가하는 추세를 보이는 것으로 나타났다.

황경후 외(2015)는 독도 관광 수요예측을 위해 시계열 모형을 이용하였다. 이때, 독도입도 관리사무소에서 제공하는 입도 시계열 자료를 활용하였다. 분석 결과, ARIMA 모형이 가장 적합한 것으로 확인되었고, 관광 수요예측 결과 독도 관광 수요가 증가하기보다는 일시적 정체현상을 보일 것으로 나타났다.

Geng *et al.*(2020)은 수질 관리정도와 관광 시스템의 메커니즘을 분석하고, 향후 4년간의 상호 작용 결과를 예측하였다. 분석 방법으로는 변수 간 상관관계를 분석하는 Coupling Coordination Degree Model을 사용하였다. 분석 결과, 향후 4년 동안 수질 관리정도와 관광산업 간의 관계는 증대될 것으로 예측되었다.

〈표 1-1〉 계량모형을 활용한 해양수산업 예측 연구

연구자	방법론	분석자료	분야
박성진(2012)	회귀분석	내항상선 선원 수급실태, 선박척수의 변화 등	해운
Lu et al.(2015)	선박 속도 측정 관련 모델	선박 운항 데이터, 해상 시험 데이터 등	해운

연구자	방법론	분석자료	분야
오진호(2016)	비중분석 선형회귀분석	물동량 데이터 등	항만
Jugović et al.(2011)	선형회귀분석	시장 변화, 항만 내륙의 변화, 상품 이동 경향의 변화 등	항만
Zhang et al.(2013)	Grey-forecast model, Logistic-growth-curve model	항만 화물 처리량 등	항만
배남규 외(2020)	다변량분석모델	조선업 회계정보 등	조선업
백진현 외(2014)	상향식 최적화 모형	조선업 최종수요, 공정 압출력 요소 등	조선업
Lu(2010)	선형회귀분석	금리, 선박 가격, 조선소 규모, 철강 가격 등	조선업
남종오·노승국 (2012)	다중회귀분석모형, SARIMA VECM	굴 월별 생산량 등	수산업
박해훈·윤갑동 (1996)	시계열 분석	멸치 월별 어획량 자료 등	수산업
이광남·정진호 (2012)	Cocharane Orcutt 반복절차 모형	어선세력, 어업용 면세유류 가격 등	수산업
Gomara et al.(2021)	시계열 분석	대기 재분석 데이터, 어획 기록 등	수산업
Velarde et al.(2004)	선형회귀분석	갈매기 식단에서의 정어리 비율, 겔프 지역 해수면 온도 등	수산업
김길훈 외(2021)	선형회귀분석, 시장접근법	제주도 관광객 수 등	해양 관광업
황경후 외(2015)	시계열 모형	독도 입도 시계열 자료 등	해양 관광업
Geng et al.(2020)	Coupling Coordination Degree Model	수질 관리 정도 등	해양 관광업

자료: 저자 작성

선행연구들을 검토한 결과, 대부분의 연구들이 한정된 크기와 변수 개수의 데이터를 사용하고 있다는 점을 발견할 수 있다. 예를 들면, 특정 선박, 특정 항만 등으로부터 얻어진 데이터를 토대로 연구가 수행되어 연구 결과

---

를 일반화하여 해석하기에는 다소 무리가 따른다고 볼 수 있다. 방법론적으로는 회귀모형, 시계열 모형 등이 주로 사용되었고, 분야별로 통계적 방법론 외의 방법론이 가미된 연구들도 있었다. 회귀모형의 경우 독립변수의 개수는 관측치 개수를 초과할 수 없고, 시계열 모형의 경우 1개의 종속변수만을 사용하는 경우가 대부분이다. 또한, 분야별 특정 방법론을 사용하는 경우 이론적으로 검증된 특정 변수들만을 사용하는 경우가 대부분이었다. 해양수산업 산업별 부가가치가 해양을 자연 환경뿐만 아니라 경제사회적 요소들이 영향을 준다는 점을 고려하고, 관련 변수의 개수와 예측력 사이에 많은 경우 양(+)의 상관관계가 있다는 사실을 받아들인다면 다양한 독립변수들을 활용할 수 있는 머신러닝 기법을 해양수산업 산업별 부가가치 전망에 활용하는 것은 새로운 결과를 기대할 수 있는 접근법이라고 보인다.

## 2. 머신러닝기법 선행연구

앞에서는 기존의 계량모형 등의 방법론으로 해양수산업 각 분야의 전망 연구들이 어떻게 수행되었는지 살펴보았다. 머신러닝 분야는 다양한 기법이 존재하며, 모형의 가정도 완화된 경우가 많으므로 적용 분야가 비교적 다양하다고 볼 수 있다.

### 1) 해운업

김동현 외(2019)는 SVM(Support Vector Machine) 알고리즘 기반의 마력해석법을 제안하여 기존 방법들의 단점인 선박 모델 데이터 준비 단계를 축소하여 부정확한 마력 예측 성능을 개선하였다. 기존의 예측 방법은 통제된 환경에서만 분석이 가능하므로 해상상태와 화물 적재



상태에 따라 분석을 할 수 없는 운항 구간이 있는 문제가 있었으나, 제안된 SVM 기반 마력해석법은 모든 항해 구간에 대한 해석이 가능한 것으로 나타났다. 또한, 기존 해석 대비 정확성에서 SVM을 활용한 예측 모델이 우수한 예측 성능을 보였고, 데이터가 많을수록 정확도가 올라가는 장점이 있는 것으로 나타났다.

Abebe *et al.*(2020)은 선박 예상 도착시간을 정확하게 예측하여 경제성이 높은 수송 경로를 탐색하는 모델을 개발하였다. 1년간 76척의 선박 자동식별시스템(Automatic Identification system, AIS) 및 해양 기상 데이터를 이용하였으며, 분석방법으로는 변수선별 기법, DTR(Decision Tree Regressor), 앙상블 방법(Ensemble Methods)들을 결합한 방법을 사용하였다. 분석 결과, 제안된 데이터 기반 모델은 약 96% 이상의 정확도로 선박 속도를 예측할 수 있는 높은 성능을 보였다.

Hu *et al.*(2019)은 선박 연료 소비 데이터에 BPNN(Back Propagation Neural Network)과 GPR(Gaussian Process Regression)을 적용하여, 두 기법의 예측 성능을 비교·분석하였다. 분석 결과, 두 기법 모두 선박 연료 소비량을 정확하게 예측할 수 있었으며, GPR의 평균 예측 정확도(0.9887)가 BPNN(0.9817)보다 약간 높지만, GPR의 러닝 타임(2236.4초)이 BPNN의 러닝 타임(14.7초)에 비해 훨씬 더 길기 때문에 BPNN이 선박 연료 소비량의 실시간 예측 방법으로 더욱 적합하다고 주장하였다.

## 2) 항만업

김두환 외(2020)는 2004년부터 2018년까지의 월별 부산항 컨테이너 물동량 자료를 활용하여 부산항 컨테이너 물동량 예측을 시도하였다. 또한

---

해운 및 항만물류산업의 급격한 변화로 인한 기존 시계열 예측 모형인 ARIMA(Autoregressive integrated moving average)의 모형의 한계점을 지적하며 컨테이너 물동량 흐름의 변동성과 복잡성이 증가하는 상황에서 예측 정확도를 향상시키기 위해 딥러닝 기법 중 LSTM(Long Short-Term Memory) 모형을 활용한 컨테이너 물동량 예측 모형을 제안하였다. SARIMA(Seasonal ARIMA) 모형과 LSTM 모형의 예측 정확도를 비교한 결과, LSTM 모형이 SARIMA 모형보다 예측 정확도가 높고, 예측치가 실측치의 특성을 잘 반영하였다고 주장하였다.

Ighravwe *et al.*(2019)은 ANN(Artificial Neural Network)과 ARIMA를 이용하여 항만의 생산성과 항만의 정박 효과 예측을 시도하였다. 예측에는 나이지리아 라고스(Lagos)의 어파(Apapa) 항만에서 수집한 항만 정박 일 수, 정박지에서 보낸 일 수 및 선박에서 처리한 톤 수 데이터가 사용되었다. ANN 모델을 적절히 학습시키기 위해 부트스트랩 샘플링 기법을 이용하였다. 그 결과, ANN을 사용한 항만 생산성 예측 상관 계수는 0.998로 ARIMA를 사용한 0.9862보다 높게 나타났고, 이에 따라 ANN 모델이 더 적합하다고 주장하였다.

### 3) 조선업

이상원 외(2019)는 한국의 조선 산업은 대내외 환경 변화로 인해 위기 상황에 처해 있으며, 이 위기를 극복하기 위해서 선박 신수요 예측을 통한 제품 및 기술의 선제적 개발이 필요하다고 주장하였다. 이에 따라 선박 신수요 예측을 위해 선박 빅데이터에 기반한 인공지능 알고리즘 개발을 시도하였다. AIS 데이터를 기반으로 선박 운항 패턴을 분석하였고, 기술 동향에 대한 빅데이터 텍스트 마이닝 분석 또한 시도하

였다. 그 결과 장비 및 기자재 제조업체를 위한 지속 가능한 신사업 모델 개발로 조선소 및 선박 기자재 업체에 대한 신성장동력을 창출할 수 있을 것이라고 주장하였다.

Kholief *et al.*(2017)은 선박 등의 표면 결함에 대한 비접촉 검사의 필요성 증대에 따라, 머신러닝 기법을 활용하여 열간압표면 결함을 감식하고 분류하였다. 균열, 구멍, 긁힘 등의 300개의 샘플로 이루어진 표면 결점 데이터를 이용하여 전진 방향(Feed-forward) 인공신경망과 DAN(Deep Auto-encoder Network)을 이용하여 결함 분석을 시도하였다. 그 결과, 아주 낮은 오판율로 우수한 결함 감지 결과를 보였다고 주장하였다.

Cai *et al.*(2019)은 키홀과 용융지의 움직임이 용접층의 품질과 관련이 있다고 보고 고속 카메라를 기반으로 한 온라인 모니터링 시스템을 구축하여 키홀 및 용융지의 이미지를 수집하여 용접 품질을 예측하였다. BPNN 모델과 RBFNN(Radial Based Function Neural Network) 모델, SVR(Support Vector Regression) 모델을 사용했으며 그 결과, BPNN 모델이 RBFNN 모델과 SVR 모델에 비해 가장 예측성능을 나타내었다고 보고하였다.

#### 4) 수산업

Enkhjargal Unuzaya *et al.*(2020)은 적조의 일종인 *Cochlodinium Polykrikoide*를 기계학습 방법과 정지궤도 위성 영상을 활용하여 탐지하는 방법을 제안하였다. 이 연구에서는 국립수산과학원의 적조 속보 자료를 활용하였으며, 기계학습 모형을 학습시키기 위해 GOCI(Geostationary Ocean Color Imager) Level 2 자료를 활용하였

---

다. 기계학습 모델은 로지스틱 회귀모형, 의사결정나무 모형, 랜덤포레스트 모형을 사용하였다. 성능 평가 결과 전통적인 GOCI 영상 기반 적조 탐지 알고리즘 결과(75%)와 비교해 보았을 때 88~98%로 정확도 향상이 기계학습을 통하여 이루어졌다고 주장하였다.

## 5) 해양관광업

CHANG *et al.*(2020)은 제주도 해변에 플라스틱 파편 등으로 인한 환경문제가 빈번하게 발생하고 있고, 플라스틱 폐기물의 양이 계속 증가함에 따라 다양한 측면에서 관리 및 보호가 필요하다고 주장하며 오염된 해변으로부터 얻은 이미지 데이터를 이용하여 CNN(Convolutional Neural Network) 알고리즘과 이전 학습(Transfer Learning)을 통한 분석을 시도하였다. 분석 결과, 97% 정도의 예측정확도를 얻어 CCTV 확대 보급 등의 시사점을 주었다.

Xie *et al.*(2021)은 기하급수적으로 발전해 온 중국 크루즈 관광객 증가세가 둔화되고 있는 상황에 착안하여, 중국의 구글이라고 할 수 있는 Baidu의 SQD(Search Query Data)와 경제지수 등을 이용하여 크루즈 관광 수요 예측을 시도하였다. 분석 방법으로는 LSSVR-GSA(Least Squares Support Vector Regression model with Gravitational Search Algorithm)모형을 이용하였다. 시계열 모형 등 다른 방법들과 비교한 결과 LSSVR-GSA 기반 모델이 중국 크루즈 관광 수요를 예측하는 데 가장 유용하다고 주장하였다.

Goncalves *et al.*(2020)은 무인 항공 시스템(Unmanned Aerial Systems, UAS)이 수집한 해변 모래사장과 모래 언덕의 고해상도 이미지 데이터를 이용하여 해양 쓰레기를 탐지하기 위한 랜덤포레스트 기

법과 CNN 기법을 비교하였다. 그 결과, 랜덤포레스트 기법의 탐지율(F-점수 70%)이 CNN 기법의 탐지율(F-점수 60%)보다 높게 나타났다고 보고하며 해변가의 쓰레기 탐지에 머신러닝 기법의 활용 가능성을 시사했다.

〈표 1-2〉 머신러닝 기법에 의한 해양수산업 예측연구

연구자	방법론	분석자료	분야
김동현 외(2019)	SVM 알고리즘 기반 마력해석법	벌크캐리어의 운항 데이터 등	해운
Abebe et al.(2020)	변수 선별 기법, DTR, 앙상블 방법들을 결합한 방법	선박 자동식별시스템, 해양 기상 데이터 등	해운
Hu et al.(2019)	BPNN, GPR	선박 연료 소비 데이터 등	해운
김두환 외(2020)	LSTM, SARIMA	월별 부산항 컨테이너 물동량 자료 등	항만
Ighrawwe et al.(2019)	ANN, ARIMA	항만 정박 일 수, 정박지에서 보낸 일 수, 등	항만
이상원 외(2019)	텍스트 마이닝	AIS 데이터 등	조선업
Cai et al.(2019)	BPNN, RBFNN, SVR	키홀 및 용융지의 이미지 데이터 등	조선업
Kholief et al.(2017)	전진 방향 인공신경망, DAN 등	선박 표면 결점 데이터 등	조선업
Enkhjargal Unuzaya et al.(2020)	로지스틱 회귀모형, 의사결정나무 모형, 랜덤포레스트 모형	적조 속보 자료, GOCI 자료 등	수산업
CHANG et al.(2020)	CNN 알고리즘, 이전 학습	해변의 폐기물 이미지 데이터 등	해양 관광업
Goncalves et al.(2020)	랜덤포레스트, CNN	해변 모래사장, 모래 언덕의 고해상도 이미지 등	해양 관광업
Xie et al.(2021)	LSSVR-GSA	Baidu의 SQD, 경제지수 등	해양 관광업

자료: 저자 작성

---

### 3. 시사점

본 절에서는 해양수산업 전망 관련 선행연구를 고찰하였다. 방법론적으로 계량경제모형과 머신러닝 기법으로 분류하였으며 각 방법론별 시사점은 다음과 같다.

계량경제모형을 활용한 연구는 한정된 데이터를 활용하였다. 예를 들어, 특정 선박 혹은 특정 항만 등의 데이터를 활용하여 분석한 연구가 대부분이었다. 방법론적으로는 회귀모형, 시계열 모형이 주를 이루었으며 연구에 따라 통계적 방법론 외의 방법론을 활용한 연구도 있었다.

머신러닝 관련 연구들은 살펴보면 다양한 머신러닝 기법들이 다양한 데이터들에 적용되어 수행된 것을 알 수 있다. 이것은 연구자마다 새로운 방법론을 제시하는 것이 기존의 계량분야보다 용이하며, 머신러닝 방법론은 예측정확도를 우선시하므로 분야별 변수들에 대한 가정이나, 제약이 기존의 계량분야보다 상대적으로 적다는 점에 기인한다고 볼 수 있다. 특히, 딥러닝 기법 중 시계열 분석이 용이한 LSTM 모형 등이 사용된 연구를 고려하면, 기존의 자기회귀모형 계열의 전통적인 시계열 분석을 대체하거나 보완하는 머신러닝, 딥러닝 분석 방법론들이 많아질 것으로 기대된다.

한편, 기존의 연구들은 보통 몇 가지의 머신러닝 방법론을 사용하여 주어진 데이터를 분석 및 비교하고, 특정한 머신러닝 방법론이 우수한 예측정확도를 보인다고 주장하였다. 하지만 머신러닝 방법론이 다양한 만큼 데이터 특성을 고려하여 다양한 머신러닝 방법론을 사용하면 보다 풍부한 분석결과를 얻을 수 있고, 높은 정확도의 예측치를 확보할 수 있는 방법론 적용도 가능하다고 할 수 있다. 또한, 다양한 머신러닝 방법론을 사용하는 경우에는 머신러닝 방법론의 다양한 분석결과들을 종합하는 접근법도 동시에 필요하다고 볼 수 있다.

## 02

# 전망모형 운영사례

### 제1절 주요기관 전망모형 운영사례

---

선행연구에서는 방법론에 따른 해양수산업 전망 연구를 고찰하였다. 해양수산업에 관한 전망 연구는 주로 계량경제모형과 머신러닝 기법을 사용하였는데, 신뢰성 있는 전망치를 얻기 위해서는 연구 목적이나 분석 대상에 따라 적합한 방법론을 적용하여야 할 것이다.<sup>1)</sup> 계량경제모형은 경제학적 이론에 근거하여 전망치를 도출하기 때문에 경기변동의 원인이나 정책 효과와 같은 이론적 설명가능성에 무게를 둘 경우에 적합하다 할 수 있다. 반면 머신러닝 기법은 변수 간 인과관계보다 높은 예측력 확보가 중요한 연구목적일 때 적합한 방법론이라 할 수 있다.<sup>2)</sup> 본 절에서는 앞서 살펴본 다양한 방법론을 적용하여 전망모형을 개발·운용하고 있는 우리나라 주요 기관 전망시스템에 대하여 알아보고자 한다.

---

1) 황상필 외(2005), p. 26.

2) 최필선·민인식(2018), p. 50.

---

## 1. 한국은행<sup>3)</sup>

한국은행은 경제전망의 정도를 제고하고, 통화정책의 효율성을 높이기 위해 전망의 시계 및 목적별로 다양한 형태의 모형을 개발·운용해 오고 있다.<sup>4)</sup> 다모형 접근법(multi-model approach)는 경제전망에서 발생하는 상충관계를 해소하기 위한 하나의 전략이다. 거시계량모형을 통한 전망은 이론적 정합성과 현실 설명력사이에 상충(trade-off)관계가 있다. 즉 현실 설명력이 높으면 이론적 정합성이 약하고 이론적 정합성이 높으면 현실 설명력이 낮아지는 것이다. 이론체계는 유지하되 현실설명력을 강화하기 위해 다양한 전략이 사용되는데, 그 중 다모형 접근법은 이론적 정합성이 높은 핵심모형을 중심으로 활용하되 예측력이 좋은 시계열모형을 보조모형으로 활용하는 것을 의미한다.<sup>5)</sup>

거시계량모형은 크게 대규모 연립방정식 모형과 DSGE(Dynamic Stochastic General Equilibrium, 동태확률 일반균형)모형으로 나눌 수 있다. 대규모 연립방정식 모형은 수정 및 확장이 용이하다는 점과 높은 데이터 설명력 등의 장점이 있으나 대규모 방정식 추정에 대한 부담과 합리적 기대(rational expectation)를 반영하지 못한다는 점 등이 단점으로 꼽힌다. 한편, DSGE 모형은 미시적 기초(micro-foundation)에 근거하여 경제주체들의 최적화 문제를 바탕으로 경제를 분석한다.<sup>6)</sup> DSGE 모형은 이론적 정합성 측면에서 비교우위가 있으나 복잡한 추정과정, 결과 해석의 어려움 등의 단점이 있다.<sup>7)</sup>

---

3) 박양수(2011), pp. 104~118.

4) 황상필 외(2005), pp. 25~26.

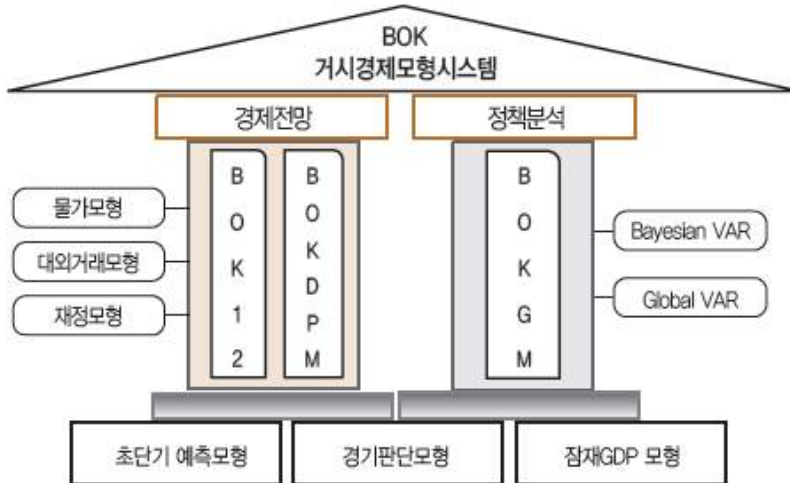
5) 박양수(2011), p. 94.

6) 장정인 외(2021a), p.10.

7) 박양수(2011), p.104.



〈그림 2-1〉 한국은행 거시경제모형체계



자료: 손민규, 김대용, 황상필(2013), p. 17.

## 1) 대규모 연립방정식 모형

대규모 연립방정식 모형은 케인즈(Keynes)의 거시경제학 이론을 토대로 한 구조방정식 체계이다. 즉 가격변동이 수반되지 않고 총수요에 의해 생산이 결정되는 IS-LM 분석에 기반하여 대규모 방정식 체계를 구축하는 것이다.<sup>8)</sup> 대규모 방정식 체계에서 개별 방정식의 모수는 주로 회귀분석을 통해 추정하는데, 추정된 계수의 크기는 과거의 평균적인 행태를 반영하기 때문에 경제의 구조적 변화나 경제 충격이 있는 경우 모형의 유효성이 크게 저하된다. 또한 미시적 근거에 기초한 경제주체들의 최적화에 기반하지 않기 때문에 이론적 기반도 취약하다는 비판도 받고 있다.<sup>9)</sup> 이러한 논란에도 불구하고 거시경제 현상과 그 흐름을 체계적으로 파악하고, 장·단기 예

8) 위의 책, p. 104.

9) 박양수(2011), pp.104~105.

---

측을 일관되게 수행할 수 있다는 장점으로 인해 한국은행을 비롯한 주요기관은 대규모 연립방정식 모형을 경제전망과 정책모의실험에 적극적으로 활용하고 있다.<sup>10)</sup>

한국은행 1970년대부터 대규모 연립방정식 모형을 개발·운용해 오고 있으며 경제구조의 변화 및 경제이론의 발전 등을 반영하여 약 5년 주기로 모형을 개선해 오고 있다. BOK04는 외환위기 이후의 경제구조를 반영하여 구축한 대규모 연립방정식 모형이며, BOK12는 글로벌 금융위기 이후 경제여건 및 경제주체의 행태 변화 등을 반영하여 재구축한 모형이다.<sup>11)</sup> 가장 최근에 개선이 이루어진 연립방정식 모형은 BOK20으로 2012년 이후 경제구조의 변화와 한국은행 국민계정의 기준년 개편으로 인한 경제변수의 변화 및 통화정책의 내생성, 인구구조 변화, 불확실성 확대 등 최근의 경제이론 및 경제구조의 변화도 반영하여 모형을 개편하였다.<sup>12)</sup>

## (1) BOK20 모형

BOK20모형은 기존의 BOK12 모형이 2012년 이후의 경제상황을 반영하는데 한계가 있기 때문에 최근의 경제이론 및 경제구조 변화와 재정부문과 국민경제의 정합성을 유지하도록 보강한 모형이다.

경제구조 변화는 모형에 표본기간 확대 및 더미변수를 재설정, 인구구조 변화, 불확실성 고려, 대외교역 여건 변화를 반영하여 개선하였다. 모형에 테일러 준칙(Taylor's rule) 형태의 금리준칙을 설정하여 기준금리 결정과정을 내생화하였다. 또한 모형이 예측하지 못한 외생적 충격을 통화정책 충격으로 식별하도록 구조화하였다. 원/달러환율을 내생변수로 포함시켜

---

10) 박경훈·임현준·노경서(2020), p.17.

11) 손민규·김대용·황살필(2013), p.14.

12) 박경훈·임현준·노경서(2020), p.16.

소규모 개방경제에서 거시변수의 파급경로를 반영할 수 있도록 하였는데, 환율의 변화를 경상수지, 자본수지 등을 이용하여 설명할 수 있도록 모형을 구성하였다. 마지막으로, 모형에 재정블록을 구축하여 재정부문을 개선하였다.<sup>13)</sup>

BOK20 모형은 최종수요, 대외거래, 물가, 임금 및 고용, 금융 및 자산시장, 재정 부문 등 총 6개 부문으로 구성되었다. 이들 부문은 모두 37개의 행태방정식과 10개의 정의식과 47개의 내생변수와 18개의 외생변수로 구성되었다.

## (2) 거시계량투입산출모형

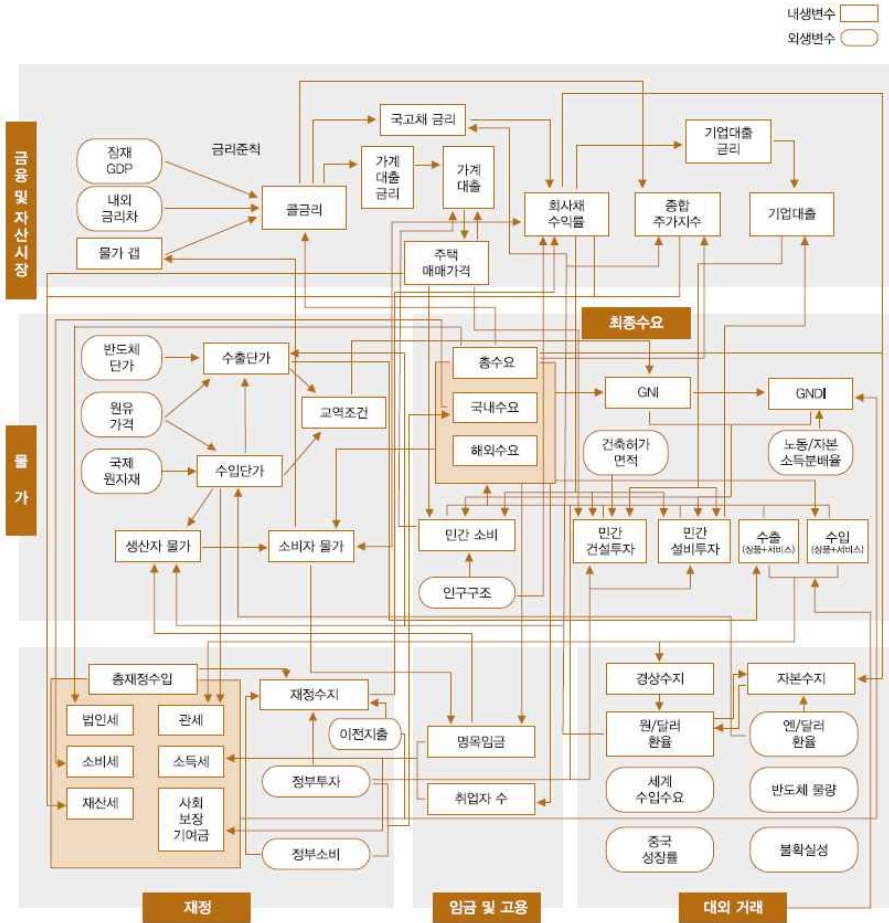
거시계량투입산출모형은 국민경제를 분석하는 연립방정식모형과 산업의 생산을 강조하는 투입산출모형을 결합한 모형으로, 최종수요의 변화나 특정산업의 생산변화가 전체 산업부문에 미치는 영향을 계측할 수 있는 모형이다.<sup>14)</sup> 동 모형은 투입산출모형의 투입계수가 특정연도에 고정되어 있다는 점을 시간변동 투입계수를 추정하여 추정산출량과 실제산출량의 차이를 조정한다는 특징이 있다.

거시계량투입산출모형은 추정생산량블록과 생산블록, 고용블록 및 소득블록에 최종수요블록을 추가한 총 5개의 블록으로 구성되어 있으며, 각 블록은 28개 산업으로 세분되어 있으므로 총 143개의 방정식으로 구성된 대규모 모형이다.

13) 위의 논문, pp. 18~20.

14) 박양수(2011), pp. 107~108.

〈그림 2-2〉 한국은행 거시계량모형(BOK20) 체계

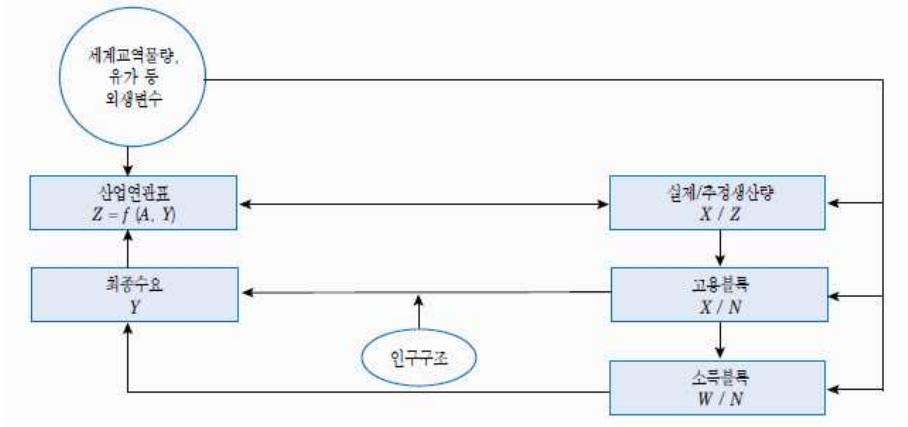


자료: 박경훈·임현준·노경서(2020), p. 31.

거시계량투입산출모형은 중장기 최종수요에 대한 전망뿐만 아니라 중장기 수요에 부합하는 산업별 생산 및 고용수준을 예측할 때 사용된다.<sup>15)</sup>

15) 박양수(2011), p.109.

〈그림 2-3〉 거시계량투입산출모형 흐름도



자료: 황상필, 박양수, 최강욱(2006), p. 35.

## 2) DSGE 모형

대규모 연립방정식 모형은 변수선정의 자의성, 적응적 기대(adaptive expectation) 가정 및 미시적 기초의 부재 등으로 인해 이론적 기반이 취약하다는 약점이 있다. 이러한 약점을 보완하기 위해 등장한 고전학과 DSGE 모형은 합리적 기대와 동태적 최적화(dynamic optimization)을 구현함으로써 연립방정식 모형에 비해 이론적 정합성에 우위를 지니고 있다. 특히 1990년대 들어서 시장의 불완전성을 추가한 뉴케이지안 모형이 등장하면서 DSGE 모형은 거시경제학의 기본적인 분석모형으로 자리잡았다.<sup>16)</sup>

한국은행은 우리나라 DSGE 모형 개발에 선도적인 역할을 수행하고 있으며 2000년대 중반부터 DSGE 형태의 모형 개발에 착수하였다. 2007년에 BOKDSGE(BOKDSM)를 개발했으며, 2009년에는 BOKDPM를 개발

16) 위의 책, p.110.

---

하였다. BOKDSM은 주로 정책효과분석에 초점을 맞춘 모형이며, BOKDPM은 경제전망에 중점을 둔 모형이다.

## 2. 산업연구원

산업연구원은 세계경제 위기의 상시화 및 장기화 경향에 대응을 위해 국내외 경제변수의 변동성을 반영하면서 거시경제와 산업부문 간 연계와 정합성이 담보되는 계량경제모형을 개발·운용해오고 있다. 우리나라의 거시경제에 대한 전망은 적극적으로 이루어지고 있으나 산업부문에 대한 전망은 극히 제한적으로 수행되고 있는데 산업연구원의 대표적인 산업·거시계량모형인 KIET-DIMM(Dynamic Interindustry Macroeconomic Model)은 거시경제와 산업부문을 통합하여 분석하는 모형이다.<sup>17)</sup>

KIET-DIMM은 한국의 경제 상황을 반영하기 위해 소규모 개방경제를 가정한다. 미시적 측면에서 왈라스의 일반균형에 근거한 산업별 수요와 공급의 균형을 유도하고, 거시적 측면에서 소득소비 승수원리, 필립스 곡선 등에 기초하여 총수요와 총공급의 균형으로 수렴하도록 모형이 설계되었다. 통계적 측면에서는 거시경제 변수인 총량변수와 미시경제변수인 산업별 변수간의 통계적 정합성을 유도하는 구조로 설계되었다.<sup>18)</sup>

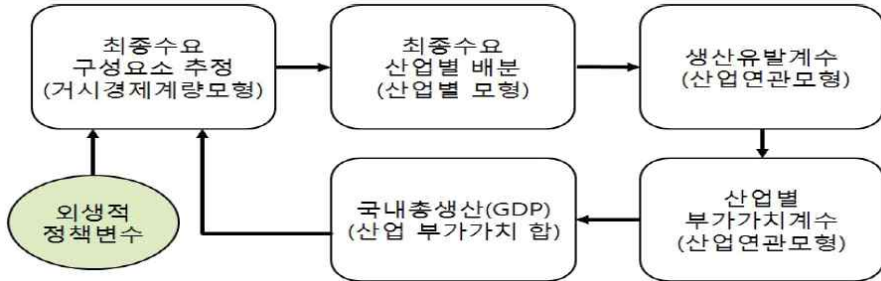
KIET-DIMM은 케인즈(Keynes) 모형과 투입산출모형이 결합된 형태로 “GDP → 최종수요 → 산업별 생산 → 부가가치 → GDP”의 순환구조를 지닌다. 이와 같은 순환구조를 통해 투입산출모형에서 외생적으로 취급하던 최종수요를 거시계량모형을 통해 산출함으로써 거시경제와 산업부문에 대한 상호 정합적인 예측치를 산출한다.<sup>19)</sup>

---

17) KIET 산업·거시경제 단기 계량전망모형 개발 연구 발표자료, p. 1.을 바탕으로 저자 정리

18) KIET-DIMM의 DB 구축현황 발표자료, p. 1.을 바탕으로 저자 정리

〈그림 2-4〉 통합모형의 순환구조



자료: KIET 산업-거시경제 단기 계량전망모형 개발 연구 발표자료, p. 6.

모형의 기본 구조는 거시, 수요·공급, 가격·기술, 대외, 소비의 5개 하부 블록으로 구성되었으며 모형의 발전함에 따라 구성이 변하기도 한다. 기본적인 구조를 구성하는 블록의 주요 기능은 다음과 같다. 거시블록에서 주요 거시경제변수인 국내총생산, 총소득 및 총소비, 조세수입이 결정되며, 물가인 GDP디플레이터, 전산업 임금 및 실업률이 결정된다. 수요·공급 블록은 산업별 소비, 투자 등 최종수요와 함께 산업별 중간수요가 유도되어 총산출을 결정한다. 가격·기술 블록은 수요·공급 블록에서 유도된 산업별 노동생산성과 거시블록에서 결정된 전산업 임금을 매개로 산업별 임금, 자본재 가격, 중간투입재가격, 생산가격, 수출가격 등이 결정되며 산업별 중간투입계수, 자본계수, 노동계수가 유도된다. 대외블록은 소규모개방경제 하에서 외생적으로 주어지는 해외의 산업별 생산 및 가격, 산업별 수출가격 등이 주요 설명변수로 작용하여 산업별 수출과 수입이 결정된다. 대외블록에서 결정된 변수는 수요·공급 블록에서 최종수요를 산출하는데 이용된다.<sup>20)</sup>

19) KIET 산업-거시경제 단기 계량전망모형 개발 연구 발표자료, p. 6.을 바탕으로 저자정리

20) 이진면 외(2007), p. 74.

**산업연관 모형**

**수요·공급목록**

- 산업별 민간소비
- 산업별 중간투입·중간수요
- 산업별 노동수요
- 산업별 투자·자본소통
- 산업별 부가가치생산
- 산업별 산출

**가시목록**

- 산업별 중간재 가격
- 산업별 임금, 사용자비용
- 산업별 자본재 가격
- 산업별 총산출 가격
- 산업별 중간투입계수
- 산업별 노동계수
- 산업별 자본계수

**대외목록**

- 산업별 가격
- 산업별 수출
- 산업별 수입

**가시목록 (속도·물가)**

- 총소비(민간, 정부)
- 총투자, 총수출, 총수입
- GDP, 잠재GDP
- 디플레이터(GDP, 소비, 투자, 수출, 수입, 산출)
- 전산업 임금

**가시목록 (재정·금융)**

- 재정수지
- 조세수입
- 이자율

**민간 소비**

- 민간 소비

**소비목록**

- 비목별 소비지출
- 민간소비 비목별 가격
- 소비비목과 산업 간 전환행렬

**산업연관 모형(78부문 생산유발, 부가가치 유발)**

산업연구원의 산업·거시계량모형은 1990년대 초에 KDI에서 개발한 다 부문모형을 시작으로 동 모형을 2000년대 초까지 활용하였다. 2007년에 산업 부문 확대와 생산, 소비, 투자를 도입하면서 모형의 이름을 KIET-DIMM으로 명명하였다. 이후, 고령화 분석을 위한 인구가계 블록을 추가한 KIET-DIMM12을 개발하였으며 4차 산업혁명의 파급효과 분석을 주요 목적으로 한 KIET-DIMM18을 개발하였다.

모형	분석대상	내용	분석 시계
KIET-DIMM07 (KIET 산업·거시 계량모형)	국민계정의 78개 산업 중 농 림어업을 1개의 산업으로 통 합하고, 기초통계가 부족한 서비스 부문을 대분류 산업으 로 통합한 39개 산업	• 정책 시뮬레이션을 위해 기 존의 KDI 모형에 산업부문 확대, 생산, 소비, 투자를 도 입	연간



모형	분석대상	내용	분석 시계
KIET 단기 거시·산업 전망모형	KIET의 반기별 산업전망 분류인 10대 제조업을 포함한 17개 산업	<ul style="list-style-type: none"> <li>• 국내외 경제변수의 단기적 영향을 살피기 위해 분기 데이터베이스를 구축</li> <li>• 대외블록 확장</li> </ul>	분기
KIET-DIMM12	한국은행 국민계정 및 산업연관표의 분류기준에서 대분류 수준으로 통합한 78개 산업	<ul style="list-style-type: none"> <li>• 고령화 분석을 위해 인구·가계 블록을 추가</li> <li>• 소비블록 확장</li> </ul>	연간
KIET-DIMM18	KIET-DIMM12와 동일한 78개 산업	<ul style="list-style-type: none"> <li>• 4차산업혁명 파급효과 분석을 위해 시나리오를 다양화</li> </ul>	연간

자료: 장정인 외(2021a), p. 18

### 3. 에너지경제연구원

에너지경제연구원의 에너지정보통계센터는 에너지 데이터를 생산하고, 에너지 수요를 전망하는 작업을 하고 있다. 전망은 단기 및 중기 전망모형과 장기 전망모형으로 이루어져 있다. 단기 전망 모형의 전망 시계는 1년에서 1.5년이며 연간 2회 전망하여 발간한다. 중기 전망의 경우 5년의 시계를 전망하여 연간 1회 전망하여 발간한다. 단기 전망 모형은 ARDL(Autoregressive distributed lags) 모형을 활용한다. 단기 전망은 에너지 산업 정책 및 정부정책을 위한 기초자료로 활용되며, 세수추계를 위한 자료로도 활용된다. 한편, 장기 전망 모형의 시계는 향후 20년 이상이며 연간 1회 전망해서 발표한다. 전망모형으로는 KEEI-EGMS(Energy and Greenhouse gas Modelling System)을 사용한다. 장기 전망은 에너지 기본계획, 전력수급 기본계획, 장기 천연가스 수급계획 등의 우리나라의 에너지 계획 수립의 기초자료로 활용된다.<sup>21)</sup>

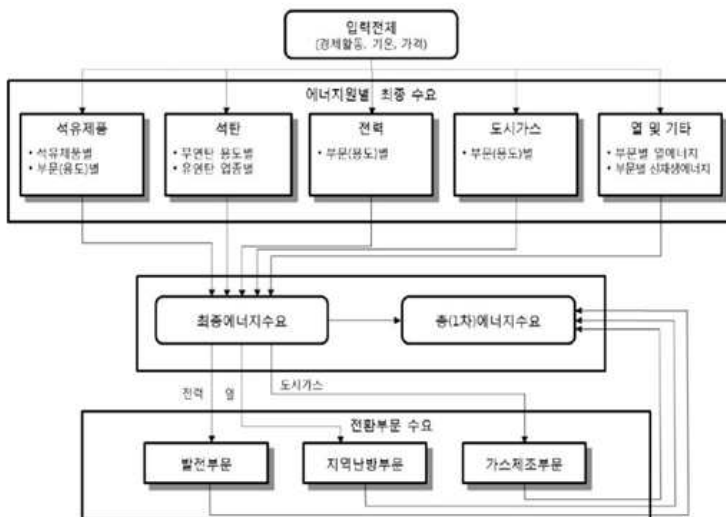
21) 에너지 수급전망 전문가 세미나(KEEI) 발표자료(2021.7.13.), p. 6.을 바탕으로 저자 정리

## 1) 단기 전망 모형

### (1) 모형 구조

단기 전망 모형의 기본 구조는 에너지원별·부문별 에너지수요를 개별적으로 전망하고 최종 에너지수요는 개별적으로 전망한 에너지원별·부문별 합계로 추정한다. 전환부문 수요를 추정한 후 최종 에너지수요에 더해서 총(1차)에너지수요를 전망한다. 에너지원별로는 석유, 석탄, 전력, 도시가스, 열에너지 및 신재생에너지·기타로 이루어지며, 부문별로는 산업, 수송, 가정, 상업, 공공부문으로 나누어진다.

〈그림 2-6〉 단기 전망 모형 구조



자료: 에너지 수급전망 전문가 세미나(KEEI) 발표자료(2021.7.13.), p. 8.

## (2) 최종 소비 부문 모형: ARDL 모형

최종 소비 부문의 모형은 앞서 언급한 ARDL 모형을 활용한다. ARDL 모형은 종속 변수와 독립변수의 시차 변수를 모형에 추가해서 추정해서 모형의 안정성(robustness)을 높일 수 있다. 단기 및 중기 전망에서는 추정치의 안정성이 매우 중요하므로 ARDL 모형을 기본모형으로 활용했다. 시차의 크기는 Henry의 general to specific 방법을 활용해서 결정했으며, 전년 동기(12)를 기준으로 하나씩 줄여나가는 방식을 활용했다. ARDL 모형은 추정식에 단위근이 존재하거나, 변수들의 공적분 관계가 있는 경우에도 유효한데, 에너지 수요를 추정을 위한 변수들이 단위근을 가지며, 추정식이 공적분관계에 있으므로 ARDL 모형의 적용이 적합하다.<sup>22)</sup> 또한 ARDL모형은 비교적 단순하고 탄력성을 쉽게 구할 수 있어서 단기 에너지 수요전망의 기본 모형으로 이용하기에 적합하다.<sup>23)</sup>

$$\ln(y_t) = \beta_0 + \beta_1 \ln(p_t) + \beta_2 \ln(HDD_t) + \beta_3 \ln(CDD_t) + \beta_4 \ln(GDP_t) + \sum_j \beta_5 \ln(y_{t-j}) + \sum_k \beta_{6k} S_k + u_t$$

$$u_t = \sum_l \rho_l u_{t-l} + \epsilon_t$$

$y$ : 에너지원별 • 부문별 수요,  $p$ : 실질가격변수,  $HDD$ : 난방도일,  $CDD$ : 냉방도일,  $GDP$ : 국내총생산

에너지원별·부문별 상황에 맞게 경제변수와 경제외적변수를 조정해서 모형을 추정하고 전망한다. 에너지수요에 영향을 미치는 경제변수로는 기본적으로 국내총생산(GDP), 산업생산지수, 에너지 가격이 있으며, 경제외적 변수로는 기상여건(기온, 냉난방도일) 등이 있다.<sup>24)</sup> GDP는 한국은행 및 KDI, 국회 예산정책처의 경제전망을 활용하며, 산업생산지수는 GDP를 이용해서 계산하고, 가격변수는 국제유가로부터 석유제품, 도시가스 가격을

22)나인강류지철(2000), pp.122~123.

23)에너지경제연구원(2018), p.67

24)에너지경제연구원(2018), p.66

계산한다. 기온변수는 전망기간 과거 10년의 일평균기온을 가정해서 냉난방도일을 계산해서 사용한다.

ARDL의 기본구조에서 에너지원과 부문별 특성에 맞게 식을 변형해서 추정하는데, 예를 들어 석유의 경우는 산업부문에서 6개의 석유제품(등유, 경유, 중유, LPG, 납사, 아스팔트), 수송부문에서 5개의 석유제품(휘발유, 경유, 중유, 항공유, LPG), 건물 부문에서 4개의 석유제품(등유, 경유, 중유, LPG)로 총 15개의 식을 추정한다.

〈표 2-2〉 에너지원별 전망 구조

에너지원	부문	추정식수
석유	산업, 수송, 건물	15
전력	산업, 수송, 가정, 상업	4
가스	산업, 가정, 상업, 수송, 공공	4
석탄	산업, 가정, 상업	5
열에너지 및 신재생·기타	산업, 가정, 상업, 수송, 공공	6
전환부문	발전, 열생산, 가스제조	LP(선형계획법)

자료: 에너지경제연구원(2018), p.67

## (2) 전환부문 모형

전환부문의 에너지 수요는 각 모형에서 전망한 최종 에너지수요 중 전환 단계를 거쳐서 소비되는 에너지인 전력, 열에너지, 도시가스에 투입되는 에너지 수요를 산출하는 단계이다.<sup>25)</sup> 전력, 열에너지, 도시가스 등의 이차 에너지수요를 생산은 발전, 열에너지 생산연료(지역난방), 도시가스 제조 부문별로 나누어서 산출한다. 전력 생산에 필요한 연료투입량은 다음의 단계로 산출된다. 먼저 자기소비와 송배전 손실을 고려해서 총발전량을 산출하

25) 나인강·류지철(2000), p. 159

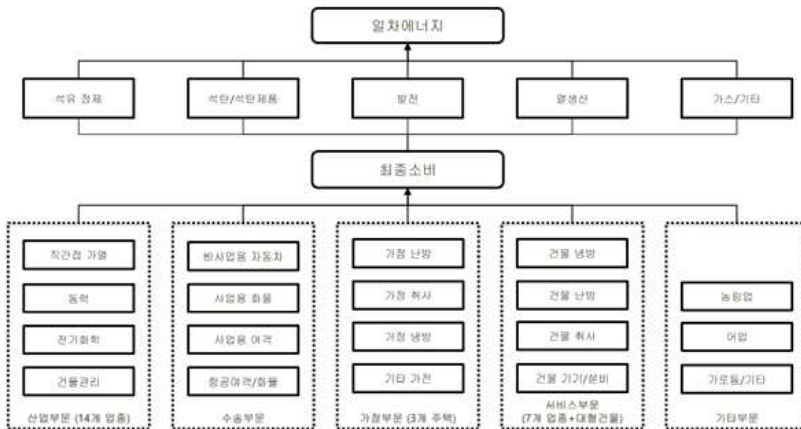
고, 선형계획법(LP: Linear Programming) 모형을 적용해서 총발전량을 충족하는 원별 발전량을 전망한다. 원별 발전량 전망치에 발전효율을 적용해서 최종적으로 연료투입량을 산출한다.<sup>26)</sup> 마지막으로 총에너지는 전환 부문의 소요에너지를 추정해서 최종에너지를 합해서 전망된다.

## 2) 장기 전망 모형

### (1) 모형 구조

20년 이상의 장기에 국가 에너지 수요 및 에너지 부문 온실가스 배출 전망을 목적으로 에너지경제연구원은 장기 전망 모형인 KEEI-EGMS 모형을 운영하고 있다. 부문별 모듈로 구성된 미시적·구조적 전망모형이다.

〈그림 2-7〉 KEEI-EGMS의 구조



자료: 에너지 수급전망 전문가 세미나(KEEI) 발표자료(2021.7.13.), p. 14.

26) 에너지경제연구원(2018), p. 66

---

최종 소비부문(산업, 가정, 서비스, 수송)은 최종에너지를 소비하는 최종 소비자로 구성되며, 각 부문의 용도별·에너지원별 최종에너지 수요를 결정한다. 전환부문은 최종에너지인 석유제품, 석탄제품, 전력, 열에너지, 도시가스를 생산하는데 투입되는 연료 및 원료를 계산한다. 일차에너지 수요는 최종에너지와 전환 순투입, 자가소비, 유통손실의 합계로 계산된다. 부문별 에너지 수요 전망을 위해 경제성장률, 통계청의 인구수 및 가구수 자료, 에너지경제연구원, EIA, IEA 등의 국제유가 자료, 산업연구원의 업종별 성장률 등을 외생변수를 사용한다.

## 4. 대외경제정책연구원

### (1) 세계경제전망 연혁

세계경제전망은 EC통합 및 세계경제의 블록화현상 등 1990년대 가속화된 국제경제 환경변화에 대응하기 위한 정책보고서에서 출발하였다.<sup>27)</sup> 이후 IMF와 와튼 계량경제연구소(WEFA) 등 타기관의 세계 및 각 지역의 성장률 전망치를 인용하고, 환경변화에 따른 정책대응방안을 제시하였다. 1994년부터는 세계경제전망을 명시한 보고서가 발간되었고<sup>28)</sup>, 금융·자본 시장의 개방 압력이 높아짐에 따라 국제금리, 환율, 유가 전망을 포함하였다. IMF, OECD, World Bank, 중국국가통계국 등 다양한 기관의 전망치를 인용하였다. 1999년부터 2006년까지는 월간 『KIEP 세계경제』의 12월 특별호로 세계경제전망 보고서가 발간되었다.<sup>29)</sup> 이때부터는 정책보고서로서의 성격이 약화되고 현황, 전망 및 이슈 중심의 전망보고서로 전환

---

27) 「1990년대 국제 경제 환경 변화와 정책대응」(1990), 「1993년도 국제 경제 환경 변화전망과 대응전략」(1992)

28) 「1994년도 세계경제 전망과 대외경제정책과제」(1994), 동 보고서는 1997년까지 연속 발간되었다.

29) 「1999년 세계경제전망」, ~ 「2006년 세계경제전망과 한국의 대외경제정책방향」

되었다. 2007년 이후에는 국제금융시장, 국제통상환경, 세계지역경제 등의 3개의 발간물로 구성된 「KIEP 오늘의 세계경제」로 발간되었다. 2012년 전망부터 KIEP 자체 전망치를 발표하기 시작했으며 2013년 전망부터 연중에 수정전망 발간을 추가하여 연2회로 발간되고 있다.<sup>30)31)</sup> 세계경제 전망 보고서의 총괄은 KIEP 국제거시금융실이 하고 있으며, 무역통상 및 지역연구자 등 약 35명이 연구진으로 참여하고 있다.

## (2) 분석모형

대외경제정책연구원은 전망작업의 최종 목적에 따라서 다양한 모델을 사용하고 있다. 먼저 세계경제전망을 수행하는 데에는 IMF의 Oxford Global Economic Model(OGEM)을 활용한다. OGEM은 전망에 필요한 세계 각국의 기본 자료를 자동 업데이트 하고 있으며, 세계 각국의 주요 변수 사이에 선형관계를 가정하여 경제이론을 기반으로 모형화했다. OGEM을 세계경제전망의 기본 모형으로 사용하고 있으며, 각 지역의 성장률 전망을 반영하여 수출입을 포함한 국민계정의 주요항목을 조정해서 사용한다. 한편 시나리오분석을 위해서는 IMF의 GIMF(Global Integrated Monetary and Fiscal Model) 수정모형을 활용한다. GIMF 모형은 합리적 기대를 가정한 DSGE 접근의 글로벌 모형이다. GIMF 모형에 우리나라의 상황을 반영해서 모형을 수정해서 사용하고 있다. 분석 시나리오 구성을 위해서는 VAR모형 등을 필요에 따라 개발해서 사용한다. 한편 대외 충격에 따른 산업별, 품목별 영향을 파악하기 위해서 CGE 모형을 사용하고 있다.<sup>32)</sup>

30) 매년 11월에 익년 세계경제 전망보고서를 발표하며, 5월에 수정전망을 반영한 업데이트 보고서를 발표해서 연간 2회 발표된다.

31) KMI-KIEP 업무 교류협력 협의회 발표자료(2021.6.24.), pp. 7~8.

32) KMI-KIEP 업무 교류협력 협의회 발표자료(2021.6.24.), p.17.

---

## 제2절 소결

---

본 장에서는 한국은행을 비롯한 우리나라 주요기관의 경제전망시스템에 대해서 알아보았다. 기관마다 전망에 사용하는 방법론에는 차이가 있었지만 전망 시계 및 분석 목적에 따라 다양한 방법론을 적용하고 있음을 확인할 수 있었다. 특히 한국은행은 전망 시계에 따라 초단기, 분기, 연간 모형을 운용하고 있으며 분석 목적에 따라서 마코프전환모형, VAR모형을 적용하는 다모형 접근법을 사용하고 있다.

지금까지 주요기관 전망모형의 특징과 장단점에 대해 살펴보았다. 해양수산업 전망모형 개발을 위해 고려하여야 할 사항은 다음과 같다. 첫째, 해양수산업 전망을 위해선 분석 대상의 수준이 산업이어야 할 것이다. 거시계량모형은 총량변수를 대상으로 하기 때문에 보다 거시적인 국민경제 수준의 전망을 수행한다. 해양수산업 대분류를 기준으로 보았을 때 이러한 거시모형은 전망에 사용하기 적합하지 않다. 산업 수준의 경제활동을 전망하는 모형은 시계열 모형과 산업·거시계량모형을 들 수 있다. 산업·거시계량모형은 다양한 연립방정식 체계를 구성하여 이론적 배경에 근거한 전망을 수행할 수 있다는 장점이 있다. 한편 시계열 모형은 변수 간의 유기적 관계를 바탕으로 전망을 수행하고 높은 예측력을 지녔다는 장점이 있다.

둘째, 전망의 목적에 따라 모형을 달리 적용하는 다모형 접근법을 활용해야 할 것이다. 구조모형을 핵심모형으로 설정하되 전망 시계, 분석 목적에 따른 보조모형을 활용하여 상호보완적인 전망체계를 구축해야 할 것이다. 특히 예측력 향상이 전망의 목적일 경우엔 가용 데이터의 모든 정보를 활용하여 예측하는 머신러닝 기법의 적용이 유용할 것이다.

마지막으로, 해양수산업 전망을 위해서는 데이터베이스 구축이 선행되어



야 할 것이다. 현재 해양수산업 데이터베이스는 한국은행의 산업연관표를 기준으로 작성되고 있다. 해양수산업 데이터베이스 구축을 통해 우리나라 국민경제에서 해양수산업의 위상을 파악하고, 현실 설명력 높은 전망치를 시산할 수 있을 것이다.



## 03

# 해양수산업 데이터베이스 구축

## 제1절 해양수산업 DB 구성

### 1. 해양수산업 DB 활용 통계<sup>33)</sup>

해양수산 전망 데이터베이스는 한국은행의 투입산출표(Input-Output table)를 기반으로 하여 작성되며, 한국은행 국민계정의 부가가치 및 산출 데이터와 정합성을 확보하도록 구축되었다. 한국은행 국민계정의 최종수요(민간소비지출, 정부소비지출, 고정자본형성, 수출, 수입), 순생산물세 등의 데이터와 시계열적인 정합성을 확보하는 작성을 수행하여 구축된다. 고용 데이터는 한국은행의 산업연관표의 부속표인 고용표를 활용하여 취업자수를 도출하였으며, 자본스톡은 한국은행의 국민계정의 순자본스톡 자료와 한국생산성본부의 실질자본스톡 자료를 활용해서 구축했다.

#### 1) 산업연관표

33) 본 연구는 장정인 외(2020) 해양수산업 전망모형 개발연구(Ⅰ)와 장정인 외(2021a) 해양수산업 전망 모형 개발연구(Ⅱ)의 후속연구임. 연차별로 진행되는 연구의 특성상 기존에 구축된 DB의 기본 틀은 유지 하되 수정·보완해서 사용하였으며, 시계열을 연장하여 활용했음. 따라서 본 절의 내용은 장정인 외(2021a) 제3장(pp.29~39)의 내용을 바탕으로 일부 내용을 수정·보완해서 작성하였음

한국은행은 1964년에 최초로 기준연도 1960년의 산업연관표(실측표)를 작성해서 공포했다. 이후 2~3년 간격으로 불규칙적으로 산업연관표가 작성되어 발표되었으며, 연속시계열로 발표되기 시작한 것은 2005년 실측표부터이다. 2005년 실측표부터 매 5년(끝자리가 0, 5로 끝나는 연도)마다 실측표가 작성되어 발표되고 있으며, 그 외의 연도에는 연장표를 작성해서 발표하고 있다. 최근 발표된 연장표는 기준연도 2019년의 산업연관표이다. 현재까지 발표된 산업연관표의 최장 연속 시계열은 2005~2019년이다. 전망 DB의 시계열은 연속성이 담보되는 2005~2019년까지 구축했다.<sup>34)</sup>

〈표 3-1〉 한국은행 산업연관표 현황(1960~2019)

기준 연도	실측/연장	기본 부문	가격기준	기준 연도	실측/연장	기본 부문	가격기준
1960	실측	109	생산자	2005	실측	403	구매자/생산자/기초
1963	실측	109	생산자	2006	연장	403	생산자/기초
1966	실측	117	생산자	2007	연장	403	생산자/기초
1968	연장	117	생산자	2008	연장	403	생산자/기초
1970	실측	153	생산자	2009	연장	403	생산자/기초
1973	연장	153	생산자	2010	실측	384	구매자/생산자/기초
1975	실측	392	생산자	2011	연장	384	구매자/기초
1978	연장	392	생산자	2012	연장	384	구매자/기초
1980	실측	396	생산자	2013	연장	384	구매자/기초
1983	연장	396	생산자	2014	연장	384	구매자/기초
1985	실측	402	생산자	2015	실측	381	구매자/생산자/기초
1986	연장	402	생산자	2016	연장	381	생산자
1987	연장	402	생산자	2017	연장	381	생산자
1988	연장	402	생산자	2018	연장	381	생산자

34) 후속 연구에서 연속시계열을 아니지만, 시계열 연장을 위해 2005년 이전 연도에 대한 시계열을 순차적으로 업데이트하는 작업을 계획하고 있다.

기준 연도	실측/ 연장	기본 부문	가격기준	기준 연도	실측/ 연장	기본 부문	가격기준
1990	실측	405	구매자/생산자	2019	연장	381	생산자
1993	연장	405	생산자				
1995	실측	402	구매자/생산자				
1998	연장	402	생산자				
2000	실측	404	구매자/생산자				
2003	실측	404	생산자/기초				

주: 실측표의 경우 음영으로 표시함

자료: 한국은행경제통계시스템(ECOS)(검색일:2021.12.15.); 장정인 외(2021a), pp. 30~31 일부 참조,

## 2) 국민계정

한국은행의 국민경제 통계는 산업연관표의 총액을 국민계정에서 발표되는 수치와 정합성 있게 조정하고, 데이터를 최근 연도까지 업데이트하는데 활용한다. 국민계정 통계에서 활용하는 자료는 최종수요 관련 자료(민간정부소비지출, 고정자본형성, 수출입), 총산출, 부가가치 및 부가가치 상세내역(피용자보수, 영업잉여, 고정자본소모, 생산세)자료와 순생산물세, 국내총생산, 순자본스톡 등의 자료이다. 최종수요와 부가가치관련 자료는 일반적으로 시차가 1년 존재하여 2021년 기준으로 2020년 자료까지 확보가능하며, 총산출과 순자본스톡은 시차가 2년 존재하여 2021년 기준으로 2019년 자료까지 확보 가능하다. 국민계정에서 확보한 가장 최근 시계열의 총액자료를 바탕으로 산업연관표의 총액을 업데이트해서 활용한다. 각 부분의 총액을 국민계정의 총액으로 조정함으로써 국민계정과와의 정합성을 확보한다.

〈표 3-2〉 국민계정 활용 통계 및 시계열

구분	항목	명목/실질	연도	비고
최종수요	민간소비지출	명목/실질	1970~2020	총액
	정부소비지출	명목/실질	1970~2020	총액
	고정자본형성	명목/실질	1970~2020	총액
	수출	명목/실질	1970~2020	총액
	수입	명목/실질	1970~2020	총액
총산출		명목	1970~2019	27부문(77부문)*
부가가치	총부가가치 (기초가격)	명목/실질	1970~2020	27부문/39부문 (77부문)*
	피용자보수	명목	1970~2020	27부문
	영업잉여	명목	1970~2020	27부문
	고정자본소모	명목	1970~2020	27부문
	생산세(보조금 공제)	명목	1970~2020	27부문
순생산물세		명목/실질	1970~2020	총액
국내총생산(GDP)		명목/실질	1970~2020	총액
순자본스톡		명목/실질	1970~2019	31부문

주: \* 총산출과 총부가가치(명목/실질)의 경우 한국은행 경제통계시스템에서 제공되는 부문 보다 세분화된 분류(77부문)를 한국은행에서 연구용 목적으로 제한적으로 협조 받아 활용함  
 자료: 장정인 외(2021a), p. 32를 바탕으로 최근시계열을 업데이트함

### 3) 기타 통계

산업별 디플레이터는 한국은행 국민계정에서 경제활동별 국내총생산 디플레이터를 제공하고 있으나, 산업 부문이 15개로 제한되어 있다.<sup>35)</sup> 따라서 세부 산업별 디플레이터를 적용하기 위해서 한국은행의 협조로 제공받은 77부문 명목 및 실질 부가가치와 산출데이터를 활용해서 디플레이터를 산출해서 사용한다.<sup>36)</sup> 다음으로 산업별 자본스톡을 도출하기 위해서는 국민계정의 순자본스톡(31부문) 자료를 기본으로 사용하되, 산업별 세분화를

35) 15개 산업 부문은 농림어업, 광업, 제조업, 전기, 가스 및 수도사업, 건설업, 도소매 및 숙박음식업, 운수업, 금융 및 보험업, 부동산업, 정보통신업, 사업서비스업, 공공행정, 국방 및 사회보장, 교육서비스업, 의료, 보건업 및 사회복지서비스업, 문화 및 기타서비스업이다.

36) 디플레이터= 명목/실질\*100

위해서 한국생산성본부의 실질자본스톡 자료를 활용하였다. 산업별 투자를 도출하기 위해서는 한국은행의 고정자본형성표(2005년, 2010년)를 활용하여 산업별 고정자본 전환행렬을 도출했다. 마지막으로 해양수산부문을 산업연관표의 기본부문으로부터 식별하기 위해서 통계청의 도소매업조사, 서비스업조사, 해양수산부의 해양수산업통계조사 등의 다양한 통계를 활용하고 있다.<sup>37)</sup>

〈표 3-3〉 기타 활용 통계 및 시계열

구분	출처	연도	부문수
디플레이터	한국은행 (명목 및 실질 부가가치, 총산출)	2000~2020	77부문*
실질자본스톡	한국생산성본부	1970~2012 (2000년 기준)	72부문
		1995~2018 (2010년 기준)	38부문
투자 (고정자본 전환행렬)	한국은행 (고정자본형성표)	2005, 2010	95 X 78(2005) 96 X 82(2010)
취업자수, 피용자수	한국은행 (고용표)	2005~2019**	168부문
해양수산 비중산출	통계청 해양수산부 등	2005~2019	산업연관표 기본부문 기준으로 비중 산출

주:1)\* 한국은행의 협조를 받아 연구목적으로 받은 부가가치, 총산출 자료 활용

2)\*\* 고용표는 1975년부터 발표되고 있으나, 전망 DB를 위해 활용한 시계열은 2005~2019년임  
자료: 저자 작성

## 2. 해양수산업 DB 산업분류

전망 DB는 시계열의 연속성이 담보되는 2005~2019년까지 구축했다. 전망 DB구축의 가장 기본적인 자료인 한국은행 산업연관표 중 투입산출표

37) 비중도출방법은 장정인 외(2021b)의 제2장 제2절을 참고. 다만 전망 DB는 총거래표를 활용하므로 국산표를 활용하는 장정인 외(2021b)에서 적용하는 비중의 수치와는 다소 차이가 있음

는 실측표가 발표되는 5년 단위로 부문분류가 조정이 된다. 따라서 2005년, 2010년, 2015년의 부문분류의 변화를 반영해서 일관성 있게 일부 부문을 통합 조정해야한다. 2015년의 산업연관표의 부문분류를 기준으로 통합조정을 위해 2005~2010년 부문연계표와 2010~2015년 부문연계표를 활용했다. 각 년도의 기본부문을 기준으로 변동사항을 반영해서 연계하되, 국민계정의 77부문과 연계해서 부문간 조정이 필요한 부문을 통합해서 총 65개로 통합하였다. 최종적으로 2015년의 대분류(33개)를 기준으로 기타 제조업제품(C13)과 제조임가공 및 산업용 장비수리(C14)부문을 하나의 부문으로 통합했고, 기타서비스(S)와 기타(T)부문을 하나의 부문으로 통합했다. 이에 따라 총 31개 부문으로 집계하였다. 해양수산업은 총 10개 부문으로 분류하였는데, 이는 2018년도에 통계청 승인을 받은 해양수산업 특수분류의 기준(대분류 9개)을 따르되 이중 해운항만업을 해운업과 항만업으로 분리했다.

〈표 3-4〉 전망 DB 구축을 위한 산업연관표 대분류

코드	대분류 산업명	비고	코드	대분류 산업명	비고
1	해양자원 개발 및 건설	해양 수산업	21	컴퓨터, 전자 및 광학기기	제조업
2	해운		22	전기장비	
3	항만		23	기계 및 장비	
4	선박 및 해양플랜트 건조 수리		24	운송장비	
5	수산물 생산		25	기타 제조업 제품 및 제조임가공 및 산업용 장비수리 (C13-C14통합)	
6	수산물 가공		26	전력, 가스 및 증기	전력, 가스·수도·폐기물 처리업
7	수산물 유통		27	수도, 폐기물 및 재활용서비스	
8	해양수산 레저관광		28	건설	건설업
9	해양수산 기자재 제조		29	도소매 및 상품중개서비스	서비스업



코드	대분류 산업명	비고	코드	대분류 산업명	비고
10	해양수산 관련 서비스		30	운송서비스	
11	농림품	농림축 산·광업	31	음식점 및 숙박서비스	
12	광산품		32	정보통신 및 방송 서비스	
13	음식료품		33	금융 및 보험 서비스	
14	섬유 및 가죽제품	제조업	34	부동산서비스	
15	목재 및 종이, 인쇄		35	전문, 과학 및 기술 서비스	
16	석탄 및 석유제품		36	사업지원 서비스	
17	화학제품	제조업	37	공공행정, 국방 및 사회보장	서비스업
18	비금속광물제품		38	교육서비스	
19	1차 금속제품		39	보건 및 사회복지서비스	
20	금속가공제품		40	예술, 스포츠 및 여가 관련 서비스	
			41	기타서비스 및 기타 (S-T 통합)	

주:1) 해양수산업은 해양수산업 특수분류의 9대 대분류를 따르되, 해운항만업을 해운업과 항만업으로 구분하여 총 10개 부문으로 함

2) 한국은행 산업연관표(2015년 기준년도)의 대분류 산업 구분에 따르되 C13.기타제조업제품과 C14.제조임가공 및 산업용 장비수리 부문을 하나의 부문으로 통합(25)하였고, S.기타서비스와 T.기타부문을 하나의 부문(41)으로 통합하여 31개 부문으로 구분하여 작성함

3) 연속된 시계열 확보를 위해 기존의 해양수산업 중심의 산업연관표의 분류에서 일부 부문을 조정하여 31개 부문으로 활용함

자료: 장정인 외(2021a), p. 36 자료 일부 수정

한편 산업별 자본스톡을 구하기 위해 한국은행의 순자본스톡(31부문)과 한국생산성본부의 실질자본스톡 자료를 통합하여 구성한 통합부문은 아래와 같이 총 47개 부문으로 구성하였다.

〈표 3-5〉 순자본스톡 통합 코드

코드	대분류 산업명	코드	대분류 산업명
1	농축산업 및 관련 서비스업	25	기타 운송장비 제조업
2	임업	26	기타 제조업 및 산업용 장비 수리업
3	어업	27	전기, 가스, 증기 및 공기조절 공급업
4	석탄, 원유 및 천연가스 채굴업	28	수도, 하수 및 폐기물 처리, 원료 재생업
5	금속광업	29	건설업
6	비금속광업	30	도소매업
7	음식료품 제조업	31	육상 운송업
8	담배 제조업	32	수상 운송업
9	섬유 및 가죽제품 제조업	33	항공 운송업
10	목재 및 목제품 제조업	34	창고 및 기타 운송관련 서비스업
11	펄프 및 종이제품 제조업	35	숙박 및 음식점업
12	인쇄 및 기록매체 복제업	36	정보통신업
13	코크스 및 석유정제품 제조업	37	금융업
14	화학물질 및 화학제품 제조업	38	보험업
15	의료용 물질 및 의약품 제조업	39	금융 및 보험 관련 서비스업
16	고무 및 플라스틱 제조업	40	부동산업
17	비금속광물제품 제조업	41	전문, 과학 및 기술관련 서비스업
18	1차금속 제조업	42	사업 지원 서비스업
19	금속가공제품 제조업	43	공공행정, 국방 및 사회보장
20	컴퓨터, 전자 및 광학기기 제조업	44	교육서비스업
21	전기장비 제조업	45	의료, 보건업 및 사회복지서비스업
22	기계 및 장비 제조업	46	예술, 스포츠 및 여가관련 서비스업
23	자동차 및 트레일러 제조업	47	기타 서비스업
24	선박 및 보트 건조업		

자료: 저자 작성

### 3. 주요 해양수산업 DB 구성

해양수산 전망을 위한 주요 DB는 해양수산업 중심의 산업연관표와 산업

별 자본스톡, 산업별 투자부문으로 구성된다.<sup>38)</sup> 해양수산업 중심 산업연관표는 경상표와 실질표, 고용표로 구성된다. 먼저 경상표는 한국은행에서 제공하는 산업연관표(기본부문)를 바탕으로 국민계정의 총액 통계(총산출, 부가가치, 최종수요부문 등)를 반영해서 구축한다. 2005~2019년도의 부문분류를 통합조정된 DB를 구성한다. 해양수산업 부문을 식별하고 비중을 도출해서 적용해서 10개의 해양수산업이 포함된 해양수산업 중심의 산업연관표(경상표, 41×41부문)를 도출한다.

경상표가 도출되면 한국은행의 산업별 디플레이터를 활용해서 경상표를 실질화한다. 취업자수와 피용자수는 한국은행 고용표를 바탕으로 해양수산업 중심 산업연관표(경상표)의 산출액을 기준을 세분화하여 해양수산업부문을 식별한 후 41개 부문으로 통합한다.<sup>39)</sup>

자본스톡의 경우 한국은행의 국민대차대조표 상의 순자본스톡(31개 부문) 자료를 바탕으로 세분화를 위해서 한국생산성본부의 실질자본스톡 자료를 활용해서 총 47개 부분으로 통합해서 사용한다. 투자는 한국은행의 고정자본형성표(2005, 2010년)를 활용해서 전행렬을 작성해서 산출했다.

38) DB구축 결과에 대한 상세한 내용은 장정인 외(2020), pp.36~44의 내용을 참고

39) 고용표는 소분류(168부문)기준으로 발표되고 있으므로 기본부문(381부문)보다 부문수가 적다. 해양수산업 부문을 식별할 때, 취업계수(산출액 10억 원 당 취업자수)가 세부산업별로 안정적이라는 가정 하에 산출액을 기준으로 세분화한다.

〈표 3-6〉 주요 해양수산업 DB의 구성

구분	통계자료	출처	연도	작성 DB
해양수 산업 중심 산업 연관표	산업연관표 기본부문 (생산자가격, 총거래표)	한국은행	2005~ 2019	해양수산업 중심 산업연관표 (41×41부문) 경상표 및 실질표, 고용표
	산업별 디플레이터(총산출)	한국은행	2005~2018	
	산업별 디플레이터(부가가치)	한국은행	2005~2019	
	국민계정 통계	한국은행	2005~2020	
	고용표	한국은행	2005~2019	
	해양수산업 부문 비중산출 (도소매업 조사, 경제총조사, 해양수산업 통계조사 등)	통계청, 해양수산부 등	2005~2019	
자본 스톡	국민대차대조표 (순자본스톡)	한국은행	2005~2018	해양수산업 중심 실질자본스톡 (41×41부문)
	실질자본스톡(72부문)	한국생산성본 부		
투자	고정자본형성표	한국은행	2005, 2010	해양수산업 중심 고정자본형성행렬 (40×40부문)

주: 해양수산업 부문 비중산출을 위해 활용한 통계의 상세내용은 장정인 외(2021b)의 제2장 제2절을 참고  
자료: 장정인 외(2021a), pp. 38~39를 바탕으로 일부 수정

## 2절 해양수산업 DB 자동화

### 1. 기존 데이터베이스 구축 방법

해양수산업 데이터베이스는 기본적으로 한국은행이 제공하는 산업연관표(기본부문 분류)에서 해양수산업을 식별하여 구축된다. 각 기본부문 분류에서 해양수산업을 식별한 후 국민계정과의 통계와 일치시켜 정합성 과정을 거친다. 이후, 기본부문 분류 중 해양수산업 비중, 취업

자/비용자수, 디스플레이터 및 자본스톡 자료를 구한 후 기본부문 분류에 대응시켜 해양수산업에 해당하는 수치를 도출한다. 이러한 과정을 거쳐 10개의 해양수산부문과 31개의 대분류로 구성된 해양수산업 중심 산업연관표로 재구성하게 된다.<sup>40)</sup>

기존에는 상기의 해양수산업 데이터베이스 구축과정이 여러 엑셀과 일을 기반으로 하여 이루어졌다. 엑셀을 활용한 작업은 파일열기, 자료의 복사, 붙여넣기 등 연속적인 수작업이 수반되어 많은 시간이 소요되었다.

〈그림 3-1〉 해양수산업 중심 산업연관표 구축 엑셀 작업

자료 : 장정인 외(2020) p. 41 참조 재작성

이에 따라, 해양수산업 부가가치 전망을 위해서 데이터베이스 구축 과정이 상당한 비중을 차지하여, 해양수산업 부가가치 전망 방법론 및 결과 분석 등의 작업에 적절하게 자원이 분배되지 못하는 경우가 발생하였고 해양수산업 데이터베이스 구축의 자동화 필요성이 제기되었다.

40) 기존의 데이터베이스 구축에 대한 보다 상세한 설명은 장정인 외(2020) 참고

## 2. 개선된 데이터베이스 구축 방법

해양수산업 데이터베이스 구축 자동화는 기본적으로 Input-Output 형식의 구조로 이루어진다. 즉, 필요한 자료를 입력하면 해양수산업 산업별 부가가치 관련 데이터베이스가 즉시 도출되는 것이다. 이를 위하여 기존의 엑셀관련 수작업을 코딩하여 새로운 프로그램을 만들었다(사용 소프트웨어 : R). 이 프로그램에 입력하는 파일 목록은 아래와 같다.

〈표 3-7〉 데이터베이스 구축 자동화 입력 파일

입력파일	출처
산업연관표	한국은행
명목/실질 국민계정 항목	
총산출 및 부가가치	
고용표 자료	
전환행렬	
자본스톡 자료	한국은행, 한국 생산성본부
기본부문 코드 매핑	연구진
해양수산업 비중	해양수산업 통계조사 경제총조사 등 활용해서 연구진 도출

자료 : 저자 작성

한국은행에서 얻을 수 있는 자료들은 한국은행 경제통계시스템(Economic Statistics System, ECOS)에 업데이트되는 시점이 상이하다. 본 연구에서는 최신자료는 반영하되, 업데이트가 늦는 자료의 경우 최근 자료들을 활용하여 사용하였다. 산업 분류 상 기본부문을 해양수산업 세부 산업으로 매핑(Mapping)하는 작업은 2015년부터 사용하고 있는 381개 기본부문을 기준으로 한 자료를 사용하였다. 향후 기본부문 체계가 변경될 시에는 이를 반영하는 추가적인 프로그래밍 보완 작업이 필요할 것이다.

상기 입력 파일들을 프로그램에 입력하여 실행시키면 해양수산업 산업별

부가가치 자료(실질/명목)가 아래와 예시와 같이 도출된다. 이 프로그램을 사용하면 필요한 파일들만 준비하여 입력하면 되므로 향후 관련 연구에서 활용성이 높을 것이며, 해양수산업 부가가치 관련 연구의 핵심적인 부분에 더 많은 자원을 집중할 수 있을 것으로 기대된다. 물론, 산업연관표 체계 변경, 기본 부문 코드 변경 등의 중요한 사항이 발견되면 프로그래밍의 수정 및 보완이 필요할 것이다.

〈그림 3-2〉 데이터베이스 구축 자동화 결과 파일(예시)

연도	구분	GDP	(실질)순자취업자수	(투자	고정자본총수출	수입		
2019	해양자원	947039.7	669326.5	11711.05	122732	2089240	57025.63	33921659
2019	해운업	4061325	81337245	35172.4	4949683	2904.741	19955118	6933412
2019	선박 및 해	8294259	21332203	74088.43	5947557	7300558	27084134	5029348
2019	수산물 생	3001480	4009895	57992.24	407578.8	0	457478.9	1386462
2019	수산물 가	2011785	4192413	31182.92	432305.9	0	2280803	5726412
2019	수산물 유	3666077	3215946	79755.06	309473.2	309640.3	853972.6	90707.12
2019	해양수산	404786.6	2320054	8550.714	183414	73.2588	49262.05	219420.1
2019	해양수산	1326257	2401115	11932.08	524804.8	1096407	1641369	1832749
2019	해양수산	15918425	45607344	337223.9	3968601	3232079	807540.1	1308398
2019	항만업	3513221	8398722	48497.31	1640165	134806.5	2379116	4206052

자료 : 저자 작성





## 04

# 2021 해양수산업 전망모형

### 제1절 다모형 접근법

---

한국해양수산개발원은 2019년부터 해양수산업의 부가가치와 고용 등 주요 경제변수에 대한 경제예측시스템을 구축·활용해오고 있다. 해양수산업 시장참여자들과 정책결정자들에게 신뢰성 있는 전망을 제공하기 위해 해양수산업 경제통계 데이터베이스를 작성하여 우리나라 국민경제에서 해양수산업의 규모를 측정하고, 경제이론과 과거경험을 바탕으로 한 경제예측시스템을 구축하고 있다.

한국해양수산개발원의 경제예측시스템은 전망치의 신뢰도를 확보하고자 단일 모형에 의존하는 것이 아닌 다양한 형태의 모형을 개발하여 활용하는 다모형 접근법(multi-model approach)을 따르고 있다. 다모형 접근법은 분석 목적에 따라 다른 모형을 활용하는 방법으로, 분석의 목적이 이론적 설명가능성 또는 예측력 확보 인지에 따라 각기 다른 모형을 적용하여 경제전망을 수행하는 방법이다.<sup>41)</sup>

2021 해양수산업 전망은 시계열 모형인 베이지안(Bayesian) 벡터자기

---

41) 황상필 외(2005), p. 26.

---

회귀모형(VAR: Vector AutoRegssive)모형과 머신러닝기법을 활용하여 수행하였다. 베이지안 VAR모형은 다변량 시계열 자료를 분석하기 위해서 널리 사용하는 기법으로 단기예측에 높은 정확도를 보여주는 것으로 알려져 있다.<sup>42)</sup> 또한 베이지안 VAR은 사전분포(prior distribution)를 설정하여 일반적인 VAR모형이 지닌 과모수화(over-parameterized) 문제를 해결하고, 무엇보다도 시계열의 안정성(stationary)을 확보되어야 한다는 제약으로부터 자유롭다는 장점이 있다.<sup>43)</sup>

머신러닝기법은 해양수산업의 주요부문에 영향을 주는 모든 독립변수를 활용하기 위해 사용하였다. 해양수산업 데이터베이스를 통해 확보한 종속변수의 시계열은 2005년부터 2020년까지 총 16개인데 반해, 부문별 독립변수는 약 60여개이기 때문에 이와 같은 변수 구성을 적용하면 과소결정(underdetermined) 문제에 귀착하게 되어 일반적인 계량경제모형으로는 분석이 어렵다는 문제가 생긴다. 이러한 문제를 해결하기 위해 머신러닝기법을 적용하여 해양수산업 주요부문의 전망을 수행하였다.

## 1. 베이지안 VAR 모형

### 1) 벡터자기회귀모형<sup>44)</sup>

여러 변수사이의 복잡한 동적 상호관계(dynamic interrelationship)를 파악하여 예측에 활용하는 것은 경제전망에 중요한 이슈라 할 수 있다.<sup>45)</sup> 단순회귀모형은 변수 간의 상관관계를 종속변수에 대한 독립변수의 인과관계로 분석한다. 회귀모형에서의 종속변수와 독립변수의 관계는 경제 이론

---

42) 김윤영·박준용(2006), p. 1.

43) 노태영·최태련(2016), p. 1857., 김현석·장명희(2014), p. 80.

44) Lütkepohl (2005), pp.14 ~18.

45) 노태영·최태련(2016), p. 1857.

에 의해 구축되지만 변수들 간의 시간에 따른 변동성을 반영하는데 한계가 있다. Sims(1980)가 제안한 VAR모형은 시계열 변수 사이의 동적 상호관계를 설명하는 모형으로써, ARIMA와 같은 일변량(univariate)을 다변량(multivariate)으로 확장한 모형이다.<sup>46)</sup> p차수의 VAR모형인 VAR(p) 모형은 [식 4-1]과 같다.

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + C x_t + u_t, \quad u_t \sim N(0, \Sigma) \quad [\text{식 4-1}]$$

위 식에서,  $p$ 는 시차를 의미한다.  $A$ 는 내생변수의 계수로,  $K \times K$  행렬이다.  $C$ 는  $K \times m$  외생변수의 계수 행렬이다.  $u_t$ 는 교란항(error term)으로,  $K \times K$  공분산 행렬  $\Sigma$ 를 지닌  $K \times 1$  벡터를 의미한다.

VAR(p)모형은 [식 4-2]와 같은 동반형(companion form)으로 나타낼 수 있다.

$$Y = XB + U \quad [\text{식 4-2}]$$

단,

$$Y = \begin{pmatrix} y'_1 \\ \vdots \\ y'_T \end{pmatrix}, \quad X = \begin{pmatrix} y'_{01} y'_{01} & \cdots & y'_{1-p} x'_{11} \\ \vdots & & \vdots \\ y'_{T-1} y'_{T-2} & \cdots & y'_{T-p} x'_{T1} \end{pmatrix}, \quad B = \begin{pmatrix} A'_1 \\ \vdots \\ A'_p \\ C' \end{pmatrix}, \quad U = \begin{pmatrix} u'_1 \\ \vdots \\ u'_T \end{pmatrix},$$

[식 4-2]에서  $Y$ 는  $T \times k$  행렬이며,  $X$ 는  $T \times (Kp + m)$  행렬이다.  $B$ 는 모든 계수의 행렬로,  $(Kp + m) \times K$  행렬이다.  $U$ 는  $T \times K$  행렬이다.

$B$ 와  $\Sigma$ 의 최소제곱법(OLS: Ordinary Least Squares) 추정치는 다음과 같다.

$$\hat{B} = (X'X)^{-1} X'Y$$

46) 김현석·장명희(2015), p. 454.

$$\widehat{\Sigma}_{ols} = \widehat{U} \widehat{U}' / (T - Kp - m - 1), \quad \widehat{U} = Y - X\widehat{B}$$

[식 4-2]를 벡터화(vectorize)하면 [식 4-3]을 얻을 수 있다.

$$y = X^* \beta + u \quad \text{[식 4-3]}$$

위 식에서,  $y = \text{vec}(Y)$ 인  $KT \times 1$  행렬이며,  $X^* = I_K \otimes X$ 인  $KT \times K(Kp + m)$  행렬이다.<sup>47)</sup>  $\beta = \text{vec}(B)$ 로 모든 계수의  $K(Kp + m) \times 1$  벡터이다. 마지막으로,  $u = \text{vec}(U)$ 인  $KT \times 1$  교란벡터이며  $KT \times KT$  공분산 행렬  $\Sigma^* = \Sigma \otimes I_T$ 을 지닌다.

## 2) 베이지안 VAR

일반적인 VAR모형을 통해 [식 4-3]의 계수 행렬인  $\beta$ 를 추정한다면 종속변수의 수와 시차에 따라 추정해야 할 모수가 기하급수적으로 증가한다. 이러한 과모수화 문제를 해결하기 위해 베이지안 VAR은  $\beta$ 에 대한 알맞은 사전분포를 설정하여 VAR(p)모형을 축소한다.<sup>48)</sup>

베이지안 VAR에 사용되는 대표적인 사전분포로 자연 공액 사전분포(natural conjugate priors), 확산 사전분포(diffuse priors), 그리고 미네소타 사전분포(the Minnesota priors) 등이 있는데 본 연구에서는 미네소타 사전분포를 설정하였다.

미네소타 사전분포는 근사를 기반으로 하는 사전분포로서 공분산 행렬  $\Sigma^*$ 는 OLS의 추정치를 사용하고  $\beta$ 에 대한 사전분포만을 고려하는 방법이다. 이러한 특징으로 인해서 사후분포의 도출 과정이 단순하고 그 계산 시간이 빠르다는 장점이 있으나  $\Sigma^*$ 에 대한 불확실성이 제대로 반영되지 못한다는 단점도 존재한다.  $\beta$ 에 대한 미네소타 사전분포는 다음과 같다.<sup>49)</sup>

47)  $\otimes$ 는 크로네커 곱(Kronecker product)이고  $I_K$ 는  $K \times K$  단위행렬이다.

48) 노태영·최태련(2016), p. 1858.

$$\beta \sim N(\beta_0, \Omega_0)$$

사전분포의 초모수(hyperparameter)  $\beta_0$ 와  $\Omega_0$ 의 선택은 모형의 독립변수 형태에 따라 달라지는데, 일반적으로  $\beta_0$ 는 0 벡터로  $\Omega_0$ 는 대각행렬로 가정한다.

베이저안 VAR모형은 일반적인 VAR모형의 단점을 보완하고 다른 계량경제모형들에 비해 높은 예측력을 보이는 것으로 알려져 있어 해운 물동량, 주택시장 등 다양한 분야의 예측모형으로 사용되고 있다.<sup>50)</sup>

## 2. 머신러닝기법

### 1) 자료의 개요 및 특성

머신러닝 기법에서 사용되는 데이터는 각 산업별로 산업별 부가가치(실질)를 종속변수로 하며, 독립변수는 자본, 수출, 취업자수 등의 산업연관표 항목, 세계GDP, 회사채율 등의 거시경제변수 및 산업별 고유변수들로 구성된다. 데이터의 관측치 개수는 총 16개(2005년 ~ 2020년)로 독립변수의 개수(산업별로 약 60여개 내외)보다 적다. 이에 따라, 주어진 데이터를 사용하여 회귀분석 등을 시도하면 과소결정(underdetermined) 문제에 귀착하게 되며 일반적인 계량경제모델로는 분석이 불가능한 문제가 발생한다. 이점이 해당 데이터를 분석하는데 있어서 기존의 계량경제모델보다는 머신러닝 방법을 적용하는 것이 유리한 이유 중의 하나라고 볼 수 있다.

49) 노태영·최태련(2016), p. 1859.

50) 해운 물동량에 관한 연구(김현석·장명희 2014, 2015), 주택시장에 관한 연구(함종영·손재영 2016; 전해정 2019).

---

관측치 개수보다 독립변수의 개수가 큰 자료를 고차원 데이터(high dimensional data)라고 하는데, 현실 데이터들은 고차원 데이터인 경우가 많아 머신러닝 기법을 적용할 때에도 중요하게 다루어지는 부분이다. 고차원 데이터를 다루는 기본적인 방법은 종속변수에 대해 설명력이 높은 독립변수들만을 선별하는 방법(feature selection)이다. 여기서 기초적인 방법으로는 변수가 없는 상태에서 점차 변수 개수를 늘려가는 전진 선택법(forward selection), 모든 변수를 포함한 상태에서 변수 개수를 줄여가는 후진 소거법(backward elimination), 모든 변수를 포함한 상태에서 특정 통계치 이상(이하) 변수를 추가(삭제)하는 단계 선택법(stepwise selection) 등이 있다. 이러한 방법들은 직관적인 방법들이지만 변수가 많은 경우에는 변수 순서 설정이 결과에 영향을 주거나 장시간이 소요되는 등의 문제가 발생할 수 있다. 이에 따라, 본 연구는 비교적 최근의 머신러닝 기법을 적용한 변수 선택법을 활용하여 상기 문제들을 완화하고 종속변수를 보다 효율적으로 예측하는 독립변수들을 선별하고자 한다.

상기와 같이 변수들을 분석 전에 선별할 수도 있지만, 분석과정에서 독립변수들의 개수를 줄이는 효과를 가지는 머신러닝 기법들도 있다. 데이터를 학습하면서 영향력이 적은 독립변수의 계수를 감소하거나 0에 가깝게 만드는 방법인데 이러한 방법을 정규화(regularization)라고 한다. 본 연구에서는 정규화를 사용한 머신러닝 기법도 활용할 것이다. 또한, 관측치들 간의 근접성을 기반으로 하여 고차원 데이터를 분석하는데 유용한 일련의 머신러닝 기법들도 함께 사용할 것이다.

고차원 데이터를 사용한 예측 연구에서 발생할 수 있는 문제점은 예측값의 분산이 증가한다는 점이다. 일반적으로 관측치 개수가 많아질수록 예측값의 분산은 적어지는데, 관측치 개수는 독립변수 개수 대비 상

대적인 것이다. 그러므로 본 연구에서와 같이 관측치 개수가 독립변수 개수보다 상당히 적은 경우에는 예측치의 분산이 클 것이라고 예상할 수 있다. 예측치의 분산이 크다는 것은 다양한 머신러닝 기법을 사용할 시 예측성능이 일관적이지 않을 수 있다는 것을 내포한다. 즉, 모든 예측 시점에서 가장 예측력이 높은 머신러닝 기법이 관찰되기 보다는, 예측 시점별로 가장 예측력이 높은 머신러닝 기법이 상이할 것이라는 점이다. 이러한 문제를 완화하기 위하여 본 연구에서는 예측조합법(forecast combination) 기법을 사용할 것이다.

## 2) 변수 선별 기법

본 연구의 데이터와 같이 관측치 개수보다 독립변수 개수가 더 많은 고차원 데이터를 분석하기 위하여 변수 선별 기법이 유용하게 쓰일 수 있다. 본 연구에서는 랜덤포레스트 변수 중요도(random forest feature importance)와 지도 주성분 분석(supervised principal component analysis)에 기반하여 독립변수를 선별하고자 한다.

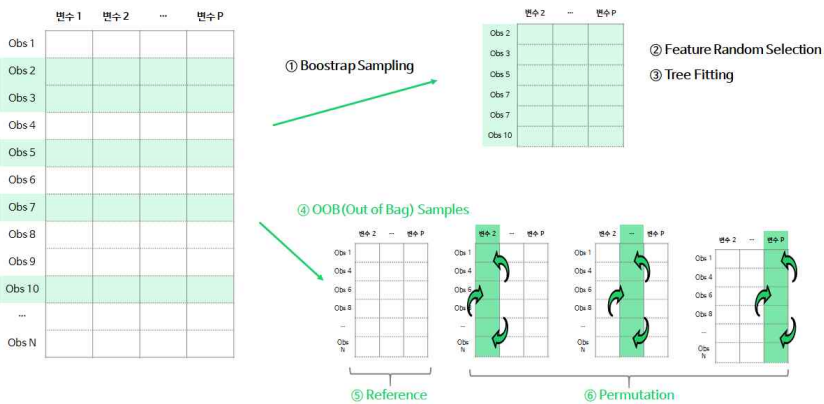
### (1) 랜덤포레스트 변수 중요도(Random forest feature importance)

랜덤포레스트(random forest)는 결정나무(decision tree)를 보다 일반화하여 학습시키는 머신러닝 기법이다<sup>51)</sup> 주어진 자료를 부트스트랩(Bootstrap) 기법으로 복제하여, 이 복제한 자료들 개수만큼 결정나무를 학습시키고 여기서 도출된 값들을 종합하여 최종 결과값을 정하는 것이 랜덤포레스트 기법으로 부트스트랩을 통하여 기존의 결정나무보다

51) Breiman(2001)

보다 일반화된 결과를 얻을 수 있는 장점이 있다. 랜덤포레스트 변수 중요도(Random forest feature importance)에 기반한 변수 선별 기법은 중요도 정의에 따라 몇 가지 기법으로 구분되지만 본 연구에서는 Breiman(2001)이 제안한 순열 중요도(permutation importance) 방법을 사용하기로 한다. 본 기법은 부트스트랩으로 복제된 자료에서 특정 변수의 순서를 무작위로 섞은 후 도출된 결과를 무작위로 섞지 않은 경우의 결과와 비교하는 방법으로, 무작위로 섞은 경우의 결과가 무작위로 섞지 않은 경우보다 낮게 나온다면 특정 변수의 중요도는 높은 것으로 간주하는 것이다. 본 기법을 도식화하면 다음과 같다.

〈그림 4-1〉 랜덤포레스트 변수 중요도



자료: vvakki(검색일: 2021.07.09.), 저자 재편집

특히, 본 기법은 주어진 자료를 훈련 샘플(training sample)과 시험 샘플(test sample)로 구분하여 시험 샘플에서만 특정 변수의 순서를 무작위로 섞음으로써 예측력 중심의 변수 선별 기법이라고 볼 수 있다.



본 기법은 RNA 분석, HIV 연구 등의 생물학, 의료 데이터, 전력량 예측 등 다양한 분야에 적용되고 있어 실용성이 높은 변수 선별 기법이라고 볼 수 있다.<sup>52)</sup>

## (2) 지도 주성분 분석(supervised principal component analysis)

지도 주성분 분석(supervised principal component analysis)에 기반한 변수 선별 기법은 Bair *et al.*(2006)에 의해 제안되었다. 이 방법은 고차원 데이터를 분석하는데 회귀분석을 단순히 적용할 수 없다는 점에서 착안되어 개발되었다. 또한, 기존의 주성분분석(principal component analysis)이 독립변수만을 고려하여 주성분(principal component)을 추출하였다면, 본 기법은 종속변수와의 관계를 고려하여 주성분을 추출한다는 점에서 차이점을 가지고 있다.

### 〈그림 4-2〉 지도 주성분 분석 알고리즘

1. Compute (univariate) standard regression coefficients for each feature.
2. Form a reduced data matrix consisting of only those features whose univariate coefficient exceeds a threshold  $\theta$  in absolute value ( $\theta$  is estimated by cross-validation).
3. Compute the first (or first few) principal components of the reduced data matrix.
4. Use these principal component(s) in a regression model to predict the outcome.

자료: Bair *et al.*(2006)의 내용을 바탕으로 저자 작성

52) HIV 연구(Altmann *et al.*, 2010), 전력량 예측 연구(Huang and Xu 2016)

---

본 기법은 우선 각 독립변수를 종속변수에 회귀하여 독립변수들의 회귀계수를 얻는다. 이후 일정한 값 이상의 회귀계수를 지닌 독립변수들만을 추출하여 새로운 독립변수 집합을 구성한다. 이 새로운 독립변수 집합에 주성분분석을 적용하여 주성분들을 추출하여 종속변수를 이 주성분들에 회귀하는 절차로 구성되어 직관적으로도 기존의 주성분분석보다 종속변수의 영향력에 초점을 둔 변수 선별 기법이라고 볼 수 있다.

본 기법은 종속변수에 주는 영향력을 기준으로 주성분분석 전에 독립변수를 선별하는 과정을 거치고 있으므로 변수를 선별하는 효과가 주성분분석보다 더 클 것이라고 예상할 수 있다. 이에 따라 본 기법이 처음 제안되었을 때 적용된 사례는 대표적인 고차원 데이터인 유전자 데이터 분석 사례로 기존의 주성분분석 대비 보다 많은 독립변수를 선별하는 것으로 나타났다. 이후 본 기법은 심장마비, 암, 치매 등 다양한 의료, 생물학 분야에서의 고차원 데이터에서 적용되었다.<sup>53)</sup>

본 연구에서는 지도 주성분 분석을 변수 선별 목적으로 사용하고 선별된 변수들을 다양한 머신러닝에 적용하여 지도 주성분 분석의 변수 선별 효과를 분석하고자 한다.

### 3) 정규화 기법

고차원 데이터의 문제점을 라그랑지안(Lagrangian) 형식으로 접근하는 방식을 정규화 기법이라고 한다. 라그랑지안 형식은 보통 목적함수를 제약 조건 하에서 최적화하는 형식을 취하고 있다. 회귀분석( $Y = X\beta + e$ )에서 목적함수를 아래 식과 같이 설정하고,

---

53) 암에 관한 연구(Barshan *et al.* 2011), 치매 관련 연구(Shigemizu *et al.* 2019)

$$\arg \min_{\beta} \| Y - X\beta \|_2^2 \quad [\text{식 4-4}]$$

제약조건을

$$\| \beta \|_q \leq t \quad [\text{식 4-5}]$$

로 두면 라그랑지안 문제는

$$\arg \min_{\beta} \| Y - X\beta \|_2^2 + \lambda \| \beta \|_q \quad [\text{식 4-6}]$$

이 된다. 여기서  $q$ 를 어떻게 설정하는가에 따라서 정규화 방식이 달라진다. 능선회귀(ridge regression)의 경우는  $q=2$  이라고 둔 것이며 LASSO(Least Absolute Shrinkage and Selection Operator)의 경우는  $q=1$  로 설정한 것이다. 또한, 이 두 가지 방법을 조합한 방법을 Elastic net이라고 한다. 본 연구에서는 LASSO 및 Elastic net과 비교적 최근의 정규화 기법으로 제안된 Generalized LASSO 방법을 사용하기로 한다.

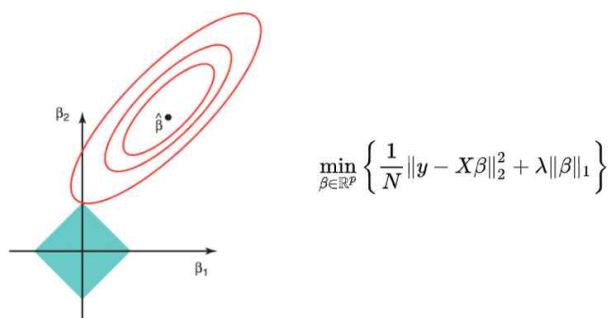
#### (1) LASSO(Least Absolute Shrinkage and Selection Operator)

제약조건 없이 회귀분석을 최소자승법으로 푸는 경우 주어진 데이터에 모델이 과적합(overfitting)되는 문제가 발생하여 예측력이 높지 않은 경우가 발생할 수 있다. LASSO 기법은 정규화 기법을 통하여 최소자승법으로 추정된 회귀계수들보다 작은 크기의 계수를 추정함으로써, 최소자승법이 주어진 자료에만 치중되는 과적합 문제를 완화시키고 예측력을 높이는 기법으로 Tibshirani(1996)에 의해 도입되었다 LASSO 식은 아래와 같다.

$$\arg \min_{\beta} \|Y - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \quad [\text{식 4-7}]$$

위 문제를  $\beta$ 공간에 도시하면 아래와 같다.

〈그림 4-3〉 LASSO



자료: Hastie *et al.* (2009), p. 71. 저자 재편집

상기 그림에서 녹색 사각형은 제약조건이며 붉은 능고선은 목적함수를 나타내고 있다. 제약조건이 사각형이므로  $\beta_1$ 이 0으로 추정되는 것을 확인할 수 있다. 즉, LASSO는 정규화 기법을 통하여 변수 선별 기능까지 제공하는 것이다.

과적합 문제 완화를 통한 예측력의 향상과 변수 선별 기능으로 LASSO는 다양한 분야에서 응용되었다. 최근에는 교육학, 지질학 등 다양한 연구 분야에서 응용되고 있는 추세이다.<sup>54)</sup>

54) 교육학 연구(구미령유진은, 2021), 지질학 연구(Wang and Zuo, 2020)

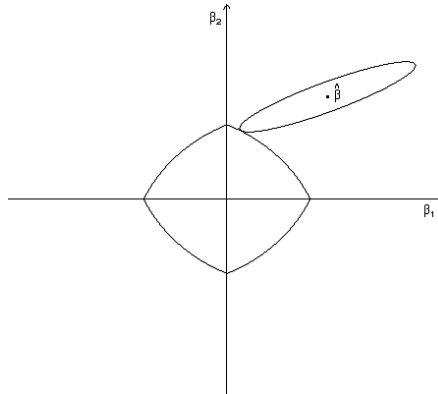
## (2) Elastic net

Elastic net은 고차원 데이터에서 LASSO가 너무 적은 수의 회귀계수만을 추정한다는 문제점을 완화하기 위하여 도입된 기법이다. 상기 라그랑지안 문제에서  $q=1$ 인 경우에는 영향력이 적은 회귀계수가 0으로 추정되는 경우가 많고  $q=2$ 인 경우에는 회귀계수가 0으로 추정되는 경우가 적다. 이 두 가지 정규화 방법을 조합한 것이 Elastic net이며 식은 아래와 같다.

$$\arg \min_{\beta} \|Y - \mathbf{X}\beta\|^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2 \quad [\text{식 4-8}]$$

위 식을 보면 능선회귀와 LASSO의 제약조건이 모두 포함된 것을 볼 수 있다. 위 문제를  $\beta$ 공간에 도식하면 아래와 같다.

〈그림 4-4〉 Elastic net



자료: Zou *et al.*(2005), p. 304. 저자 재편집

상기 그림과 같이 Elastic net에서의 제약조건은 LASSO보다 완만한 형태이며 이에 따라  $\beta_1$ 과  $\beta_2$  모두 0이 아닌 값으로 추정이 된다. 이와 같이 Elastic net은 두 가지 제약조건을 모두 포함함으로써 고차원 데이터 분석에서 LASSO가 너무 적은 수의 회귀계수만을 추정하는 문제를 완화하며 고차원 데이터에 보다 적합한 정규화 기법으로 평가되어 사용되고 있다.

### (3) Generalized LASSO

라그랑지안 형태에서 LASSO를 푸는 과정에서 회귀계수들이 개별적으로 종속변수에 미치는 영향력을 중심으로 추정되게 된다. 고차원 데이터의 경우 독립변수들이 서로 연관이 있는 경우가 있는데 연관이 높은 변수들을 묶어서 회귀계수를 추정하면 보다 효율적인 정규화 효과를 기대할 수 있다. Generalized LASSO는 기존의 LASSO를 일반화하여 제안되었으며 라그랑지안 형태는 다음과 같다.

$$\arg \min_{\beta} \| Y - \mathbf{X}\beta \|^2 + \lambda \| D\beta \|_1 \quad [\text{식 4-9}]$$

여기서 매트릭스  $D$ 가  $\beta$ 의 특성을 나타내는 페널티(Penalty) 매트릭스이다. 연관성이 높은 독립변수들은  $D$ 의 원소들을 1과 -1을 사용하여 배열할 수 있다. 참고로 일반적인 LASSO의 경우  $D$ 는 항등행렬  $I$ 이다.

〈그림 4-5〉 매트릭스  $D$ (예시)

$$D_{1d} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

자료: Tibshirani and Taylor(2011), "The solution path of the generalized lasso"

Generalized LASSO는 기존의 LASSO 및 LASSO의 변형된 모델들을 일반화한 모형이므로 다양한 분야에 적용되어 왔다. Wang *et al.*(2020)은 신호처리 분야에 동 기법을 적용했으며, Bian *et al.*(2020)은 뇌공학에 Generalized LASSO를 활용하였다.

#### 4) 기타 머신러닝 기법

##### (1) k-최근접 이웃법(k-nearest neighbor, kNN)

kNN 방법은 머신러닝 알고리즘 중 지도학습에 속하며, 거리기반 분류분석 모델로 데이터를 가장 가까운 속성에 따라 분류하여 레이블링 하는 알고리즘으로 개발되었다. 평균 식을 활용한 kNN 식은 아래와 같다.

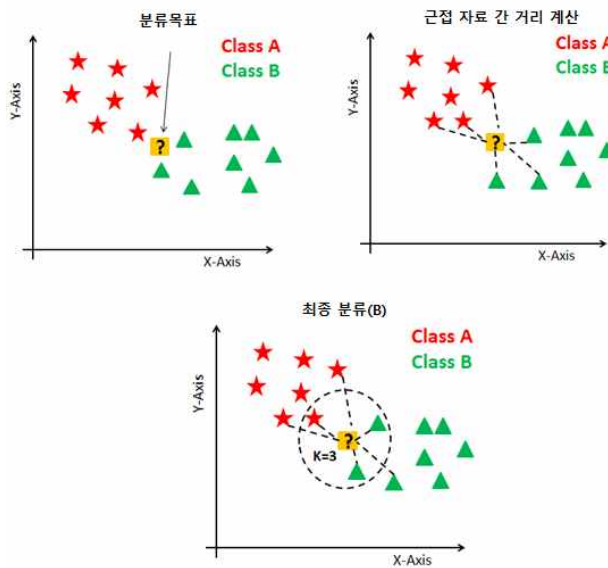
$$\hat{Y} = \frac{\sum_{i=1}^k Y_i}{k} \quad [\text{식 4-10}]$$

여기서  $Y_i$ 는 주어진 데이터,  $\hat{Y}$ 는 추정치,  $k$ 는 이웃 개수이다.

kNN 근접 거리에 있는  $k$ 개의 데이터를 활용하여 추정을 하는 방식을 따르고 있는데, 상기 식에서 나타나듯이  $k$ 의 값을 사전에 정하는 것

이 중요하다. 일반적으로는 5, 10 등의 간단한 숫자로 주어진 데이터를 학습시킨 후 가장 예측성능이 높은  $k$ 값을 결정하는 경우가 많은 편이다.

〈그림 4-6〉  $k$ -최근점 이웃법



자료: towards data science(검색일:2021.07.09), 저자 재편집

상기 그림은 분류 목표를  $k$ NN 방법을 이용하여 Class A 또는 Class B로 분류하는 모습이다. 단순하고 효율적이며 수치 기반 데이터 분류 작업에서 성능이 우수한 것으로 알려진  $k$ NN 방법은 다양한 분야에서 응용되었다. 최근에는 의료보건 분야, 교통 분야 등 다양한 분야에서 응용되고 있다.<sup>55)</sup>

55) 의료보건 분야(Xing & Bei, 2019), 교통분야(Luo *et al.*, 2019)



## (2) k-평균 군집화(k-means clustering)

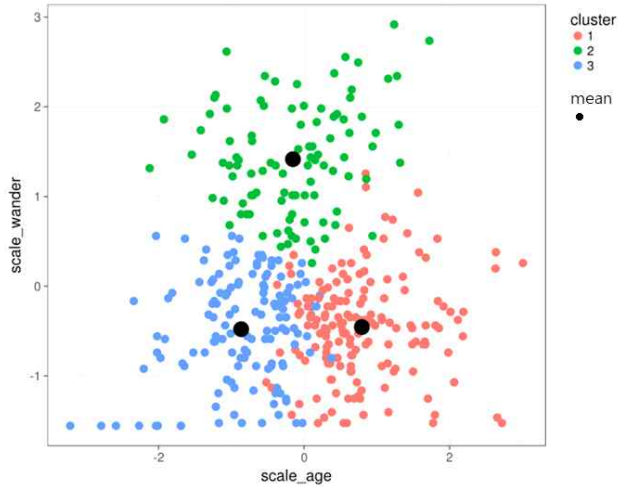
k-평균 군집화 방법의 데이터는 여러 개의 군집들로 구분될 수 있으며, 각 군집은 하나의 중심을 가진다. 각 군집에 포함된 개체들은 이 중심까지의 거리가 주어지며, 이 거리들이 가장 작게끔 군집의 개수를 정하는 알고리즘으로 개발되었다. 주어진 데이터가 적절히 군집화 된다면, 우리는 미래값과 가까운 군집들의 평균값을 활용하여 미래값을 더욱 잘 예측할 수 있으며 본 연구에서는 이 방법을 적용하여 미래값을 예측하였다.

k-평균 군집화 방법의 식은 아래와 같다.

$$\arg \min_{\mathcal{S}} \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \quad [\text{식 4-11}]$$

$n$ 개의  $d$ -차원 데이터 관측치  $(x_1, x_2, \dots, x_n)$  집합이 주어졌을 때, k-평균 군집화 알고리즘은  $n$ 개의 데이터 관측치들을 각 집합 내 관측치 간 거리를 최소화하는  $k(\leq n)$ 개의 집합  $\mathcal{S} = \{S_1, S_2, \dots, S_k\}$ 으로 분할한다. 다시 말해,  $\mu_i$ 가 집합  $S_i$ 의 중심점이라 할 때 각 집합별 중심점과 집합 내 관측치 간 거리의 제곱합을 최소로 하는 집합  $\mathcal{S}$ 를 찾는 것이 이 알고리즘의 목표이다.

〈그림 4-7〉 k-평균 군집화



자료: K-Means Clustering Tutorial (검색일: 2020.2.12), 저자 재편집

상기 그림은 레이블이 없는 데이터들을 비슷한 특성을 지닌 데이터들끼리 묶어 3개의 군집으로 군집화한 모습이다. k-평균 군집화 방법은 구현이 매우 간단한 알고리즘이면서도 실행 속도가 빠르면서도 좋은 성능을 가진 것으로 알려져 있다. 최근에는 건강과학 분야, 농업 분야 등 다양한 연구 분야에서 응용되고 있다.<sup>56)</sup>

### (3) 주성분회귀(PCR: Principal Component Regression)

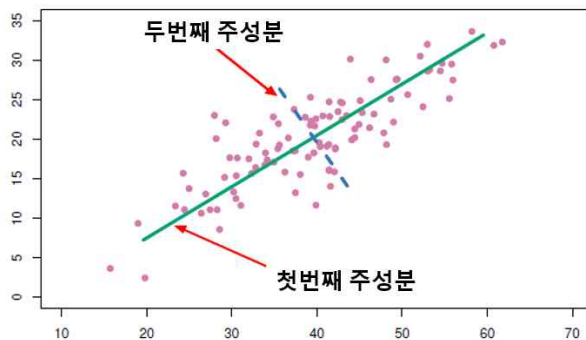
관측치 개수에 대비하여 독립변수가 많은 경우, 설명력이 높지 않은 변수가 포함되거나, 다중 공선성의 문제가 발생할 수가 있다. 주성분 분석(PCA: Principal Component Analysis)은 많은 변수를 선형 조합하

56) 건강과학 분야(Shakeel *et al.*, 2018), 농업 분야(Windarto, 2017)

여 적은 수의 변수(주성분)들로 환원시키고, 각 독립변수들의 분산크기 및 직교성을 고려하여 이러한 문제를 완화시키는 방법으로 개발되었다. PCR은 주성분을 독립변수로 회귀분석에 적용한 기법으로 본 연구에서 사용되었다.

주성분회귀의 절차는 다음과 같다. 1) 주성분 분석을 수행하여 새로운 주성분을 생성, 2) 새롭게 생성된 주성분을 입력하여 선형회귀모형을 학습, 3) 주성분들로 추정된 선형회귀모형을 사용하여 분류 및 예측 순이다.<sup>57)</sup>

〈그림 4-8〉 주성분분석



자료: Nicome - The Data Blog(검색일: 2020.2.12), 저자 재편집

상기 그림은 2차원의 데이터를 분산이 최대로 보존되는 다른 차원의 축 평면으로 투영시키는 모습이다. 이러한 주성분분석은 최근에는 공중 보건 분야(Song *et al.*, 2019), 농업 분야(Chen *et al.*, 2020) 등에서 사용되기도 하였다.

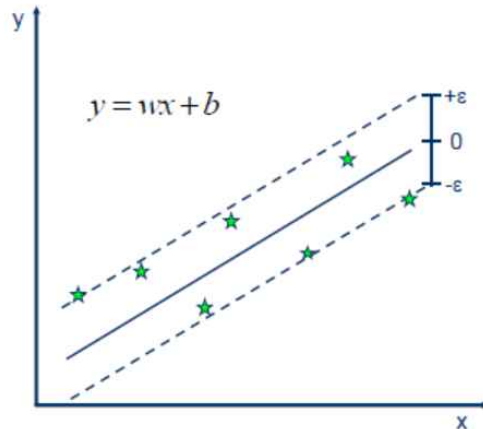
57) OpenGenus, <http://www.opengenus.org/> (검색일: 2021.08.08.)

#### (4) 서포트벡터회귀(SVR: Support Vector Regression)

이 기법은 머신러닝에서 분류(classification) 목적으로 주로 사용되는 서포트벡터머신(Support Vector Machine, SVM) 방법을 회귀분석에 적용한 것으로, 서포트벡터머신은 패턴인식, 자료 분석을 위한 모델로 개발되었다. 선형회귀분석이 오직 하나의 선형회귀계수를 도출한다면, SVM은 하이퍼파라미터를 사용하여, 적당한 간격(margin) 내에서 최적 모델을 찾는 방법으로 모델의 과적합을 피할 수 있는 방법이다. 특히, 커널(kernel) 함수를 사용하여 복잡해 보이는 데이터를 낮은 차원으로 환원하여 선형모델로 분석하게 해주는 것은 다른 머신러닝과의 차이점이라고 할 수 있다.

서포트벡터회귀는 아래 그림 및 수식과 같이 나타낼 수 있다.

〈그림 4-9〉 서포트벡터회귀



자료: saedsayad(검색일: 2020.08.08.), 저자 재편집

$$\text{목적함수} : \min_w \frac{1}{2} \|w\|^2 \quad [\text{식 4-12}]$$

$$\text{제약식} : \begin{aligned} y_i - wx_i - b &\leq \epsilon \\ wx_i + b - y_i &\leq \epsilon \end{aligned} \quad [\text{식 4-13}]$$

즉, SVR은 주어진 간격  $\epsilon$  내에서 주어진 데이터를 가장 많이 포함하는 식을 찾아내는 것이다. 이러한 SVR이 최근에 사용된 사례는 신경컴퓨팅 분야, 수문학 분야 등이 있다.<sup>58)</sup>

## (5) Bagging

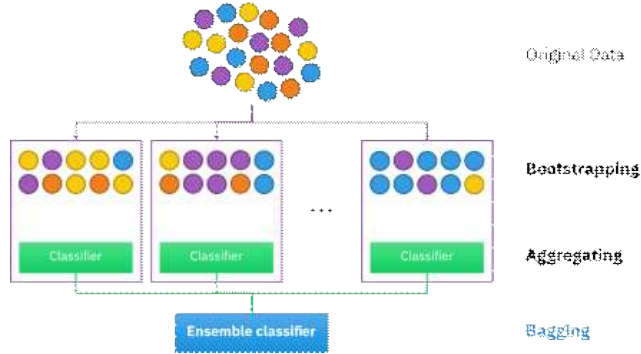
부트스트랩 방법으로 주어진 자료를 복제하고 각 복제된 샘플에 모델을 학습한 후 각 결과치를 통합하여 최종 결과치를 도출하는 방법을 Bagging이라고 한다. 이 기법으로 우리는 주어진 자료가 내포할 수 있는 편의를 감소시킬 수 있으며, 본 연구의 자료와 같은 소규모 자료를 활용하여 보다 일반화된 결과를 도출할 수 있다. 또한, Bagging을 통한 결과치는 샘플들의 평균이므로 분산이 감소되는 장점을 가지고 있다.

데이터  $Z = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 를 사용하여 Bagging 추정량을 구하는 경우, 부트스트랩 샘플들을  $Z^b$ ,  $b = 1, 2, \dots, B$ 라 하면 Bagging 추정량  $\hat{f}_{bag}$ 은 아래 식과 같다.

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x) \quad [\text{식 4-14}]$$

58) 신경컴퓨팅 분야(Quan *et al.*, 2020), 수문학 분야(Panahi *et al.*, 2020)

〈그림 4-10〉 Bagging



자료: wikipedia(검색일: 2021.7.20.)

상기 그림과 같이 Bagging은 특정한 최적화 방식을 따르는 것이 아니라 머신러닝 기법의 적용 절차를 의미하므로 Bagging의 활용분야는 상당히 방대한 편이다. 이에 따라, Bagging과 다른 통계적 기법에 적용 및 융합을 통한 연구가 많이 수행되어 왔다.<sup>59)</sup>

## 5) 예측조합법

본 연구에서 쓰인 관측치는 2005년부터 2020년까지 총 16개에 독립변수 개수보다 적다. 이에 따라 예측값의 분산이 클 수가 있다. 이러한 문제를 완화하기 위하여 우리는 다양한 예측기법들로부터 얻어진 예측값들을 조합하는 예측조합(forecasting combination) 기법을 사용할 수 있다. 예측조합기법의 기본적인 수식은 다음과 같다.

59) 비모수회귀(Barreiro-Ures *et al.*, 2021), 신경망 모형(Rew *et al.*, 2021)

$$f_c = \sum_{i=1}^m \omega_i f_i \quad [\text{식 4-15}]$$

여기서,  $f_c$  는 최종예측값,  $m$ 은 예측기법 개수,  $\omega_i$ 는 각 예측기법의 예측값에 부여되는 가중치,  $f_i$ 는 각 예측기법의 예측값임.

가중치  $\omega_i$ 를 어떻게 정하느냐에 다양한 예측조합 기법이 존재할 수 있다. 본 연구에서는 3가지 방법을 사용하였다. 각 예측조합 기법에 동등하게 가중치를 부여하는 방법(simple average),

$$\omega_i = 1/m \quad [\text{식 4-16}]$$

각 예측기법의 예측값들의 오차분산(MSE: Mean Squared Error)에 반비례하도록 가중치를 부여하는 방법(variance-based),

$$\omega_i \propto 1/\text{MSE}(f_i) \quad [\text{식 4-17}]$$

그리고 예측 직전까지의 예측값 중에서 MSE가 가장 낮은 예측기법을 선택하는 방법(best individual model),

$$\omega_i = \begin{cases} 1 & \text{MSE}_i < \text{MSE}_{-i} \\ 0 & \text{otherwise} \end{cases} \quad [\text{식 4-18}]$$

등이 본 연구에서 사용된 예측조합법들이다.

예측조합법은 경제학 특히 금융 분야에서 안정된 예측을 위하여 주로 적용되어왔는데, 최근에는 인구학, 산업 분야 등 다양한 방면에서 활용되고 있다.<sup>60)</sup>

60) 인구학(Shang and Haberman, 2018), 전력수요 예측 등의 산업 분야(Laouafi *et al.*, 2017)

## 제2절 2022 해양수산업 전망

### 1. 베이지안 VAR를 통한 전망

#### 1) 전망 절차

베이지안 VAR을 통한 해양수산업 부가가치 전망은 다음과 같은 과정에 의해 수행되었다. 첫째, 해양수산업 특수분류 중 성격이 유사한 부문을 묶어 모형을 구성하였다. 모형은 수산업, 해운·항만업, 조선·제조·건설업, 기타 해양수산업, 총 4개로 구성하였다. 부문별로 살펴보면, 수산업을 구성하는 대분류는 수산물 생산업, 수산물 가공업, 수산물 유통업으로 모두 수산물과 관련된 산업으로 구성하였다. 해운·항만업은 해운업과 항만업으로 구성되었으며 세계 교역에 의한 운송 및 보관과 관련되어있다. 조선·제조·건설업은 선박 및 해양플랜트 건조 수리업, 해양수산 기자재 제조업, 해양자원 개발 및 건설업으로 구성되어있다. 기타 해양수산업은 해양수산 레저관광업과 해양수산 관련 서비스업으로 모형 중 내수산업의 성격이 짙은 산업으로 구성하였다.

〈표 4-1〉 베이지안 VAR 모형 구성

모형	해양수산업 대분류
수산업	수산물 생산업, 수산물 가공업, 수산물 유통업
해운·항만업	해운업, 항만업
조선·제조·건설업	선박 및 해양플랜트 건조 수리업, 해양수산 기자재 제조업, 해양자원 개발 및 건설업
기타 해양수산업	해양수산 레저관광업, 해양수산 관련 서비스업

자료: 저자 작성



둘째, 모형을 구성하는 해양수산업 대분류에 공통적으로 영향을 줄 것으로 예상되는 경제변수를 모형에 추가하였다. 예를 들어, 수산업 모형의 벡터를 구성하는 변수는 수산물 생산업, 수산물 가공업, 수산물 유통업의 부가가치 외에도 수산물 생산, 수산물 소비, 그리고 수산물 수출을 추가하여 모형을 구성하였다.

〈표 4-2〉 모형별 경제변수

모형	경제변수
수산업	수산물 생산, 수산물 수출, 1인당 수산물 소비량
해운·항만업	ClarkSea Index, 국내 항만 총물동량
조선·제조·건설업	세계 GDP, 국제유가(브렌트유)
기타 해양수산업	우리나라 GDP, 크루즈 방문객 수

자료: 저자 작성

셋째, 베이지안 VAR을 추정한다. 이때 모형별 시차는 각 시차별 베이지안 모형을 비교하여 결정한다. 즉 시차별로 베이지안 모형을 추정한 다음 시차별 추정 결과의 로그-한계우도함수(log-marginal likelihoods)를 비교하여 적정 시차를 결정한다. 베이지안 VAR을 추정하기 위한 사전분포는 미네소타 사전분포를 사용하였음을 밝힌다.

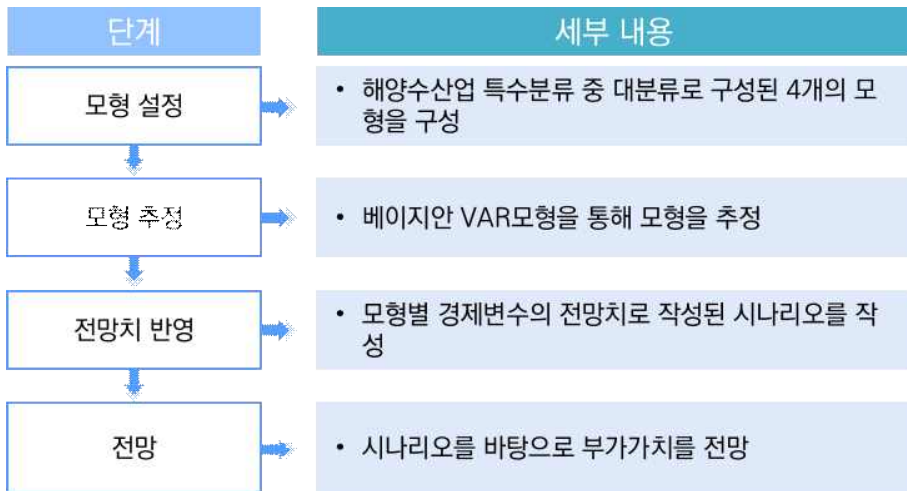
넷째, 주요기관의 모형별 경제변수의 전망치를 반영하여 시나리오를 작성한다. 각 경제변수의 전망치의 출처는 〈표 4-3〉에 수록하였다. 마지막으로, 시나리오를 바탕으로 전망을 수행한다.

〈표 4-3〉 경제변수 전망치 출처

경제변수	전망기관
수산물 생산, 수산물 수출, 1인당 수산물 소비	KMI
ClarkSea Index, 국내 항만 총물동량	KMI
세계 GDP, 국제유가(브렌트유)	IMF
우리나라 GDP, 크루즈 방문객	IMF, KMI

자료: 저자 작성

〈그림 4-11〉 베이지안 VAR 전망 절차



자료: 저자 작성

## 2) 데이터

앞서 밝혔듯이 베이지안 VAR모형을 통한 전망은 해양수산업 4개 부문 모형과 각 부문과 밀접한 상관관계를 지닐 것으로 예상되는 경제변수로 구성하였다. 경제변수의 설정은 한국해양수산개발원의 ‘2022 해양수산 주요 지표 전망’의 지표를 참고하였다.

‘2022 해양수산업 주요지표 전망’ 보고서는 해양수산업 주요부문에 대한 경제지표, 사회·현안지표 등을 작성하여 해양수산업의 객관적이며 종합적인 진단과 전망을 수행하는 것을 목적으로 한다. 해양수산업의 대분류 중 정책적 중요도가 높은 5대 해양수산업을 대상으로 산업별 주요지표를 선정하고 분석하는데, 주요지표는 각 부문별 워크숍과 국내외 문헌조사와 가용통계 현황을 점검하여 선정하였다.<sup>61)</sup> 주요 산업별 주요지표와 연구방법은 <표 4-4>와 같다.

<표 4-4> 2022 해양수산업 주요지표 전망

구분		추진 내용	방법
지표선정		• 주요지표 선정 및 검토, 자료 가용성	• 통계수집, 문헌조사, 전문가자문
지표 전망	수산	• 생산량, 수출입액, 가격, 소비량 등 전망	• KMI 수산업전망모형 FOSIM 적용
	해운	• 해상물동량, 운임지수 등 전망	• KMI 해운시황 전망모형
	항만	• 항만물동량 전망	• KMI 항만수요예측모형
	해양레저	• 관광객 수, 동력레저기구 등	• 문헌조사, 계량분석
	부가가치	• 각 부문별 실질 부가가치 전망	• KMI 부가가치 전망모형
영향요인분석		• 해양수산업 영향요인 및 경제지표 분석	• 문헌조사, 키워드분석 적용
기업전망		• 해양수산업 기업경기전망	• KMI 기업경기조사 결과분석

자료: 장정인 외(2021c), p. xii

이러한 논의를 바탕으로 2022 해양수산업 전망에 활용된 변수는 <표 4-5>와 같다. 모형별로 살펴보면, 수산물을 주요 상품으로 취급하는 수산업 모형은 수산업 관련 대분류의 부가가치 외에도 수산물 생산과 수산물 수출, 그리고 1인당 수산물 소비를 모형에 포함시켜 예측력을 제고하고자 하였다. 해운·항만업은 대분류의 부가가치와 해운 운임, 우리나라 항만의 총 물동량을 사용하여 부가가치를 전망하였다. 조선·제조·건설업은 각 대분

61) 장정인 외(2021c), p. xi.

류의 부가가치와 세계 GDP, 국제유가를 사용하였다. 마지막으로 기타 해양수산업은 부가가치와 우리나라 GDP, 크루즈 방문객을 사용하여 전망하였다.

〈표 4-5〉 변수 설명 및 출처

모형	경제변수	단위	출처
수산업	수산물 생산업 부가가치	백만 원	KMI
	수산물 가공업 부가가치		
	수산물 유통업 부가가치		
	수산물 생산	kg	통계청
	수산물 수출	US\$	
	1인당 수산물 소비	kg	한국농촌경제연구원
해운·항만업	해운업 부가가치	백만 원	KMI
	항만업 부가가치		
	ClarkSea Index	\$/day	Clarksons
	국내 항만 총물동량	TEU	해운항만물류정보시스템
조선·제조·건설업	선박 및 해양플랜트 건조 수리업 부가가치	백만 원	KMI
	해양수산 기자재 제조업 부가가치		
	해양자원 개발 및 건설업 부가가치		
	세계 GDP	US\$	IMF
	국제유가(브렌트유)	US\$	FRED
기타 해양수산업	해양수산 레저관광업 부가가치	백만 원	KMI
	해양수산 관련 서비스업 부가가치		
	우리나라 GDP	US\$	IMF
	크루즈 방문객	명	한국관광공사

자료: 저자 작성

분석을 위해 해운업과 항만업의 부가가치를 제외한 모든 변수는 로그 차분하여 전년 대비 증감률로 변환하였다. 해운업과 항만업 부가가치는 로그 수준변수를 사용하여 추정하였음을 밝힌다. 분석자료는 연간자료를 사용했으며 분석기간은 2004년부터 2020년이다. 모형에 사용된 변수의 기초통계량은 다음과 같다.

〈표 4-6〉 기초통계량

단위: %, 로그

변수	Obs.	평균	표준편차	min	max
수산물 생산업 부가가치	16	0.5	8.0	-21.0	10.9
수산물 가공업 부가가치	16	3.8	5.6	-3.8	17.8
수산물 유통업 부가가치	16	5.4	5.7	-2.1	23.9
수산물 생산	18	2.4	5.1	-5.5	13.0
수산물 수출	18	1.8	7.5	-9.8	23.5
1인당 수산물 소비	18	4.4	10.7	-9.3	25.0
해운업 부가가치	17	15.3	0.3	14.8	15.8
항만업 부가가치	17	15.3	0.3	14.6	15.6
ClarkSea Index	18	-1.3	35.7	-105.9	61.6
국내 항만 총물동량	18	3.0	4.5	-9.2	11.1
선박 및 해양플랜트 건조 수리업 부가가치	16	-0.2	16.6	-39.4	20.8
해양수산 기자재 제조업 부가가치	16	0.8	9.8	-21.2	12.2
해양자원 개발 및 건설업 부가가치	16	-2.7	13.1	-26.7	33.5
세계 GDP	18	2.9	2.2	-3.5	5.7
국제유가(브렌트유)	18	3.5	31.0	-63.6	52.6
해양수산 레저관광업 부가가치	16	-1.3	19.9	-63.3	37.9
해양수산 관련 서비스업 부가가치	16	4.1	6.5	-3.2	22.8

변수	Obs.	평균	표준편차	min	max
우리나라 GDP	18	3.2	1.7	-1.0	6.6
크루즈 방문객	14	15.7	68.4	-160.1	103.6

자료: 저자 작성

일반적인 VAR모형은 모형을 추정하기 위해 적정 시차를 선택하는 과정을 거친다. 적정시차는 주어진 데이터의 정보기준(information criteria)을 기준으로 선택하는데, 대표적인 정보기준으로는 아카이케정보기준(Akaike information criteria)과 베이즈정보기준(Bayesian information criteria)이 있다. 베이즈안 VAR은 이러한 정보기준을 통해 적정 시차를 선택하기 보다는 시차별 모형을 추정하고, 각 모형의 로그-한계우도함수를 계산하여, 사후확률(posterior probability)이 가장 높은 모형의 시차를 적정시차로 설정한다.<sup>62)</sup> 각 모형의 로그-한계우도함수와 사후확률의 검정결과는 <표 4-7>과 같다.

<표 4-7> 모형별 최적 시차

수산업			해운·항만업		
시차	로그-우도함수	사후확률	시차	로그-우도함수	사후확률
1	45.03	1.00	1	36.28	0.47
2	26.44	0.00	2	45.46	0.82
3	16.98	0.00	3	18.97	0.00
4	1.63	0.00	4	12.97	0.00
조선·제조·건설업			기타 해양수산업		
시차	로그-우도함수	사후확률	시차	로그-우도함수	사후확률
1	-30.74	0.50	1	2.08	0.00
2	-30.74	0.50	2	9.06	1.00
3	-40.65	0.00	3	-2.05	0.00
4	-51.67	0.00	4	-9.48	0.00

자료: 저자 작성

62) StataCorp(2021), p. 27.

### 3) 전망 결과

적정시차 검정결과를 바탕으로 각 모형별 베이지안 VAR모형을 추정하였다. 모형별 추정결과에 따라 2022년 해양수산업 부가가치 전망을 수행하였다. 전망치의 추정은 주요기관의 경제변수 전망치를 모형에 대입하고, 베이지안 VAR에서 얻은 변수 간 상관관계를 활용하여 해양수산업 대분류별 부가가치의 예측을 수행하였다. 2022년 해양수산업 부가가치 전망결과는 <표 4-8>과 같다.

<표 4-8> 해양수산업 부가가치 전망

단위: 백만 원, %

모형	대분류	2019	2020p	2021f	2022f
수산업	수산물생산업	2,867,088	2,937,145 (2.4)	2,971,115 (1.2)	3,020,492 (1.7)
	수산물가공업	1,904,780	1,942,521 (2.0)	2,002,662 (3.1)	2,057,898 (2.8)
	수산물유통업	3,557,231	3,484,150 (-2.1)	3,334,932 (-4.3)	3,224,299 (-3.3)
해운· 항만업	해운업	6,838,222	6,672,071 (-2.4)	7,270,771 (9.0)	7,692,505 (5.8)
	항만업	3,538,789	3,291,597 (-7.0)	3,305,052 (0.4)	3,268,250 (-1.1)
조선· 제조· 건설업	선박및해양플랜트 건조수리업	11,954,343	10,723,365 (-10.3)	11,527,295 (7.5)	12,266,947 (6.4)
	해양수산기자재제 조업	1,474,621	1,624,912 (10.2)	1,730,008 (6.5)	1,743,428 (0.8)
	해양자원개발및건 설업	782,693	799,211 (2.1)	845,382 (5.8)	814,861 (-3.6)
기타 해양수산 업	해양수산레저관광 업	413,180	369,028 (-11.3)	357,443 (-3.1)	372,882 (4.3)
	해양수산물관련서비 스업	14,300,352	14,067,834 (-1.6)	14,223,810 (1.1)	14,672,321 (3.2)

모형	대분류	2019	2020p	2021f	2022f
	해양수산업	47,631,300	45,911,834 (-3.6)	47,568,472 (3.6)	49,133,883 (3.3)
	부가가치 기여도	2.57%	2.50%	2.48%	2.48%

주: p는 잠정치, f는 전망치, ( )는 전년대비 증감  
자료: 저자작성

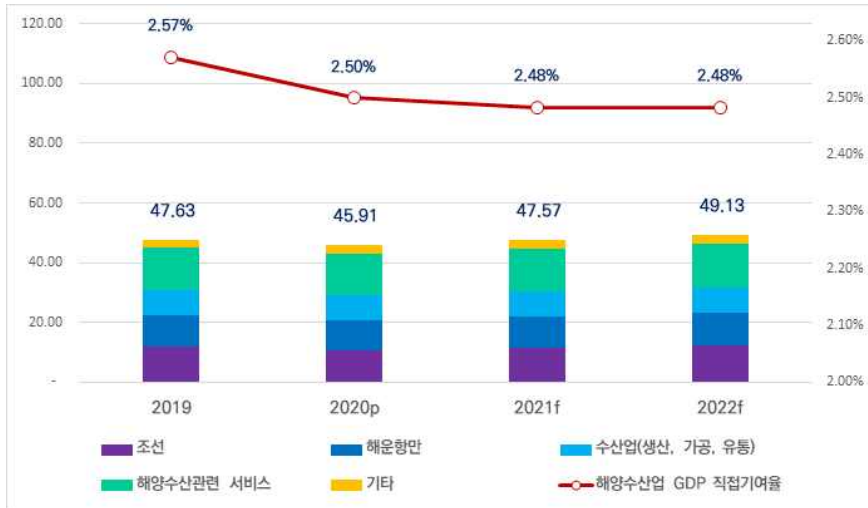
2022년은 국내외 경제의 회복세와 함께 해양수산 분야 경제지표도 양호한 흐름을 이어갈 것이나 전년도에 비해 증가폭은 감소할 것으로 전망되었다. 분야별로 살펴보면, 수산분야는 생산, 수출, 소비가 개선됨에 따라 부가가치는 역시 다소 개선될 것으로 전망되었다. 비대면 거래 및 가정간편식 시장 확대는 긍정요인으로 작용하겠으나 원자재 가격 및 인건비 상승 등 생산자 물가의 상승 압력은 수산분야에 부정적으로 작용할 우려가 있다. 2021년 유례없는 호황을 맞이한 해운 분야는 2022년에도 긍정적인 흐름을 이어갈 것으로 전망되었다. 주요 운임은 2021년에 비해 하락할 것으로 예상되나 하락폭이 크지 않을 전망이고, 물동량은 증가세가 예상된다. 항만 분야는 2021년의 긍정적인 기조가 이어질 전망이다. 2021년 세계경제 회복세로 인해 증가한 물동량은 2022년에도 비대면 소비재 수요 증가로 인해 호조를 유지할 것으로 보인다. 2021년 침체기에서 벗어난 조선 분야는 2022년에도 양호한 흐름을 이어갈 전망이다. 세계 발주량은 전년 대비 감소할 것으로 예상되나 국내업체가 강점을 지니는 LNG선 수요 증가와 환경 규제 강화로 인한 노후선 교체압력 상승은 국내 조선 분야에 긍정적인 요인으로 작용할 것으로 보인다. 해양레저관광 분야는 양적 감소가 2022년에는 다소 개선될 전망이지만 코로나19 이전 수준으로의 회복에는 시간이 소요될 것으로 전망되었다. 다만 한적한 관광지 선호현상과 가치소비 중심의 MZ세대의 수요 다변화 현상은 해양레저관광에 긍정적인 요인이라 할 수 있다. 종합해보면, 2022년 해양수산업은 양호한 성장세가 지속되어 실질부가가치 기준 약 3.3%의 성장세를 보일 것으로 예상되며, 우리나라



GDP에 대한 직접기여도는 지난해에 이어 약 2.5% 수준을 유지할 것으로 보인다.<sup>63)</sup>

〈그림 4-12〉 2022 해양수산업 전망

단위: 조 원, %



자료: 저자 작성

## 2. 머신러닝 기법에 의한 전망

### 1) 대분류별 부가가치 전망

본 연구는 다양한 머신러닝 기법으로 해양수산업 부가가치를 전망하는 것을 주요 목적으로 한다. 이에 따라, 각 산업별 부가가치 전망은 머신러닝 기법들과 이 머신러닝 기법들에 예측조합법을 활용하여 이루어질 것이다.

각 기법들의 예측성능을 비교하기 위하여 예측오차의 평균절대비오차

63) 한국해양수산개발원 보도자료(2022. 1. 25.)

---

(Mean Absolute Percentage Error, MAPE)를 사용하였다. MAPE의 수식은 다음과 같다.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - f_i|}{y_i} \quad [\text{식 4-19}]$$

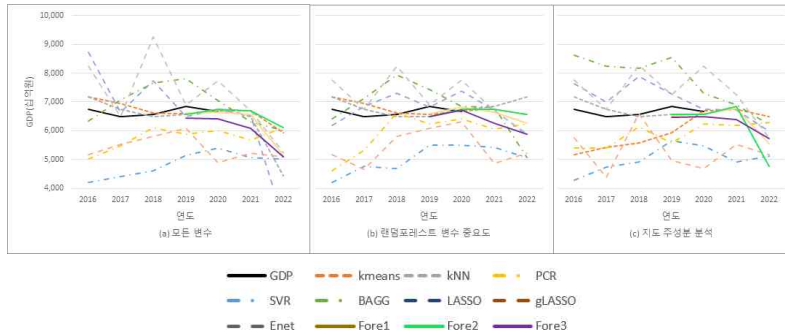
여기서,  $N$ 은 데이터 개수,  $y_i$ 는 실제값,  $f_i$ 는 예측값임.

본 연구에서는 예측조합법을 사용하므로 예측기법들로 얻어진 이전 년도들의 전망치를 우선 확보해야만 한다. 즉, 예측을 위한 관측치 개수와 예측조합법을 위한 전망치 개수가 각각 필요하며, 본 연구에서 사용되는 자료의 관측치 개수가 적은 점을 고려하여 각 예측기법에 따른 전망치는 2016년부터, 예측조합법에 따른 전망치는 2016년부터의 전망치를 사용하여 구하였다. 전망치는 2022년까지 제시하였으며 MAPE값은 마지막 관측치인 2020년까지의 데이터를 사용하여 산출하였다.

### (1) 해운업

해운업 부가가치 전망에 사용된 변수는 산업연관표 항목 및 거시경제변수 외에 운임, 유가 등의 해운업과 관련된 산업별 고유변수들로 구성된다. 우선 머신러닝 기법들과 예측조합법을 사용한 전망치들은 아래 그림과 같다.

〈그림 4-13〉 해운업 부가가치 전망치 비교



주: kmeans(k-평균 군집화), BAGG(Bagging), gLASSO(Generalized LASSO), Enet(Elastic net), Fore1(Variance-based 예측조합법), Fore2(Best 예측조합법), Fore3(Simple 예측조합법)

자료: 저자 작성

상기 각 그래프는 독립변수 사용방식에 따른 구분이다. 각 그래프마다 2016년부터 2020년까지 실제 해운업 부가가치가 나타나 있으며, 2021년과 2022년도는 각 예측기법에 따른 전망치만이 표시되어 있다. 예측조합법은 각 예측기법을 활용하여 얻어지므로 2019년부터 제시된다.

해운업 부가가치의 경우 2017년에 급락하는 모습을 보이는데 이것은 한진해운 파산, 중국의 경제성장률 정체, 불안정한 유가 등이 영향을 끼친 것으로 보인다. 이후 해운업 부가가치는 회복하는 듯이 보이나 2020년의 경우 코로나 사태에 따라 다시 감소하였다.

다음 표는 해운업 부가가치와 각 예측기법에 따른 전망치들을 나타내고 있다.

〈표 4-9〉 해운업 부가가치 전망치

단위 : 십억 원

연도	GDP	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수												
2016	6,741	7,177	7,177	5,016	4,195	6,343	8,733	5,178	8,253	-	-	-
2017	6,477	6,959	6,741	5,470	4,408	7,078	6,640	5,514	6,501	-	-	-
2018	6,555	6,609	6,477	6,109	4,612	7,647	7,726	5,811	9,252	-	-	-
2019	6,838	6,591	6,555	5,882	5,139	7,810	6,482	6,082	6,902	6,616	6,555	6,431
2020	6,672	6,697	6,741	6,001	5,391	7,052	6,752	4,879	7,745	6,636	6,741	6,407
2021	-	6,689	6,477	5,666	5,073	6,321	6,548	5,218	6,703	6,557	6,689	6,087
2022	-	5,940	4,427	6,108	5,018	5,996	2,892	5,104	5,252	5,142	6,108	5,092
랜덤포레스트 변수 중요도												
2016	6,741	7,177	7,177	4,611	4,205	6,409	6,182	5,178	7,765	-	-	-
2017	6,477	6,959	6,741	5,329	4,773	7,114	6,835	4,671	6,749	-	-	-
2018	6,555	6,609	6,477	6,618	4,700	7,944	7,289	5,811	8,243	-	-	-
2019	6,838	6,555	6,477	6,232	5,513	7,438	6,826	6,082	6,896	6,582	6,477	6,502
2020	6,672	6,697	6,741	6,425	5,511	6,856	7,390	6,317	7,752	6,811	6,741	6,711
2021	-	6,755	6,838	6,048	5,435	6,810	6,706	4,864	6,697	6,638	6,755	6,269
2022	-	6,570	7,177	6,247	5,076	5,061	5,847	5,223	5,900	6,253	6,570	5,888
지도 주성분 분석												
2016	6,741	5,172	7,177	5,405	4,281	8,627	7,625	5,741	7,752	-	-	-
2017	6,477	5,433	6,741	5,408	4,751	8,248	6,990	4,388	6,756	-	-	-
2018	6,555	5,583	6,477	6,138	4,921	8,161	7,881	6,585	8,255	-	-	-
2019	6,838	5,922	6,555	5,571	5,649	8,556	7,288	4,964	7,251	6,521	6,555	6,469
2020	6,672	6,697	6,555	6,232	5,474	7,308	6,732	4,689	8,251	6,586	6,555	6,492
2021	-	6,755	6,838	6,176	4,916	6,895	6,713	5,516	7,259	6,766	6,838	6,384
2022	-	6,493	4,759	6,291	5,108	6,220	5,973	5,178	5,749	5,545	4,759	5,721

자료: 저자 작성

아울러, 상기 전망치들을 토대로 2021년 및 2022년의 전망치의 신뢰구간은 아래 표와 같이 나타났다.

〈표 4-10〉 해운업 부가가치 전망치 신뢰구간

단위: 백만 원

연도	평균	중위값	90% 신뢰구간	95% 신뢰구간
2021	6,330,317	6,638,438	[6,142,424, 6,518,209]	[6,104,373, 6,556,261]
2022	5,596,648	5,749,293	[5,360,633, 5,832,663]	[5,312,835, 5,880,461]

자료: 저자 작성

아래 표는 연도별 각 예측기법의 절대비오차(Absolute Percentage Error)와 MAPE를 보여주고 있다. 모든 변수를 고려한 경우에서 해운업 부가가치를 잘 예측하고 있는 머신러닝 예측기법은 kNN으로 나타났다. kNN 다음으로 정확한 머신러닝 예측기법은 kmeans인데, 연도별 절대비오차를 살펴보면 2017년을 제외하고는 kNN보다 높은 예측성능을 보여주고 있다는 것을 알 수 있다. 예측조합법 중에서는 Fore1이 해운업 부가가치를 가장 잘 예측하고 있는 것으로 나타났다. Fore3의 경우는 예측성능이 높지 않은 것으로 나타났는데 이는 변동성이 높은 머신러닝 예측기법에도 동일한 가중치를 두기 때문으로 판단된다. 다음으로 랜덤포레스트 변수 중요도를 활용한 경우에도 kNN 기법이 가장 높은 예측성능을 보여주었다. 모든 변수의 경우와 마찬가지로 kmeans도 높은 예측성능을 보여주었다. 예측조합법 중에서는 Fore3가 높은 예측성능을 보여주고 있다. 지도 주성분 분석에 따라 변수를 선택한 경우에도 kNN이 높은 예측성능을 보여주었으며, 그 다음으로는 LASSO가 비교적 높은 예측성능을 보여주었다. 예측조합법 중에서는 Fore2가 가장 낮은 절대비오차를 보여주었고 이것은 kNN의 결과를 반영하는 것으로 보인다. 머신러닝 기법들 중에서는 모든 변수 선택 후 kNN 기법을 사용하는 것이 가장 높은 예측성능을 보여주고, 예측조합법 중에서는 모든 변수의 경우에 Fore1이 높은 예측성능을 보여준다고 볼 수 있다.

〈표 4-11〉 해운업 부가가치 예측성능 비교

단위 : %

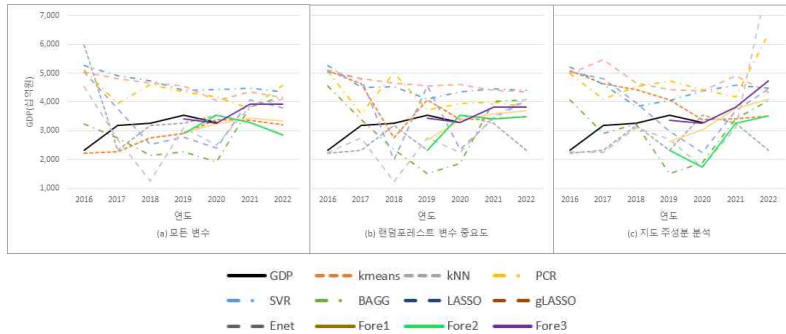
연도	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수											
2016	6.476	6.476	25.581	37.770	5.905	29.564	23.180	22.430	-	-	-
2017	7.434	4.065	15.554	31.950	9.274	2.512	14.880	0.365	-	-	-
2018	0.821	1.187	6.814	29.650	16.659	17.859	11.350	41.136	-	-	-
2019	3.614	4.139	13.978	24.850	14.213	5.203	11.050	0.938	3.244	4.139	5.961
2020	0.369	1.028	10.061	19.200	5.697	1.194	26.870	16.077	0.533	1.028	3.971
MAPE	3.743	3.379	14.398	28.680	10.350	11.266	17.470	16.189	1.889	2.584	4.966
랜덤포레스트 변수 중요도											
2016	6.476	6.476	31.601	37.610	4.915	8.293	23.183	15.194	-	-	-
2017	7.434	4.065	17.727	26.310	9.824	5.528	27.892	4.200	-	-	-
2018	0.821	1.187	0.964	28.300	21.193	11.199	11.355	25.747	-	-	-
2019	4.139	5.276	8.869	19.380	8.771	0.186	11.054	0.850	3.750	5.276	4.911
2020	0.369	1.028	3.696	17.400	2.754	10.759	5.328	16.184	2.082	1.028	0.583
MAPE	3.848	3.606	12.572	25.800	9.491	7.193	15.763	12.435	2.916	3.152	2.747
지도 주성분 분석											
2016	23.272	6.476	19.819	36.490	27.987	13.115	14.826	14.999	-	-	-
2017	16.117	4.065	16.505	26.650	27.329	7.914	32.262	4.296	-	-	-
2018	14.837	1.187	6.364	24.930	24.499	20.232	0.449	25.931	-	-	-
2019	13.405	4.139	18.531	17.400	25.114	6.572	27.408	6.031	4.638	4.139	5.395
2020	0.369	1.752	6.600	17.960	9.528	0.891	29.725	23.662	1.289	1.752	2.698
MAPE	13.600	3.523	13.564	24.690	22.891	9.745	20.934	14.984	2.964	2.945	4.047

자료: 저자 작성

## (2) 항만업

항만업 부가가치 전망에 사용된 변수는 해운업과 동일하게 구성된다. 우선 머신러닝 기법들과 예측조합법을 사용한 전망치들은 아래 그림과 같다.

〈그림 4-14〉 항만업 부가가치 전망치 비교



자료: 저자 작성

2016년부터 2020년까지 항만업 부가가치를 살펴보면 코로나 사태에 따라 부가가치가 감소한 2020년을 제외하면 상승추세에 있는 것으로 보인다. 이러한 추세는 전반적인 항만물동량의 증가추세와도 관련이 깊다고 하겠다.<sup>64)</sup> 하지만 코로나 사태에 따른 항만 적체 심화, 운임 증대에 따른 수익성 악화는 2021년과 2022년의 항만업 부가가치의 증대에 다소 부정적인 요소로 작용할 것으로 예상되며, 이는 아래 표와 같이 각 예측기법의 2021년과 2022년 전망치가 다소 낮게 나타나는 것과 관련이 있다고 하겠다.

64) 해사신문(2019.02.01),

<http://www.haesanews.com/news/articleView.html?idxno=84122>(검색일: 2021.11.09)

〈표 4-12〉 항만업 부가가치 전망치

단위: 십억 원

연도	GDP	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수												
2016	2,334	2,226	5,968	5,115	5,258	3,238	5,044	5,047	4,520	-	-	-
2017	3,173	2,280	2,334	3,957	4,910	2,754	3,769	4,798	2,697	-	-	-
2018	3,249	2,753	3,173	4,609	4,732	2,145	2,519	4,662	1,254	-	-	-
2019	3,539	2,919	3,249	4,340	4,394	2,266	2,785	4,555	3,105	2,978	2,919	3,452
2020	3,292	3,394	3,539	4,168	4,423	1,918	2,408	4,053	2,498	3,226	3,539	3,300
2021	-	3,360	3,292	3,781	4,486	3,831	4,083	4,350	3,851	3,441	3,292	3,879
2022	-	3,203	2,821	4,590	4,350	4,191	3,802	4,155	4,095	3,354	2,821	3,901
랜덤포레스트 변수 중요도												
2016	2,334	5,090	2,226	5,061	5,272	4,554	5,044	5,124	2,258	-	-	-
2017	3,173	4,631	2,334	3,579	4,498	3,392	4,798	4,802	2,751	-	-	-
2018	3,249	2,753	3,173	5,012	4,540	2,342	2,000	4,665	1,239	-	-	-
2019	3,539	4,078	2,334	3,726	4,093	1,513	4,554	4,566	2,751	2,682	2,334	3,452
2020	3,292	3,394	3,539	3,943	4,344	1,870	2,339	4,599	2,253	3,373	3,539	3,285
2021	-	3,415	3,249	4,000	4,457	4,033	3,447	4,399	3,500	3,598	3,415	3,813
2022	-	3,495	2,334	3,808	4,415	4,077	4,038	4,359	4,096	3,693	3,495	3,828
지도 주성분 분석												
2016	2,334	5,090	2,226	4,995	5,207	4,079	5,044	4,976	2,269	-	-	-
2017	3,173	4,631	2,334	4,100	4,618	2,919	4,798	5,467	2,254	-	-	-
2018	3,249	4,423	3,173	4,518	3,845	3,228	3,979	4,669	3,099	-	-	-
2019	3,539	4,078	2,334	4,744	4,054	1,520	2,982	4,437	2,697	2,616	2,334	3,356
2020	3,292	3,394	3,539	4,434	4,373	1,898	2,234	4,369	1,750	3,044	1,750	3,249
2021	-	3,415	3,249	4,174	4,570	3,418	3,686	4,919	3,101	3,757	3,249	3,816
2022	-	3,504	2,334	6,349	4,489	4,058	4,456	4,337	8,237	4,105	3,504	4,721

자료: 저자 작성

아울러, 상기 전망치들을 토대로 2021년 및 2022년의 전망치의 신뢰구간은 아래 표와 같이 나타났다.



〈표 4-13〉 항만업 부가가치 전망치 신뢰구간

단위: 백만 원

연도	평균	중위값	90% 신뢰구간	95% 신뢰구간
2021	3,777,511	3,756,949	[3,640,152, 3,914,871]	[3,612,334, 3,942,689]
2022	4,037,955	4,058,401	[3,725,607, 4,350,302]	[3,662,351, 4,413,559]

자료: 저자 작성

연도별 각 예측기법의 절대비오차와 MAPE를 살펴보면, 전반적으로 각 예측기법은 그리 높은 예측성능을 보여주고 있지 않은 것으로 관찰된다. 모든 변수를 고려한 경우에서 머신러닝 기법 중에서는 kmeans가 가장 높은 예측성능을 보여주고 있으나, 부가가치의 변동이 다소 심한 2017년에서 2019년까지의 경우 예측성능이 낮은 것으로 나타났다. 예측조합법 중에서는 Fore3이 항만업 부가가치를 가장 잘 예측하고 있는 것으로 나타났다. Fore3는 다른 모든 예측기법과 비교해서도 가장 낮은 절대비오차를 보여주고 있는데, 이것은 예측조합법에서 오래된 기법인 Simple 예측조합법이 일반적으로 우수한 예측성능을 보여준다는 의견과 일치한다고 하겠다 (Smith and Wallis, 2009). 랜덤포레스트 변수 중요도를 활용한 경우에는 kNN이 가장 우수한 예측 성능을 보여주었다. kNN과 kmeans 기법의 유사성을 고려하면 해석이 용이한 부분이라고 하겠다. 예측조합법 중에서는 Fore3가 모든 변수를 사용한 경우와 마찬가지로 높은 예측성능을 보여주었다. 이는 높은 변동성의 전망치들을 대상으로 예측조합법을 적용 시에는 Simple 예측조합법이 우수한 예측을 보여준다는 직관을 다시금 뒷받침한다고 볼 수 있다. 지도 주성분 분석에 따라 변수를 선택한 경우의 결과는 랜덤포레스트 변수 중요도와 유사하게 머신러닝 기법에서는 kNN이 예측조합법 중에서는 Fore3이 각각 우수한 예측성능을 보여주었다. 종합적으로 보면 항만업의 경우에는 모든 변수를 선택 후 kmeans 기법을 사용하는 것이 예측성능을 높일 수 있는 방법이며, 예측조합법 중에서는 랜덤포레스트

트 변수 중요도에 따라 Fore3 예측조합법을 사용하는 것이 가장 높은 예측 성능을 보여준다고 볼 수 있다.

〈표 4-14〉 항만업 부가가치 예측성능 비교

단위: %

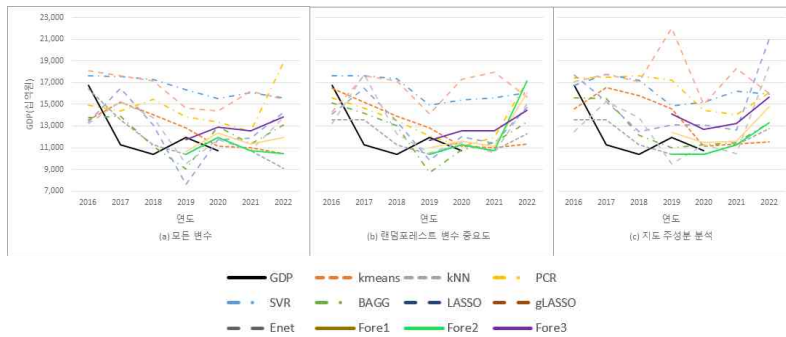
연도	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수											
2016	4.633	155.754	119.190	125.310	38.760	116.150	116.290	93.670	-	-	-
2017	28.162	26.458	24.710	54.740	13.220	18.760	51.220	15.020	-	-	-
2018	15.260	2.341	41.830	45.640	33.980	22.480	43.480	61.400	-	-	-
2019	17.523	8.181	22.650	24.160	35.970	21.290	28.720	12.260	15.847	17.520	2.463
2020	3.112	7.510	26.610	34.390	41.720	26.850	23.120	24.100	1.988	7.510	0.259
MAPE	13.738	40.049	47.000	56.850	32.730	41.110	52.570	41.290	8.917	12.520	1.361
랜덤포레스트 변수 중요도											
2016	118.135	4.633	116.852	125.910	95.128	116.150	119.560	3.231	-	-	-
2017	45.941	26.458	12.777	41.740	6.892	51.190	51.330	13.296	-	-	-
2018	15.260	2.341	54.257	39.710	27.934	38.460	43.570	61.863	-	-	-
2019	15.231	34.056	5.282	15.660	57.242	28.680	29.020	22.264	24.205	34.060	2.461
2020	3.112	7.510	19.784	31.990	43.184	28.930	39.710	31.543	2.488	7.510	0.194
MAPE	39.536	15.000	41.791	51.000	46.076	52.680	56.640	26.439	13.346	20.780	1.328
지도 주성분 분석											
2016	118.135	4.633	114.040	123.130	74.778	116.150	113.210	2.791	-	-	-
2017	45.941	26.458	29.200	45.540	8.023	51.190	72.300	28.952	-	-	-
2018	36.115	2.341	39.060	18.340	0.640	22.460	43.680	4.634	-	-	-
2019	15.231	34.056	34.060	14.550	57.038	15.730	25.370	23.798	26.070	34.060	5.176
2020	3.112	7.510	34.700	32.850	42.324	32.130	32.740	46.826	7.526	46.830	1.296
MAPE	43.707	15.000	50.210	46.880	36.560	47.530	57.460	21.400	16.798	40.440	3.236

자료: 저자 작성

### (3) 조선업

조선업 부가가치 전망에 사용된 변수는 해운업 및 항만업과 동일하게 구성된다. 우선 머신러닝 기법들과 예측조합법을 사용한 전망치들은 아래 그림과 같다.

〈그림 4-15〉 조선업 부가가치 전망치 비교



자료: 저자 작성

2016년부터 2020년까지 조선업 부가가치를 일정한 추세를 보이기보다는 상승과 하락을 반복하는 것으로 나타났다. 그러므로 2020년 다소 하락한 것으로 나타나는 2020년 조선업 부가가치도 코로나 사태의 영향만으로는 판단할 수 없는 것으로 보인다. 다만, 코로나 사태와 관련된 외부환경이 개선됨에 따라 조선업 부가가치의 변동성은 상승추세로 선회할 수도 있을 것으로 보인다. 이에 따라, 아래 표에 제시된 많은 예측기법들이 2021년 이후 증가한 조선업 부가가치 전망치를 보여주고 있다.

〈표 4-15〉 조선업 부가가치 전망치

단위: 십억 원

연도	GDP	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수												
2016	16,749	13,602	16,439	14,957	17,653	13,772	13,344	18,117	13,245	-	-	-
2017	11,290	15,175	13,602	14,402	17,575	13,911	16,470	17,622	15,256	-	-	-
2018	10,377	14,020	11,290	15,465	17,285	11,131	13,128	17,137	13,750	-	-	-
2019	11,954	12,805	10,377	13,839	16,329	9,017	7,630	14,684	9,525	10,682	10,377	11,776
2020	10,723	11,165	11,954	13,399	15,537	12,947	11,724	14,414	11,751	12,349	11,954	12,861
2021	-	11,018	10,723	12,455	16,073	11,322	11,901	16,221	10,751	11,326	10,723	12,558
2022	-	10,488	9,133	18,860	15,594	13,116	14,261	15,457	13,743	11,925	10,488	13,831
랜덤포레스트 변수 중요도												
2016	16,749	16,496	13,602	15,628	17,602	15,141	14,107	14,290	13,256	-	-	-
2017	11,290	15,175	13,602	14,647	17,603	14,202	16,466	17,628	17,755	-	-	-
2018	10,377	13,880	11,290	13,506	17,330	13,018	13,414	17,179	12,259	-	-	-
2019	11,954	12,805	10,377	12,061	14,940	8,729	9,873	14,141	10,504	10,970	10,377	11,679
2020	10,723	11,165	11,290	11,153	15,373	10,801	11,983	17,298	11,491	11,617	11,290	12,569
2021	-	11,018	10,723	11,938	15,582	11,499	11,386	17,972	10,512	11,168	10,723	12,579
2022	-	11,337	12,294	17,128	16,085	13,357	14,728	15,652	15,028	16,088	17,128	14,451
지도 주성분 분석												
2016	16,749	14,567	13,602	17,415	16,754	15,604	17,711	17,102	12,518	-	-	-
2017	11,290	16,538	13,602	17,465	17,748	15,536	15,266	17,757	15,234	-	-	-
2018	10,377	15,788	11,290	17,599	17,205	12,177	12,456	17,121	13,757	-	-	-
2019	11,954	14,619	10,377	17,214	14,876	10,981	13,086	22,001	9,503	12,428	10,377	14,082
2020	10,723	11,165	10,377	14,433	15,197	11,261	13,078	14,963	11,239	11,506	10,377	12,714
2021	-	11,339	11,290	14,070	16,181	11,456	12,624	18,337	10,482	11,639	11,290	13,222
2022	-	11,563	12,750	16,099	15,959	13,328	21,015	15,998	18,516	14,814	13,328	15,654

자료: 저자 작성

아울러, 상기 전망치들을 토대로 2021년 및 2022년의 전망치의 신뢰구간은 아래 표와 같이 나타났다.

〈표 4-16〉 조선업 부가가치 전망치 신뢰구간

단위: 백만 원

연도	평균	중위값	90% 신뢰구간	95% 신뢰구간
2021	12,488,003	11,456,178	[11,832,079, 13,143,928]	[11,699,241, 13,276,765]
2022	14,521,073	14,728,174	[13,766,112, 15,276,034]	[13,613,218, 15,428,928]

자료: 저자 작성

아래 표는 조선업 부가가치에 관한 각 예측기법의 절대비오차와 MAPE를 나타낸다. 모든 변수를 고려한 경우, 머신러닝 기법 중에서는 kNN이 가장 높은 예측성능을 보여주고 있다. kNN이 근접한 관측치를 참고하여 예측을 하는 기법이라는 점을 고려하면, 2017년과 같이 부가가치가 급변하는 경우에 예측성능이 하락하는 점은 설명이 가능하다고 하겠다. 예측조합법 중에서는 Fore3가 조선업 부가가치를 가장 잘 전망하는 것으로 나타났다으며, 개별 머신러닝 기법과 비교해도 예측성능이 뛰어난 것으로 나타났다. 랜덤포레스트 변수 중요도를 활용한 경우에는 kNN 기법의 예측성능이 가장 우수한 것으로 나타났으며, 예측조합법 중에서는 Fore1이 높은 예측성능을 보여주었으며, 이것은 개별 머신러닝 기법보다 높은 예측성능으로 나타났다. 지도 주성분 분석에 따라 변수를 선택한 경우에서도 kNN이 가장 높은 예측성능을 보였다. 머신러닝 기법 중 kNN이 계속적으로 높은 예측력을 보여주는 이유는 kNN은 근처의 자료들을 참조하는 머신러닝 기법이며, 2017년에 전년대비 부가가치가 많이 감소 후 비슷한 수준으로 유지되기 때문이라고 추정된다. 예측조합법 중에서는 이전과 마찬가지로 Fore1이 각각 우수한 예측성능을 보여주었다. 결론적으로 조선업의 경우에는 모든 변수를 선택 후, kNN을 사용하는 것이 예측성능이 높았으며, 예측조합법의 경우에는 지도주성분 분석을 사용하여 Fore1 예측조합법을 사용하는 것이 가장 높은 예측성능을 보였다.

〈표 4-17〉 조선업 부가가치 예측성능 비교

단위: %

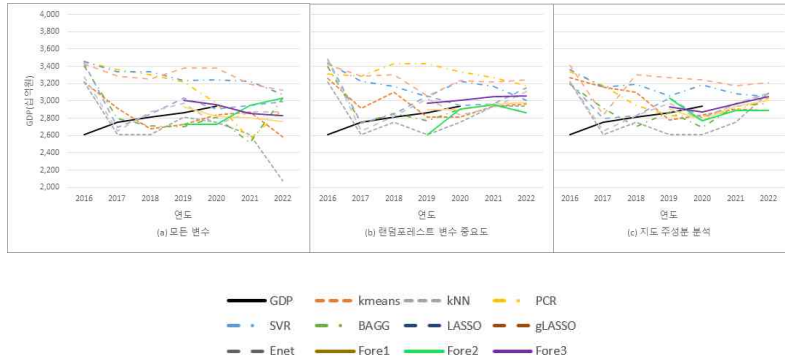
연도	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수											
2016	18.791	1.851	10.700	5.396	17.774	20.330	8.172	20.918	-	-	-
2017	34.408	20.470	27.560	55.664	23.211	45.873	56.080	35.125	-	-	-
2018	35.109	8.807	49.040	66.577	7.275	26.520	65.149	32.512	-	-	-
2019	7.118	13.199	15.760	36.593	24.572	36.171	22.837	20.325	10.640	13.200	1.495
2020	4.122	11.479	24.950	44.888	20.736	9.333	34.420	9.586	15.160	11.480	19.939
MAPE	19.910	11.161	25.600	41.824	18.713	27.645	37.331	23.693	12.900	12.340	10.717
랜덤포레스트 변수 중요도											
2016	1.509	18.791	6.693	5.097	9.597	15.780	14.680	20.852	-	-	-
2017	34.408	20.470	29.728	55.910	25.785	45.840	56.140	57.259	-	-	-
2018	33.766	8.807	30.164	67.013	25.456	29.270	65.550	18.139	-	-	-
2019	7.118	13.199	0.891	24.976	26.981	17.410	18.290	12.135	8.237	13.199	2.306
2020	4.122	5.287	4.007	43.357	0.721	11.750	61.310	7.162	8.334	5.287	17.214
MAPE	16.185	13.311	14.297	39.271	17.708	24.010	43.200	23.109	8.286	9.243	9.760
지도 주성분 분석											
2016	13.028	18.791	3.979	0.029	6.837	5.747	2.106	25.262	-	-	-
2017	46.480	20.470	54.687	57.195	37.608	35.213	57.277	34.927	-	-	-
2018	52.156	8.807	69.603	65.808	17.348	20.043	65.000	32.579	-	-	-
2019	22.289	13.199	43.999	24.439	8.141	9.468	84.041	20.504	3.961	13.199	17.800
2020	4.122	3.235	34.594	41.721	5.009	21.959	39.532	4.808	7.295	3.235	18.560
MAPE	27.615	12.900	41.373	37.838	14.989	18.486	49.591	23.616	5.628	8.217	18.180

자료: 저자 작성

#### (4) 수산업

수산업(수산물 생산업) 부가가치 전망에 사용된 변수는 산업연관표 항목 및 거시경제변수 외에 어가수, 어가인구 등의 수산업과 관련된 산업별 고유변수들로 구성된다. 머신러닝 기법들과 예측조합법을 사용한 전망치들은 아래 그림과 같다.

〈그림 4-16〉 수산업 부가가치 전망치 비교



자료: 저자 작성

2016년부터 2020년까지 수산업 부가가치는 점진적으로 증가하는 추세를 보이고 있다. 하지만 2017년 부가가치 증가율이 5.4%인데 반해 2020년 부가가치 증가율은 2.4%에 그쳐 수산업 분야도 코로나 사태를 피해갈 수는 없었던 것으로 보인다. 하지만 웰빙 문화와 더불어 수산물 수요는 전반적으로 증대될 가능성이 크며, 아래와 같이 많은 예측기법들이 증가하는 추세의 수산업 부가가치 전망치를 보여주고 있다.

〈표 4-18〉 수산업 부가가치 전망치

단위: 십억 원

연도	GDP	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수												
2016	2,610	3,220	3,220	3,458	3,457	3,405	3,442	3,431	3,275	-	-	-
2017	2,752	2,915	2,610	3,361	3,341	2,798	2,694	3,285	2,650	-	-	-
2018	2,812	2,681	2,610	3,306	3,335	2,713	2,845	3,255	2,875	-	-	-
2019	2,867	2,725	2,812	3,221	3,239	2,699	3,035	3,379	2,977	2,927	2,725	3,011
2020	2,937	2,840	2,752	2,976	3,248	2,813	2,906	3,379	2,726	2,817	2,726	2,955
2021	-	2,872	2,610	2,562	3,230	2,517	2,950	3,198	2,876	2,808	2,950	2,852
2022	-	2,584	2,074	2,896	3,077	3,029	2,995	3,123	2,874	2,759	3,029	2,832

랜덤포레스트 변수 중요도												
2016	2,610	3,258	3,220	3,316	3,439	3,398	3,478	3,426	3,475	-	-	-
2017	2,752	2,915	2,610	3,289	3,231	2,748	2,746	3,289	2,650	-	-	-
2018	2,812	3,099	2,752	3,427	3,171	2,856	2,851	3,307	2,824	-	-	-
2019	2,867	2,812	2,610	3,430	3,053	2,769	3,055	3,060	3,025	2,900	2,610	2,977
2020	2,937	2,810	2,752	3,341	3,224	2,905	2,948	3,236	2,841	2,897	2,905	3,007
2021	-	2,937	2,937	3,272	3,168	2,955	2,966	3,219	2,925	2,961	2,955	3,047
2022	-	2,968	2,937	3,176	3,009	2,864	3,150	3,246	3,105	2,979	2,864	3,057
지도 주성분 분석												
2016	2,610	3,267	3,220	3,334	3,362	3,197	3,212	3,414	3,224	-	-	-
2017	2,752	3,157	2,610	3,174	3,161	2,916	2,795	2,842	2,649	-	-	-
2018	2,812	3,099	2,752	2,950	3,190	2,704	2,831	3,302	2,827	-	-	-
2019	2,867	2,782	2,610	2,834	3,059	2,874	3,031	3,268	2,976	2,905	3,031	2,929
2020	2,937	2,840	2,610	2,818	3,189	2,698	2,807	3,241	2,774	2,775	2,774	2,872
2021	-	2,902	2,752	2,889	3,087	2,937	2,968	3,180	2,976	2,937	2,889	2,961
2022	-	2,890	3,090	3,029	3,051	2,972	3,083	3,212	3,090	3,009	2,890	3,052

자료: 저자 작성

아울러, 상기 전망치들을 토대로 2021년 및 2022년의 전망치의 신뢰구간은 아래 표와 같이 나타났다.

〈표 4-19〉 수산업 부가가치 전망치 신뢰구간

단위: 백만 원

연도	평균	중위값	90% 신뢰구간	95% 신뢰구간
2021	2,946,760	2,950,013	[2,894,878, 2,998,641]	[2,884,371, 3,009,148]
2022	2,969,598	3,008,968	[2,907,805, 3,031,391]	[2,895,290, 3,043,906]

자료: 저자 작성

수산업 부가가치의 절대비오차와 MAPE는 아래 표와 같다. 모든 변수를 고려한 경우, 머신러닝 기법 중에서는 LASSO 기법이 가장 높은 예측성능을 보여주고 있다. Enet도 비교적 높은 예측성능을 보여주고 있는데 LASSO와 Enet이 선형모델을 사용하고 있다는 점에서 해석이 가능한 부분



이라고 할 수 있다. 예측조합법 중에서는 Fore3가 수산업 부가가치를 가장 잘 전망하는 것으로 나타났으며 Fore1도 비견할만한 예측성능을 보여주고 있다. 하지만 두 예측조합법의 연도별 절대비오차는 차이가 있는데 Fore1의 경우 2019년과 2020년 비교적 비슷한 절대비오차를 보여주고 있으나, Fore3의 경우에는 2019년 대비 2020년 절대비오차가 상당히 적은 것으로 나타났다. 두 예측조합법 모두 각 예측기법들에 가중치를 둔다는 점에서 안정적인 예측조합법들이라고 볼 수 있으나 Fore1이 이전의 전망치들을 사용하여 분산을 최소화한다는 점에서 Fore3보다 상대적으로 안정성을 중시하는 방법이라고 볼 수 있다. 랜덤포레스트 변수 중요도를 활용한 경우에는 BAGG의 예측성능이 높게 나타났다. 또한, 모든 변수를 고려한 경우와 마찬가지로 LASSO와 Enet의 예측성능도 높게 나타났다. 예측조합법 중에서는 Fore1이 높은 예측성능을 보여주었다. 지도 주성분 분석에 따라 변수를 선택한 경우, 머신러닝 기법에서는 LASSO 기법이 가장 정확하였으며, 예측조합법에서는 Fore3가 우수한 정확한 예측성능을 보여주었다. 종합적으로 머신러닝 기법에서는 지도 주성분 분석에 따른 LASSO가 예측성능이 높으며 예측조합법 중에서는 랜덤포레스트 변수 중요도에 따른 Fore1이 높은 예측성능을 보이는 기법으로 나타났다.

〈표 4-20〉 수산업 부가가치 예측성능 비교

단위: %

연도	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수											
2016	23.380	23.380	32.498	32.440	30.439	31.870	31.430	25.481	-	-	-
2017	5.937	5.151	22.114	21.400	1.656	2.100	19.380	3.698	-	-	-
2018	4.656	7.176	17.566	18.610	3.527	1.155	15.770	2.240	-	-	-
2019	4.965	1.921	12.347	12.980	5.848	5.844	17.870	3.823	2.101	4.965	5.016
2020	3.323	6.305	1.333	10.600	4.239	1.062	15.050	7.185	4.077	7.185	0.608
MAPE	8.452	8.787	17.172	19.210	9.142	8.406	19.900	8.486	3.089	6.075	2.812

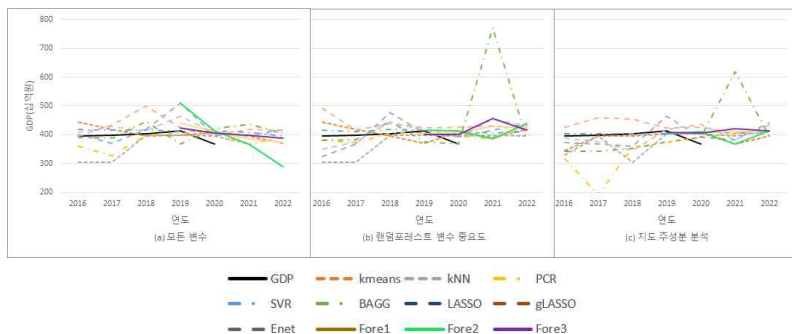
랜덤포레스트 변수 중요도											
2016	24.801	23.380	27.040	31.751	30.167	33.229	31.246	33.131	-	-	-
2017	5.937	5.151	19.520	17.400	0.133	0.208	19.511	3.690	-	-	-
2018	10.221	2.136	21.860	12.754	1.550	1.379	17.609	0.435	-	-	-
2019	1.921	8.959	19.620	6.497	3.409	6.555	6.712	5.507	1.139	8.959	3.825
2020	4.317	6.305	13.760	9.763	1.110	0.359	10.166	3.274	1.379	1.110	2.381
MAPE	9.440	9.186	20.360	15.633	7.274	8.346	17.049	9.207	1.259	5.035	3.103
지도 주성분 분석											
2016	25.153	23.380	27.732	28.801	22.479	23.037	30.795	23.529	-	-	-
2017	14.731	5.151	15.348	14.861	5.977	1.563	3.259	3.740	-	-	-
2018	10.221	2.136	4.893	13.427	3.828	0.668	17.414	0.526	-	-	-
2019	2.968	8.959	1.168	6.690	0.248	5.720	13.988	3.782	1.335	5.720	2.167
2020	3.323	11.131	4.067	8.575	8.136	4.434	10.361	5.541	5.507	5.541	2.212
MAPE	11.279	10.151	10.642	14.471	8.134	7.084	15.163	7.424	3.421	5.631	2.189

자료: 저자 작성

## (5) 해양관광업

해양관광업 부가가치 전망에 사용된 변수는 산업연관표 항목 및 거시경제변수 외에 크루즈 관광객수, 연강수량 등의 해양관광업과 관련된 산업별 고유변수들로 구성된다. 머신러닝 기법들과 예측조합법을 사용한 전망치들은 아래 그림과 같다.

〈그림 4-17〉 해양관광업 부가가치 전망치 비교



자료: 저자 작성

해양관광업 부가가치의 경우 코로나 사태의 영향이 뚜렷하게 드러난다. 즉 2019년까지 완만한 증가세를 보이던 부가가치가 2020년 상당히 많이 감소한 것이다.

〈표 4-21〉 해양관광업 부가가치 전망치

단위: 십억 원

연도	GDP	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수												
2016	394	444	304	360	420	391	401	395	409	-	-	-
2017	399	419	304	326	417	389	371	434	428	-	-	-
2018	405	397	399	398	415	447	418	500	413	-	-	-
2019	413	399	399	405	414	368	510	424	465	442	510	423
2020	369	409	394	404	413	421	399	397	405	410	413	405
2021	-	396	369	368	409	437	398	419	393	385	369	398
2022	-	369	290	391	396	405	417	416	412	374	290	387
랜덤포레스트 변수 중요도												
2016	394	444	304	383	416	380	325	493	350	-	-	-
2017	399	419	304	370	411	384	369	415	375	-	-	-
2018	405	397	399	405	417	447	476	439	445	-	-	-
2019	413	370	399	405	415	375	406	425	415	403	415	401
2020	369	409	394	395	414	369	393	426	398	407	414	400
2021	-	391	399	385	406	771	417	430	455	429	385	457
2022	-	418	394	427	412	384	439	421	431	425	439	416
지도 주성분 분석												
2016	394	344	330	318	404	343	373	427	389	-	-	-
2017	399	394	394	190	403	343	368	459	375	-	-	-
2018	405	397	304	357	402	354	360	455	345	-	-	-
2019	413	405	405	373	403	376	464	424	417	405	403	408
2020	369	405	405	393	411	390	391	426	435	409	411	407
2021	-	369	394	404	405	620	384	407	383	412	369	421
2022	-	395	413	417	409	392	441	410	433	415	413	414

자료: 저자 작성

아울러, 상기 전망치들을 토대로 2021년 및 2022년의 전망치의 신뢰구간은 아래 표와 같이 나타났다.

〈표 4-22〉 해양관광업 부가가치 전망치 신뢰구간

단위: 백만 원

연도	평균	중위값	90% 신뢰구간	95% 신뢰구간
2021	419,144	398,590	[396,393, 441,895]	[391,785, 446,503]
2022	403,166	413,180	[393,093, 413,239]	[391,053, 415,279]

자료: 저자 작성

해양관광업 부가가치의 절대비오차와 MAPE는 아래 표와 같다. 모든 변수를 고려한 경우, 머신러닝 기법 중에서는 SVR 기법이 가장 높은 예측성능을 보여주고 있다. 이것은 2019년도의 낮은 절대비오차에 많은 부분 기인하고 있는 것으로 보인다. 예측조합법 중에서는 Fore3가 수산업 부가가치를 가장 잘 전망하는 것으로 나타났다. 랜덤포레스트 변수 중요도를 활용한 경우에는 PCR의 예측성능이 높게 나타났다. SVR도 비교적 정확한 예측성능을 보여주었다. 예측조합법 중에서는 Fore3가 높은 예측성능을 보여주었다. 지도 주성분 분석에 따라 변수를 선택한 경우 SVR이 가장 높은 예측성능을 나타내었다. 예측조합법에서는 Fore3가 우수한 정확한 예측성능을 보여주었다. 해양관광업의 경우 지도 주성분 분석에 따른 변수 선택기법을 사용하여 SVR을 사용하는 것이 예측기법이 높으며, 예측조합법으로는 랜덤포레스트 변수 중요도에 따라 변수를 선별 후 Fore3를 사용하는 것이 예측성능이 높은 것으로 판단된다.

〈표 4-23〉 해양관광업 부가가치 예측성능 비교

단위: %

연도	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수											
2016	12.681	22.827	8.686	6.374	0.956	1.769	0.041	3.600	-	-	-
2017	5.240	23.625	18.159	4.596	2.373	7.017	8.898	7.257	-	-	-
2018	1.973	1.463	1.687	2.658	10.604	3.216	23.687	1.977	-	-	-
2019	3.386	3.531	2.076	0.157	11.011	23.522	2.542	12.514	7.007	23.520	2.341
2020	10.790	6.893	9.364	12.010	14.148	8.153	7.706	9.732	11.219	12.010	9.849
MAPE	6.814	11.668	7.994	5.159	7.818	8.735	8.575	7.016	9.113	17.770	6.095
랜덤포레스트 변수 중요도											
2016	12.681	22.827	2.958	5.513	3.700	17.612	24.964	11.221	-	-	-
2017	5.240	23.625	7.102	3.223	3.715	7.522	4.236	5.941	-	-	-
2018	1.973	1.463	0.082	3.154	10.530	17.583	8.554	9.998	-	-	-
2019	10.355	3.531	1.963	0.426	9.204	1.739	2.762	0.448	2.551	0.426	2.894
2020	10.790	6.893	7.148	12.167	0.037	6.617	15.367	7.740	10.338	12.167	8.336
MAPE	8.208	11.668	3.851	4.897	5.437	10.215	11.177	7.070	6.445	6.297	5.615
지도 주성분 분석											
2016	12.848	16.466	19.445	2.387	13.153	5.548	8.239	1.350	-	-	-
2017	1.035	1.035	52.279	1.012	14.007	7.741	15.199	5.902	-	-	-
2018	1.973	24.743	11.702	0.529	12.572	11.070	12.395	14.721	-	-	-
2019	2.099	2.099	9.733	2.384	9.112	12.191	2.551	1.042	1.908	2.384	1.205
2020	9.863	9.615	6.387	11.439	5.566	5.957	15.329	17.903	10.802	11.439	10.257
MAPE	5.563	10.791	19.909	3.550	10.882	8.501	10.743	8.184	6.355	6.912	5.731

자료: 저자 작성

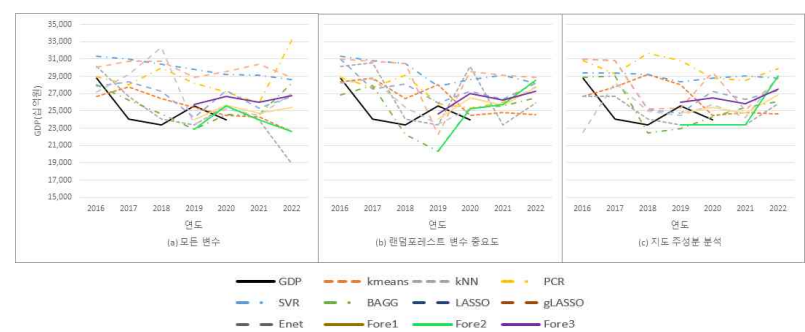
## 2) 전체 산업 부가가치 전망

해운업, 항만업, 조선업, 수산업, 해양관광업을 종합한 해양수산업 전체 산업 부가가치 전망은 두 가지 방식으로 수행되었다. 첫 번째는 개별 산업 부가가치 전망과 동일한 방법으로 수행되었으며, 두 번째는 개별 산업의 각 연도의 절대비오차가 가장 적은 경우의 변수 선택 방식과 예측기법을 선택하여 예측성능을 검증하였다.

우선 개별 산업 부가가치의 합을 종속변수로 하고 공통되는 독립변수와

산업별 고유 변수들을 모두 독립변수로 설정하여 머신러닝 기법을 적용하였다. 머신러닝 기법들과 예측조합법을 사용한 전망치들은 아래 그림과 같다.

〈그림 4-18〉 전 산업 부가가치 전망치 비교



자료: 저자 작성

전체 산업 부가가치의 경우 2017년 해운업 관련 악재와 코로나 사태가 여실히 반영된 것으로 나타났다. 이처럼 전체 산업 부가가치가 2차례 하락하는 것은 향후 전망에 부정적인 영향을 미쳐 아래 표에서와 같이 큰 폭으로 증대된 부가가치 전망치는 관찰되지 않았다.

〈표 4-24〉 전 산업 부가가치 전망치

단위: 십억 원

연도	GDP	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수												
2016	28,828	26,669	30,130	28,989	31,316	28,071	27,875	30,106	27,214	-	-	-
2017	24,092	27,748	26,669	28,047	31,017	26,279	28,377	30,712	29,237	-	-	-
2018	23,397	26,460	24,092	29,989	30,438	24,663	27,252	30,725	32,386	-	-	-
2019	25,612	25,439	23,397	28,236	29,838	22,857	24,213	28,863	23,503	23,996	22,857	25,793
2020	23,993	24,505	25,612	27,187	29,260	24,540	27,381	29,579	25,247	25,664	25,612	26,664
2021	-	24,334	23,993	25,930	29,110	24,598	25,358	30,425	24,492	24,725	23,993	26,030
2022	-	22,647	18,931	33,229	28,602	28,167	26,674	28,914	27,027	25,450	22,647	26,774

랜덤포레스트 변수 중요도												
2016	28,828	28,400	30,130	28,883	31,302	26,864	31,012	31,034	28,235	-	-	-
2017	24,092	28,828	30,585	27,679	30,842	27,929	27,496	30,754	28,750	-	-	-
2018	23,397	26,460	24,092	29,094	30,505	22,263	28,146	30,454	25,428	-	-	-
2019	25,612	28,001	23,397	25,823	27,865	20,381	25,920	22,284	23,489	24,095	20,381	24,645
2020	23,993	24,505	30,130	25,361	28,601	25,292	27,315	29,596	25,260	26,519	25,260	27,008
2021	-	24,802	23,397	26,000	29,113	25,610	26,355	29,100	25,755	25,700	25,755	26,267
2022	-	24,598	25,898	28,509	28,176	26,516	28,166	28,866	27,753	27,695	28,509	27,310
지도 주성분 분석												
2016	28,828	26,669	26,669	30,860	29,415	28,915	31,012	31,028	22,512	-	-	-
2017	24,092	27,748	26,669	29,320	29,415	29,045	30,814	30,832	28,251	-	-	-
2018	23,397	29,252	24,092	31,634	29,230	22,428	25,192	25,216	25,024	-	-	-
2019	25,612	28,001	23,397	30,795	28,404	22,952	24,744	25,305	24,509	24,760	23,397	26,013
2020	23,993	24,505	23,397	28,775	28,779	24,333	27,299	29,438	25,752	25,391	23,397	26,535
2021	-	24,802	23,397	28,545	29,035	25,453	26,338	25,099	24,245	24,742	23,397	25,864
2022	-	24,637	25,876	29,906	28,820	26,128	27,344	28,809	29,033	26,891	29,033	27,569

자료: 저자 작성

아울러, 상기 전망치들을 토대로 2021년 및 2022년의 전망치의 신뢰구간은 아래 표와 같이 나타났다.

〈표 4-25〉 전 산업 부가가치 전망치 신뢰구간

단위: 백만 원

연도	평균	중위값	90% 신뢰구간	95% 신뢰구간
2021	25,811,034	25,609,940	[25,264,772, 26,357,295]	[25,154,144, 26,467,923]
2022	27,124,374	27,569,219	[26,383,618, 27,865,131]	[26,233,601, 28,015,148]

자료: 저자 작성

전체 산업 부가가치의 절대비오차와 MAPE는 아래 표와 같다. 모든 변수를 고려한 경우, 머신러닝 기법 중에서는 BAGG 기법이 가장 높은 예측성능을 보여주고 있다. 두 번째로 높은 예측성능을 보이는 기법은 선형모델이 아닌 kNN 기법으로 나타났다. 예측조합법 중에서는 Fore3가 예측성능이 가장 높았다. 랜덤포레스트 변수 중요도를 활용한 경우에는 Enet이 높

은 예측성능을 보이는 것으로 나타났다. 예측조합법 중에서는 Fore3가 가장 높은 예측성능을 보여주었다. 지도 주성분 분석의 경우 kNN 기법이 가장 높은 예측성능을 나타내었으며, BAGG 기법이 그 뒤를 따랐다. 예측조합법에서는 Fore1이 가장 높은 예측성능을 보여주었다. 전체 산업의 경우에는 모든 변수를 사용하여 BAGG 기법을 사용하는 것이 예측성능이 높으며, 예측조합법 중에서는 지도 주성분 분석에 따른 변수들로 Fore1을 사용하는 것이 예측성능이 우수한 것으로 나타났다.

〈표 4-26〉 전 산업 부가가치 예측성능 비교

단위: %

연도	kmeans	kNN	PCR	SVR	BAGG	LASSO	gLASSO	Enet	Fore1	Fore2	Fore3
모든 변수											
2016	7.488	4.518	0.560	8.633	2.623	3.304	4.432	5.598	-	-	-
2017	15.180	10.700	16.420	28.747	9.080	17.788	27.480	21.356	-	-	-
2018	13.088	2.966	28.173	30.093	5.409	16.472	31.316	38.417	-	-	-
2019	0.674	8.645	10.245	16.501	10.755	5.462	12.695	8.235	6.307	10.755	0.709
2020	2.131	6.745	13.313	21.952	2.280	14.119	23.280	5.226	6.962	6.745	11.131
MAPE	7.712	6.715	13.742	21.185	6.029	11.429	19.841	15.766	6.634	8.750	5.920
랜덤포레스트 변수 중요도											
2016	1.485	4.518	0.191	8.584	6.812	7.578	7.652	2.058	-	-	-
2017	19.659	26.954	14.889	28.019	15.929	14.133	27.654	19.337	-	-	-
2018	13.088	2.966	24.345	30.376	4.850	20.295	30.160	8.677	-	-	-
2019	9.331	8.645	0.827	8.797	20.423	1.203	12.992	8.287	5.922	20.420	3.774
2020	2.131	25.579	5.701	19.205	5.412	13.845	23.353	5.280	10.527	5.280	12.563
MAPE	9.139	13.733	9.190	18.996	10.685	11.411	20.362	8.728	8.225	12.850	8.168
지도 주성분 분석											
2016	7.488	7.488	7.049	2.036	0.303	7.578	7.631	21.909	-	-	-
2017	15.180	10.700	21.704	22.099	20.562	27.903	27.979	17.264	-	-	-
2018	25.022	2.966	35.202	24.927	4.145	7.671	7.772	6.950	-	-	-
2019	9.331	8.645	20.238	10.901	10.385	3.388	1.197	4.303	3.326	8.645	1.569
2020	2.131	2.483	19.929	19.946	1.417	13.780	22.693	7.330	5.824	2.483	10.593
MAPE	11.830	6.456	20.824	15.982	7.362	12.064	13.454	11.551	4.575	5.564	6.081

자료: 저자 작성



다음으로 개별 산업의 각 연도의 절대비오차가 가장 적은 경우의 변수 선택 방식과 예측기법을 선택하여 전체 산업의 부가가치를 전망한 경우를 살펴보고자 한다. 우선 상기 개별 산업에서 2019년과 2020년 각 연도별로 가장 적은 절대비오차를 나타내는 변수 선택 방식과 예측기법을 선택하여 부가가치를 전망한 수치는 아래 표와 같다. 이 경우에는 이전 개별 산업 분석의 결과와 다소 다른 결과를 보이는 데, 아래 표는 MAPE가 아닌 해당 연도의 절대비오차만을 고려한 결과이기 때문이다. 충분한 데이터가 주어지는 경우에는 MAPE를 얻을 수 있으므로 이전 개별 산업 분석과 같이 MAPE를 기준으로 변수 선택 방식과 예측기법을 선택하는 것이 바람직하나, 2019년의 경우 2개 이상의 관측치로 MAPE를 생성할 수 없으므로 절대비오차만을 고려하여 아래와 같은 결과를 산출하였다. 2021년과 2022년의 경우에는 2020년의 변수 선택 방식과 예측기법을 기준으로 전망치를 산출하였다.

〈표 4-27〉 개별산업 부가가치 전망치(최저 절대비오차 기준)

단위: 백만 원

연도	해운업	항만업	조선업	수산업	해양관광업
2019	6,838,223 (RF/LASSO)	3,469,568 (RF/Fore3)	11,164,168 (RF/PCR)	2,867,089 (SPC/BAGG)	412,949 (ALL/SVR)
2020	6,672,072 (ALL/kNN)	3,274,513 (RF/Fore3)	10,941,951 (RF/BAGG)	2,937,146 (RF/LASSO)	381,286 (RF/BAGG)
2021	5,758,671 (ALL/kNN)	3,707,119 (RF/Fore3)	12,368,334 (RF/BAGG)	2,836,659 (RF/LASSO)	407,577 (RF/BAGG)
2022	5,392,660 (ALL/kNN)	3,915,575 (RF/Fore3)	13,267,952 (RF/BAGG)	3,161,127 (RF/LASSO)	396,055 (RF/BAGG)

주: ALL(모든 변수), RF(랜덤포레스트 변수 중요도), SPC(지도 주성분 분석), 2021년과 2022년은 2020년 예측방법을 따름

자료: 저자 작성

상기 결과를 바탕으로 2019년부터 2022년까지 전체 산업의 부가가치와 절대비오차를 구하면 아래 표와 같다.

〈표 4-28〉 전 산업 부가가치 전망치 및 절대비오차

단위: 백만 원

연도	전망 부가가치	실제 부가가치	절대비오차
2019	25,626,122	25,611,623	0.057
2020	24,143,108	23,993,206	0.625
2021	25,525,284	-	-
2022	25,145,428	-	-

자료: 저자 작성

상기 표와 같이 개별 산업 부가가치 전망치의 합을 사용하여 전체 산업 부가가치를 추정하는 방식은 첫 번째 접근 방식보다 정확한 예측성능을 보여준다고 보인다. 첫 번째 접근 방식의 경우는 각 산업별 특성을 고려하지 않고 모든 산업의 고유한 독립변수를 활용하여 너무 많은 독립변수가 구성되었고, 이 과정에서 특정 산업과 무관한 독립변수가 포함되는 등의 요인들이 작용했다고 보인다. 그러므로 현재 분석 결과만을 고려한다면, 첫 번째 접근 방식을 활용하여 연속적인 기간에서 지속적으로 높은 예측성능을 주는 변수 선택방식과 예측기법이 관찰되지 않는 한, 두 번째 접근 방식을 따르는 것이 바람직해 보인다.

# 05

## 결론

### 제1절 요약 및 결론

---

#### 1. 요약

2022년 해양수산업 전망모형 개발을 위해 올해는 해양수산업 데이터베이스 구축 자동화와 베이지안 VAR 모형 및 머신러닝을 통한 해양수산업 전망을 수행하였다.

해양수산업 데이터베이스 자동화는 기존의 데이터베이스 구축과정이 많은 시간이 소요되는 엑셀(Excel)을 기반으로 작업이 이루어져 해양수산업 전망모형 개발과 결과 분석 등에 적절한 자원이 분배되지 못하는 문제가 있었다. 이에 따라, 본 연구에서는 통계패키지인 R을 활용하여 해양수산업 데이터베이스 구축을 자동화하였다. 자동화는 input-output 형식의 프로그램으로 데이터베이스 구축에 필요한 파일을 입력하면 해양수산업 부가가치 데이터베이스가 즉시 도출된다.

해양수산업 데이터베이스를 활용하여 2022 해양수산업 부가가치 전망을 수행하였다. 다모형 접근법을 기반으로 하여 시계열모형인 베이지안

---

VAR 모형과 다양한 머신러닝 기법을 적용하여 부가가치를 전망하였다.

베이지안 VAR를 통한 전망작업은 해양수산업을 4개 부문으로 구분하여 모형을 구축하였고, 모형마다 공통적으로 영향을 받는 경제지표를 추가하여 모형의 현실 설명력을 확보하고자 하였다. 주요 경제변수는 한국해양수산개발원의 연구진의 협의와 선행연구 고찰을 통해 선별하였다. 2022 해양수산업 부가가치 전망 결과를 살펴보면, 2022년은 2021년의 양호한 흐름을 유지할 것으로 예상되나 증가 폭은 전년에 비해 감소할 것으로 전망되었다.

다음으로 머신러닝을 기법을 활용하여 해양수산업 주요 산업의 부가가치를 전망하였다. 부가가치 전망을 위하여 사용되는 데이터는 산업별 부가가치(실질)를 종속변수로 하며, 독립변수는 자본, 수출, 취업자수 등의 산업연관표 항목, 세계GDP, 회사채율 등의 거시경제변수 및 산업별 고유 변수들로 구성되었다. 데이터 관측치는 2005년부터 2020년까지로 총 16개이나 독립변수의 개수는 이를 상회하여 주어진 데이터는 관측치 개수보다 독립변수의 개수가 큰 고차원 데이터이다. 이에 따라, 일반적인 계량경제모델로는 분석이 불가능한 문제가 발생하여 본 연구에서는 머신러닝 방법을 적용하여 부가가치를 전망하였다. 또한, 고차원 데이터의 경우 전망치의 분산이 클 수 있으므로 안정적인 전망에 애로가 발생할 수 있으므로, 이러한 문제를 완화하기 위한 예측조합법을 예측기법으로 포함하였다. 또한, 고차원 데이터임을 고려하여 머신러닝 기법을 활용한 변수선택 방식들도 활용하였다. 이에 따라, 3가지 방식의 변수선택 하에서 8가지의 머신러닝 기법과 3가지 예측조합법으로 해양수산업 각 산업 및 전체산업의 부가가치를 전망하였으며, 2016년부터 2020년까지의 각 예측기법의 전망치로 절대비오차를 구하여 예측기법들 간의 예측성능을 비교하였다. 개별 산업과 전체 산업에서 가장 높은 예측성능을 보이는 변수선택 방식, 머신러닝 기법 및 예측조합법의 종류는 상이했으나, kNN과 Fore1 등이 가장 높은 예측성능을 보

이는 경우가 많이 관찰되었다. 추가적으로 개별 산업에서 가장 높은 예측성능을 보이는 경우의 부가가치 전망치를 합하여 전체 산업의 전망치로 간주하고 절대비오차를 구하였는데, 이 경우에 첫 번째 접근 방식보다 높은 예측성능을 관찰할 수 있었다.

## 2. 결론

본 연구의 결론은 다음과 같다. 첫째, 데이터베이스 자동화 프로그램을 통해 해양수산업 데이터베이스 구축에 대한 시간과 비용을 크게 단축시켰다. 이와 같은 자동화 프로그램을 활용하여 향후 연구에서는 해양수산업 부가가치 외에도 고용, 자본 등의 데이터베이스를 구축하고 전망에 활용할 수 있을 것이다.

둘째, 해양수산업 부가가치 전망을 위해 전망모형의 방법론을 고도화하였다. 한국해양수산개발원은 앞서 수행한 2번의 연구를 통해 해양수산업 부가가치 전망을 생산함수 접근법과 조건부예측모형을 통해 수행한 바 있다. 두 모형 모두 해양수산업의 특성을 반영하고 있으나 변수 간 유기적 관계를 고려한 전망치를 도출하는 데는 약점이 있었다. 올해 구축한 시계열모형은 충분한 시계열이 확보되지 않은 상황에서도 추정의 정확성을 확보할 수 있다는 장점이 있다. 또한 한국해양수산개발원의 전망치를 전망 시나리오에 반영하여 주요 경제지표와 해양수산업 대분류의 부가가치 간의 상관관계가 반영된 전망치를 추정할 수 있었다.

한편, 본 연구는 해양수산업 분야 고차원 데이터를 머신러닝 기법으로 분석하고자 하였다. 기존의 방법론으로는 고차원 데이터의 경우에는 독립변수를 연구자가 임의로 선택하는 경우가 많았으나, 머신러닝 기반의 변수선택 기법을 활용하여 이러한 임의성을 배제한 연구로써 의미가 있다고 볼

---

수 있다. 관측치가 상대적으로 부족함에 따라, 특정 산업에 특정한 머신러닝 기법이 언제나 적절하다고 할 수는 없다. 다만, 산업별로 예측조합법이 비교적 안정적이며 양호한 예측성능을 보이는 것으로 나타났으므로 다양한 방법들을 기반으로 한 예측조합법을 해양수산업 부가가치 전망에 활용하는 것이 바람직해 보인다. 전체 산업의 경우에는 개별 산업의 전망치를 합하는 것이 비교적 안정적이며 양호한 예측성능을 기대할 수 있는 것으로 보인다.

## 제2절 향후 추진 과제

---

### 1. 연구 기여

본 연구는 해양수산업 데이터베이스의 자동화와 시계열 모형인 베이지안 VAR 모형과 다양한 머신러닝 기법을 적용한 전망 모형을 개발하였다.

해양수산업 데이터베이스 자동화 프로그램을 개발하여 기존의 데이터베이스 구축 소요 시간을 획기적으로 단축하였으며, 향후 데이터베이스 확장 및 잠정치 추계에 경제적 효율성이 담보된 시스템을 마련하였다.

전망 모형의 방법론적인 측면에서는 다른 방법론에 비해 높은 예측력을 지닌 것으로 알려진 시계열 모형과 머신러닝 기법을 적용하였으며, 특히 베이지안 VAR 모형에 주요 부문의 시장 지표를 모형에 추가하여 한국해양수산개발원의 주요 지표 전망과 주요기관의 거시경제 전망의 시나리오와 부합하는 전망치를 도출한 것이 본 연구의 기여라 할 수 있다.

그럼에도 불구하고 후속 연구를 통해 보완하여야 할 한계는 다음과 같다. 첫째, 현재 해양수산업 주요 전망은 부가가치를 중심으로 이루어지고 있다. 부가가치는 우리나라 국민경제에서 해양수산업의 위상을 파악하고,

산업의 생산성을 진단하는데 매우 중요한 지표이기 때문에 해양수산업 전망체계에서 가장 중요한 경제지표라 할 수 있을 것이다. 이러한 부가가치 전망 외에도 해양수산업의 경제상황을 보다 포괄적이고 종합적으로 진단·전망하기 위해서 전망 대상을 해양수산업 생산, 고용, 수출입으로 확대할 필요가 있다.

둘째, 1차 년도와 2차 년도, 그리고 이번 3차 년도의 전망 모형 개발을 위해 생산함수 접근법, 조건부 예측, 시계열 모형, 머신러닝 기법 등 다양한 방법론을 적용하여 해양수산업 데이터 특성에 적합한 모형을 발굴하고자 하였다. 이러한 다모형 접근법의 고도화는 해양수산업 전망체계의 완성도를 고도화하고 전망치의 신뢰성을 향상시켰으나 해양수산업 전망체계의 핵심모형 설계를 위한 구조모형의 개발이 필요하다. 하향식(top-down) 방식의 구조모형을 개발하여 해양수산업 전망의 이론적 정합성을 확보하고, 대외여건 및 정책 변화가 해양수산업에 미치는 영향을 시뮬레이션을 수행한다면 현재 전망 모형과 더불어 정책목표 수립에 활용될 수 있을 것이다.

## 2. 향후 추진 과제

향후 추진될 해양수산업 전망체계는 구조모형을 개발하고, 핵심모형으로 활용하여 이론적 정합성이 담보된 해양수산업 전망을 수행하는 것을 목적으로 한다. 이는 현재 1~2년 주기의 비교적 단기 전망에서 확대하여 3~5년 주기의 중장기전망이 가능케 할 것이다. 또한 중앙정부의 주요 정책의 변화 또는 대외 충격이 해양수산업에 미치는 영향을 시뮬레이션을 통해 분석함으로써 분석 목적과 시나리오에 따라 유연한 모형 운용이 가능할 것으로 기대할 수 있다.

〈그림 5-1〉 해양수산업 전망모형 구축 단계(안)

## 단계

국민계정 통계와 정합성 있는  
데이터베이스 구축

해양수산업 중심 데이터베이스  
이전(예시)  
-해양수산업 부문별  
분할 비중 도출-

거시-산업 계량경제 모형  
Eviews 및 R 프로그래밍

머신러닝 모형  
R 프로그래밍

모형 적용 흐름도

## 예시

**국민계정 통계와 정합성 있는 데이터베이스 구축**

해양수산업 중심 데이터베이스 이전(예시)

구분	항목	단위	2019년	2020년
1 해양자원 개발 및 건설업	13 생산 및 해상교량 건설업	292	1	1
	01 해양자원을 처리 및 정제	294(상하수도시설)	0.036	0.036
	13 생산 및 해상교량 건설업	295	0.190	0.190
	6 수산물 유통업	302(도매서비스)	0.030	0.030
9 해양수산 관련 서비스업	61 수산물 유통업	302(도매서비스)	7.42613E-05	7.42613E-05
	61 수산물 유통업	303(소매서비스)	0.006	0.006
	61 수산물 유통업	303(소매서비스)	7.42613E-05	7.42613E-05
	2 해운물만업	309(건설 및 내륙수상 운	1	1
2 해운물만업	21 해운업	310(외항운송서비스)	1	1
	2 해운물만업	315(항공운송서비스)	1	1
	2 해운물만업	315(항공운송서비스)	0.546	0.546
	2 해운물만업	316(보관 및 창고 서비스)	0.17	0.17

거시-산업 계량경제 모형 Eviews 및 R 프로그래밍

머신러닝 모형 R 프로그래밍

투입산출표 기반 해양수산업 중심 데이터베이스

해양수산 거시계량투입산출모형(안)

조건부배출모형(예)

총격반응함수분석(예)

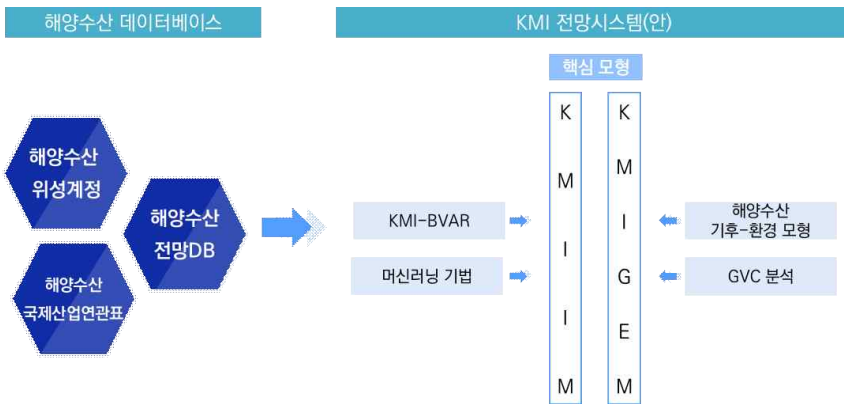
자료: 장정인 외(2021a), p. 61.

본 연구에서 고찰한 다양한 전망 방법 중 해양수산업 전망모형에 적합한 방법론은 거시경제와 산업 간의 유기적인 관계를 고려한 산업·거시계량모형이다. 동 모형은 거시경제 및 주요 지표가 연립방정식 시스템에서 내생적으로 추정됨으로써 거시경제와 산업부문을 통합하여 분석할 수 있다는 장점이 있다. 산업·거시계량모형인 KMI-IM(가제)을 핵심모형으로 설정하



고, 예측력 확보를 위한 시계열 모형을 보조 모형으로 설정하여 단기 전망에 활용한다면 모형 간 상호보완적인 역할을 수행할 수 있을 것이다. 정량적·정성적 데이터를 모두 활용 가능한 머신러닝 기법을 활용하여 종합적이고 포괄적인 전망을 수행할 것이다.

〈그림 5-2〉 KMI 전망시스템(안)



자료: 저자작성

이 외에도 한국해양수산개발원에서 해양수산업을 중심으로 구축된 CGE(Computable General Equilibrium, 연산가능일반균형) 모형인 KMI-GEM과의 연계를 통해 체계적인 정책 영향을 분석할 것이다. 또한, 기후변화와 탄소중립이 해양수산업에 미치는 경제적 영향을 분석하기 위해 개발 중인 해양수산 기후-경제모형(가제)과 해양수산업 국제산업연관표에 기반한 글로벌 밸류 체인(GVC: Global Value Chain) 분석을 활용하여 해양수산업을 둘러싼 주요 이슈와 그 파급효과에 대해 다각적이고 종합적인 분석을 수행할 것이다.<sup>65)</sup>

65) KMI-GEM에 대한 자세한 내용은 장정인 외(2021d)를, 해양수산 분야 국제산업연관표에 대한 자세한 내용은 정수빈 외(2021)를 참고하기 바란다.

---

이를 위해선 해양수산업 데이터베이스 작성과 현실 설명력 제고를 위한 개선이 꾸준히 이루어져야 할 것이다. 해양수산 위성계정 구축을 통해 해양수산 경제통계의 시의성, 국민계정과의 정합성이 확보된 해양수산 데이터를 제공할 것이며, 실질자본스톡 추계방법 개선, 동태적 추정방정식을 통한 데이터 작성방법에 대한 개선도 필요할 것이다.

---

## 참고문헌

### 국내 문헌

---

- 구미령·유진은, 「집단내 상관계수 (ICC) 크기에 따른 별점회귀모형의 변수 선택 및 예측 성능 비교: TALIS 2018 자료의 교사협력 예측 변수 탐색」, 『아시아교육연구』, 제22권, 제1호, 2021, 31-59.
- 나인강·류지철(2000), 『에너지 수요분석 및 전망-2001년 수요전망과 정책이슈』-, 에너지경제연구원
- 김길훈·고성효, 「제주도 해양관광 수요예측에 관한 연구」, 『산경논집』, 제23권, 제1호, 2008, 19-42.
- 김동현·한승재·정봉규·한승훈·이상봉, 「머신러닝을 이용한 기관 출력 예측 방법에 관한 연구」, 『해양환경안전학회지』, 제25권, 제7호, 2019, 851-857.
- 김두환·이강배, 「LSTM을 활용한 부산항 컨테이너 물동량 예측」, 『한국항만경제학회지』, 제36권, 제2호, 2020, 53-62.
- 김윤영·박준용, 「혼합주기자료 VAR모형을 이용한 경제성장률 예측」, 『금융경제연구』, 제268호, 2006.
- 김현석, 「VAR 모형과 Bayesian VAR 모형의 물동량 예측력 비교 연구」, 『해운물류연구』, 제31권 제2호, 2015, 449-466.
- 김현석·장명희, 「Bayesian VAR 를 이용한 해운경기, 환율 그리고 산업생산 간의 동태적 상관분석」, 『한국항만경제학회지』, 제30권, 제2호, 2014, 77-92.
- 남종오·노승국, 「시계열모형을 이용한 굴 생산량 예측 가능성에 관한 연구」, 『한국해양수산개발원 수산업관측센터』, 제34권, 제2호, 2012, 185-195.

- 
- 노태영·최태련, 「백터자기회귀 모형 추정을 위한 베이지안 축소 방법론 비교 연구」, 『Journal of The Korean Data Analysis Society』, 제18권, 제4호, 2016, 1857-1870.
- 박경훈·임현준·노경서, 「한국은행 거시계량모형(BOK20) 구축결과」, 『조사통계월보』, 2020
- 박성진·배후석·신용준, 「연안해운 선원인력 수요예측에 관한 연구」, 『한국항해항만학회지』, 제36권, 제3호, 2012, 205-213.
- 박양수, 『경제전망의 실제』, 서울: 한티미디어, 2011.
- 박해훈·윤갑동, 「한국 멸치어업의 어획량 분석과 예측」, 『한국수산과학회지』, 제29권, 제2호, 1996, 143-149.
- 배남규·최영문, 「중소조선업계의 도산예측변수에 관한 연구」, 『국제회계연구』, 제9권, 2020, 1-24.
- 배성완·유정석, 「머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측」, 『주택연구』, 제26권 제1호, 2018, 107-133.
- 백천현·김후곤·김영진·정용주, 「상향식 모형을 이용한 국내 조선업의 온실가스 배출 분석」, 『경영과학』, 제31권, 제1호, 2014, 41-48.
- KIET 산업·거시경제 단기 계량전망모형 개발 연구 발표자료(2019. 5. 4.)
- KIET-DIMM의 DB 구축현황 발표자료(2020. 7. 29.)
- 손민규·김대용·황상필, 「한국은행 분기거시계량모형 (BOK12) 재정모형 구축결과」, 『조사통계월보』, 2013년 6월호, 한국은행, 2013.
- 에너지경제연구원, 『에너지수요전망 2018 겨울호』
- 에너지 수급전망 전문가 세미나(KEEI) 발표자료(2021.7.13.)
- 앵호자리갈 운자야·박수호·황도현·정민지·김나경·윤홍주, 「GOCI 영상과 기계학습 기법을 이용한 *Cochlodinium polykrikoides* 적조 탐지 기법 연구」, 『한국전자동신학회논문지』, 제15권, 제6호, 2020, 1089-1098.
- 오진호, 「제주도 항만 물동량 예측: 주요 품목을 대상으로」, 『제주발전연구』, 제20호, 2016, 147-167.
- 이광남·정진호, 「어업용 면세유류 사용량 예측에 관한 연구」, 『수산경영론집』, 제43권, 제1호, 2012, 49-61.
- 이상원·정인환, 「선박 신수요 예측을 위한 빅데이터 기반 인공지능 알고리즘을 활용

- 한 플랫폼 개발」, 『한국인터넷방송통신학회 논문지』, 제19권, 제1호, 2019, 171-178.
- 이진면·김동석·김민수, 『다부문모형에 의한 산업구조 변화의 장기전망』, 한국개발연구원, 2001.
- 이진면·변창욱·최용재·김진웅·이상호, 『KIET 산업경제계량모형』, 산업연구원, 2007.
- 이진면·이용호·김재진·김바우·한정민, 『KIET 산업·거시경제 단기 계량전망모형 개발 연구』, 산업연구원, 2013.
- 이진면·이용호·김재진, 『4차 산업혁명과 우리 산업의 중장기 구조변화 전망』, 산업연구원, 2018.
- 장정인·권장한·정수빈, 『해양수산업 전망모형 개발 연구(II)』, 한국해양수산개발원, 2021a.
- 장정인·김태진·정수빈·박희대·고병욱·이현동·하태영, 『해양수산업 위성계정 기초연구(I)』, 한국해양수산개발원, 2020.
- 장정인·김태진·정수빈·김정우, 『해양수산업 전망모형 개발 연구(I)』, 한국해양수산개발원, 2020.
- 장정인·정수빈·박동욱·김태한, 『2020 해양수산업 분석 보고서 -2016~2017년 산업연관표 기준-』, 한국해양수산개발원, 2021b.
- 장정인·박동욱·권장한·정수빈·김주현·김태한·김성은·최일선·마창모·허수진·하현정·최건우·이기열·이수영, 『2022 해양수산 주요지표 전망』, 한국해양수산개발원, 2021c.
- 장정인·임병호·정인교·권장한·조정란·정수빈·김성아·최건우, 『해양수산업 부문 정책효과 분석을 위한 KMI-GEM 구축(2)』, 한국해양수산개발원, 2021d.
- 전해정, 「베이지안 패널 VAR모형을 이용한 거시경제변수가 지역 주택가격에 미치는 영향」, 『인문사회 21』, 제10권, 제6호, 2019, 1349-1362.
- 정수빈·장정인·권장한·김성은·박동욱, 『해양수산 분야 국제산업연관표 구축 기반 연구』, 한국해양수산개발원, 2021.
- 최필선·민인식, 「머신러닝 기법을 이용한 대졸자 취업예측 모형」, 『직업능력개발연구』, 제21권, 제1호, 2018, 31-54.
- 함종영·손재영, 「사전확률분포를 이용한 주택시장 예측모형 비교 연구: Bayesian VA

- 
- R 모형을 중심으로, 『부동산·도시연구』, 제8권, 제2호, 2016, 25-38.
- 황경후·김원표·정철, 「일반량 시계열 모형 비교를 통한 독도관광 수요예측」, 『관광레저연구』, 제27권, 제2호, 2015, 59-77.
- 황상필·박양수·최강욱, 「한국은행 거시계량투입산출모형」, 『조사통계월보』, 2006년 9월호, 한국은행, 2006.
- KMI-KIEP 업무 교류협력 협의회 발표자료(2021.6.24.)

## 국외 문헌

---

- A. Jugović, S. Hess and T. Poletan Jugović, "Traffic Demand Forecasting for Port Services", *Promet-Traffic & Transportation*, Vol. 23, No. 1, 2011, pp. 59-69.
- Agus Perdana Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method", *International Journal of Artificial Intelligence Research*, Vol. 1, No. 2, 2017, pp. 26-33.
- Altmann, A., Tološi, L., Sander, O. and Lengauer, T., "Permutation importance: a corrected feature importance measure", *Bioinformatics*, Vol. 26, No. 10, 2010, pp. 1340-1347.
- Bair, E., Hastie, T., Paul, D. and Tibshirani, R., "Prediction by supervised principal components", *Journal of the American Statistical Association*, Vol. 101, No. 473, 2006, pp. 119-137.
- Barreiro-Ures, D., Cao, R. and Francisco-Fernandez, M. "Bagging cross-validated bandwidth selection in nonparametric regression estimation with applications to large-sized samples", *arXiv preprint arXiv:2105.04134*, 2021.
- Barshan, E., Ghodsi, A., Azimifar, Z. and Jahromi, M. Z., "Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds", *Pattern Recognition*, Vol. 44, No. 7, 2011, pp. 1357-1371.
- Breiman, L., "Random Forests", *Machine Learning*, Vol. 45, 2001, pp. 5-32.
- Breiman, Leo, "Bagging predictors", *Machine Learning*, Vol. 24, 1996, pp. 123-140.
- CHANG, M., XING, Y. Y., ZHANG, Q. Y., HAN, S. J., & Kim, M., "A CNN Image Classifi

- cation Analysis for 'Clean-Coast Detector' as Tourism Service Distribution", *Journal of Distribution Science*, Vol. 18, No. 1, 2020, pp. 15-26.
- Chi Zhang, Lei Huang and Zhichao Zhao, "Research on combination forecast of port cargo throughput based on time series and causality analysis", *Journal of Industrial Engineering and Management*, Vol. 6, No. 1, 2013, pp. 124-134.
- Desmond Eseoghene Ighravwe and Christopher Osita Anyaeche, "A comparison of ARIMA and ANN techniques in predicting port productivity and berth effectiveness", *International Journal of Data and Network Science*, Vol. 3, No. 1, 2019, pp. 13-22.
- Ehab A. Kholief, Samy H. Darwish and M. Nashat Fors, "Detection of Steel Surface Defect Based on Machine Learning Using Deep Auto-encoder Network", *Industrial Engineering and Operations Management*, 2017, pp. 218-229.
- Enriqueta Velarde, Exequiel Ezcurra, Miguel A. Cisneros-Mata-Miguel, and F. Lavi, "Seabird Ecology, El Nino Anomalies, and Prediction Of Sardine Fisheries In The Gulf Of California", *Ecological Applications*, Vol. 14, No. 2, 2004, pp. 607-615.
- Fix, Evelyn and Hodges, Joseph L., "Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties", *USAF School of Aviation Medicine*, 1951.
- Gang Xie, Yatong Qian and Shouyang Wang, "Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach", *Tourism Management*, Vol. 82, 2021.
- Gil Goncalves, Umberto Andriolo, Luis Pinto and Diogo Duarte, "Mapping marine litter with Unmanned Aerial Systems: A showcase comparison among manual image screening and machine learning techniques", *Marine Pollution Bulletin*, Vol. 155, 2020.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, "The elements of statistical learning: data mining, inference, and prediction", *Springer Science & Business Media*, 2009.
- Huang, N., Lu, G. and Xu, D., "A permutation importance-based feature selection

- 
- method for short-term electricity load forecasting using random forest”, *Energies*, Vol. 9, No. 10, 2016, pp. 767-790.
- Inigo Gomara, Belen Rodriguez-Fonseca, Elsa Mohino, Teresa Losada, Irene Polo and Marta Coll, “Skillful prediction of tropical Pacific fisheries provided by Atlantic Ninos”, *Environmental Research Letters*, Vol. 16, No. 5, 2021.
- J. M. Bates and C. W. J. Granger, “The Combination of Forecasts”, *Operational Research*, Vol. 20, No. 4, 1969, pp. 451-468.
- Jeremy Smith and Kenneth F. Wallis, “A Simple Explanation of the Forecast Combination Puzzle”, *OXFORD BULLETIN OF ECONOMICS AND STATISTICS*, Vol. 71, No. 3, 2009, pp. 331-355.
- Junhui Li, “An Econometric Analysis of Shipbuilding Market in China”, Rotterdam: Erasmus University of Rotterdam, 2011.
- Laouafi, A., Mordjaoui, M., Haddad, S., Boukelia, T. E. and Ganouche, A., “Online electricity demand forecasting based on an effective forecast combination methodology”, *Electric Power Systems Research*, Vol. 148, 2017, pp. 35-47.
- Li, J., Bian, C., Chen, D., Meng, X., Luo, H., Liang, H. and Shen, L., “Persistent Feature Analysis of Multimodal Brain Networks Using Generalized Fused Lasso for EMCI Identification”, In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 44-52.
- Lloyd, S. P., “Least squares quantization in PCM”, Technical Report RR-5497, Bell Lab, 1957.
- Mahdi Panahi, Nitheshnirmal Sadhasivam, Hamid Reza Pourghasemi, Fatemeh Rezaie and Saro Lee, “Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR)”, *Journal of Hydrology*, Vol. 588, 2020.
- Mengting Chen, Yufeng Luo, Yingying Shen, Zhenzhong Han and Yuanlai Cui, “Driving force analysis of irrigation water consumption using principal component regression analysis”, *Agricultural Water Management*, Vol. 234, 2020.
- Misganaw Abebe, Yongwoo Shin, Yoojeong Noh, Sangbong Lee and Inwon Lee, “Machine Learning Approaches for Ship Speed Prediction towards Energy Efficient Shipping”, *Applied Sciences*, Vol. 10, No. 7, 2020, pp. 2325-2342.
-



- P. Mohamed Shakeel, S. Baskar, V. R. Sarma Dhulipala and Mustafa Musa Jaber, "Cloud based framework for diagnosis of diabetes mellitus using K-means clustering", *Health Information Science and Systems*, Vol. 6, No. 1, 2018, pp. 1-7.
- Pearson, K., "On Lines and Planes of Closest Fit to Systems of Points in Space", *Philosophical Magazine Series 6*, Vol. 2, No. 11, 1901, pp. 559-572.
- Quan Quan, Zou Hao, Huang Xifeng and Lei Jingchun, "Research on water temperature prediction based on improved support vector regression", *Neural Computing and Applications*, 2020, pp. 1-10.
- Rew, J., Cho, Y. and Hwang, E., "A Robust Prediction Model for Species Distribution Using Bagging Ensembles with Deep Neural Networks", *Remote Sensing*, Vol. 13, No. 8, 2021.
- Ruihua Lu, Osman Turan, Evangelos Boulougouris, Charlotte Banks and Atilla Incek, "A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping", *Ocean Engineering*, Vol. 110, 2015, pp. 18-28.
- Shang, H. L. and Haberman, S., "Model confidence sets and forecast combination: an application to age-specific mortality", *Genus*, Vol. 74, No. 1, 2018, pp. 1-23.
- Shigemizu, D., Akiyama, S., Asanomi, Y., Boroevich, K. A., Sharma, A., Tsunoda, T., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Sakurai, T., Ozaki, K., Ochiya, T. and Niida, S., "Risk prediction models for dementia constructed by supervised principal component analysis using miRNA expression data", *Communications biology*, Vol. 2, No. 1, 2019, pp. 1-8.
- StataCorp, L. P., *Stata base reference manual*, College Station: StataCorp LLC, 2021
- Tibshirani, R. J. and Taylor, J., "The solution path of the generalized lasso", *The annals of statistics*, Vol. 39, No. 3, 2011, pp. 1335-1371.
- Tibshirani, Robert, "Regression Shrinkage and Selection via the lasso", *Journal of the Royal Statistical Society. Series B (methodological)*, Vol. 58, No. 1, 1996, pp. 267-288.
- V. N. Vapnik and A. Ya. Lerner, "Recognition of Patterns with help of Generalized

- 
- Portraits”, *Avtomatika i Telemekhanika*, Vol. 24, No. 6, 1963, pp. 774-780.
- Viyada Suriyakul Na Ayudhaya and Praew Ritthirungrat, “The econometric analysis of the factors affecting the revenue of Bangkok Port”, *World Maritime University MS Thesis*, 2018.
- Waldmann, P., Meszaros, G., Gredler, B., Fuerst, C. and Solkner, J., “Evaluation of the lasso and the elastic net in genome-wide association studies.”, *Frontiers in genetics*, Vol. 4, 2013.
- Wang Cai, Jianzhuang Wang, Longchao Cao, Gaoyang Mi, Leshi Shu, Qi Zhou and Ping Jiang, “Predicting the weld width from high-speed successive images of the weld zone using different machine learning algorithms during laser welding”, *Mathematical Biosciences and Engineering*, Vol. 16, No. 5, 2019, pp. 5595-5612.
- Wang, J. and Zuo, R., “Assessing geochemical anomalies using geographically weighted lasso”, *Applied Geochemistry*, Vol. 119, 2020.
- Wang, X., Benesty, J., Chen, J. and Cohen, I., “Beamforming with small-spacing microphone arrays using constrained/generalized lasso”, *IEEE Signal Processing Letters*, Vol. 27, 2020, pp. 356-360.
- Weiss, C. E., Raviv, E. and Roetzer, G., “Forecast Combinations in R using the ForecastComb Package”, *R Journal*, Vol. 10, No. 2, 2018, pp. 262-281.
- Wenchao Xing and Yilin Bei, “Medical Health Big Data Classification Based on KNN Classification Algorithm”, *IEEE Access*, Vol. 8, 2019, pp. 28808-28819.
- Xianglong Luo, Danyang Li, Yu Yang and Shengrui Zhang, “Spatiotemporal Traffic Flow Prediction with KNN and LSTM”, *Journal of Advanced Transportation*, Vol. 2019, 2019, pp. 1-10.
- Yuqing Geng, Mukasar Maimaituerxun and Han Zhang, “Coupling Coordination of Water Governance and Tourism: Measurement and Prediction”, *Discrete Dynamics in Nature and Society*, Vol. 2020, 2020, pp. 1-13.
- Zhihui Hu, Yongxin Jin, Qinyou Hu, Sukanta Sen, Tianrui Zhou and Mohd Tarmizi Osman, “Prediction of Fuel Consumption for Enroute Ship Based on Machine Learning”, *IEEE Access*, Vol. 7, 2019, pp. 119497-119505.

- Zhe Song, Qinli Deng and Zhigang Ren, "Correlation and principal component regression analysis for studying air quality and meteorological elements in Wuhan, China", Environmental Progress & Sustainable Energy, Vol. 39, No. 1, 2019.
- Zou, Hui and Hastie, Trevor, "Regularization and Variable Selection via the Elastic Net", Journal of the Royal Statistical Society, Series B, Vol. 67, No. 2, 2005, pp. 301-320.

## 인터넷 자료

- 한국은행경제통계시스템(ECOS), <http://ecos.bok.kr/>, (검색일:2021.12.15.)
- 한국해양수산개발원 보도자료(2022. 1. 25.), 「KMI 2022 해양수산 전망대회, 해양수산의 코로나 팬데믹 이후 대응과 재도약을 말하다」
- 해사신문(2019.02.01.), <http://www.haesaneews.com/news/articleView.html?idxno=84122>(검색일: 2021.11.09)
- OpenGenus, <https://iq.opengenus.org/principal-component-regression/>(검색일 : 2021.08.08)
- saedsayad, [https://www.saedsayad.com/support\\_vector\\_machine\\_reg.htm](https://www.saedsayad.com/support_vector_machine_reg.htm)(검색일 : 2021.08.08)
- towards data science, <https://towardsdatascience.com/> (검색일 : 2021.07.09)
- vvakki, [https://velog.io/@vvakki\\_](https://velog.io/@vvakki_)(검색일 : 2021.07.09.)
- wikipedia, [https://en.wikipedia.org/wiki/Bootstrap\\_aggregating](https://en.wikipedia.org/wiki/Bootstrap_aggregating)(검색일 : 2021.07.20.)



## 부록

### 1. 해양수산업 데이터베이스 자동화 프로그램

```
Sys.setenv(LANG = "en_US.UTF-8")
library("readxl")
setwd("E:/Mine/경통실/2021/IO")

### 입력 파일들
iotable <- read_excel ("투입산출표_생산자가격_기본부문_2017.xlsx", sheet =1)[ -c(1:3) , ]
# 산업연관표
input <- read_excel ("input.xlsx", sheet =1 ) # 총산출 등 명목 국민계정 항목
rtoarr <- read_excel("2. 해양수산 중심 IO_(2차)해양수산업 중심
IO_2017_1216_200421.xlsx", sheet = 4) # 해양수산업 비율 추출
sumrc <- read_excel("2. 해양수산 중심 IO_(2차)해양수산업 중심
IO_2017_1216_200421.xlsx", sheet = 7) # 80(분류+해양수산업)코드 매핑
mt1524r <- read_excel("3. (종합) 해양수산업 IO_2015_2017_실질표.xlsx", sheet = 6) # 디
플레이터 116 코드 추출
def <- read.csv("deflator.csv") # 디플레이터
inputr <- read_excel ("inputr.xlsx", sheet =1 ) # 총산출 등 실질 국민계정 항목
mtchcd <- read_excel("3.(고용표)_해양수산업 IO_2015_2017_1213.xlsx", sheet = 5) # 취업
자피용자 코드 매핑
empstat <- read.csv("취업피용.csv") # 취업자 피용자 데이터
mtchcdct <- read_excel("3.1.(자본스톡)_해양중심_취합_연계표_2015.xlsx", sheet = 2) # 자
```

본스톡 코드 매핑

```
catstat <- read.csv("자본스톡.csv") # 자본스톡 데이터
iyvmt <- read_excel("5.고정자본형성_전환행렬_2005_2010_1230.xlsx", sheet = 3) # 전환행
렬 비율
cname <- read_excel("1.전망모형DB_산업별자료_명목_실질_18년포함__1230.xlsx", sheet =
2) # 최종결과 항목들

### 1. 경상표 : 명목GDP 결과 만들기
### 1.2 총거래표 조정
iotable <- read_excel ("투입산출표_생산자가격_기본부문_2018.xlsx", sheet =1)[ -c(1:3) , ]
# 산업연관표
write.csv(iotable, "iptable.csv")
iotable0 <- read.csv("iptable.csv", fileEncoding = "euc-kr")[,~1]

input <- read_excel ("input.xlsx", sheet =1 ) # 총산출 등 명목 국민계정 항목
iotable1 <- matrix(as.numeric(unlist(iotable0[1:390, 1:400])), 390, 400) # make a matrix
ipt <- unlist(input[5:9, 3]) # 한국은행 국민계정 중간투입 ~ 생산세
iotbl <- iotable1[384:388, 3:384] # 중간수요계까지 382
iotbl1 <- iotbl[, 1:381]*iotbl[, 382]^(-1) # 비율
iotbl2 <- iotbl1*ipt # 비율*한국은행 국민계정 수치
iotable0[which(iotable0[,2]=="중간투입계"), 3:383] <-
(as.numeric(iotable0[which(iotable0[,2]=="중간투입계"),
3:383])/sum(as.numeric(iotable0[which(iotable0[,2]=="중간투입계"),
3:383])))*as.numeric(input[which(input$item=="중간투입(총산출-부가가치)" ,3)]) # 중간투입
계 조정
iotbl3 <- rbind(iotbl2, colSums(iotbl2[2:5,]), colSums(iotbl2[1:5,])) # 부가가치와 총투입
추가
iotable0[385:390, 3:383] <- iotbl3[2:7, ]
iotable0[3:383, which(iotable0[2,]=="총산출")]<- as.numeric(iotable0[390, 3:383]) # 총투
입행 ->총산출열
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[2,]=="총산출")] <- sum(
as.numeric(iotable0[3:383, which(iotable0[2,]=="총산출"])))
```

```

#### 3. Total control
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="정부소비지출")] <-
as.numeric(input[which(input$item=="정부소비지출"),3])
## 민간고정 = 고정자본 in input
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="민간고정Wn자본형성")]
<- as.numeric(input[which(input$item=="고정자본형성"),3])
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="수출")] <-
as.numeric(input[which(input$item=="수출"),3])
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="수입")] <-
as.numeric(input[which(input$item=="수입"),3])
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="총공급계Wn")] <-
as.numeric(iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="총산출"])] +
as.numeric(input[which(input$item=="수입"),3])
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="총수요계")] <-
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="총공급계Wn")]
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="중간수요계")] <- sum(
as.numeric(iotable0[which(iotable0[,2]=="중간투입계"), 3:383]))
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="최종수요계")] <-
as.numeric(iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="총수요계"])]
- as.numeric(iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="중간수요계
")]))
iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="재고증감")] <-
as.numeric(iotable0[which(iotable0[,2]=="중간투입계"), which(iotable0[,2, ]=="최종수요계
")]))
- as.numeric(input[which(input$item=="수출"),3])
as.numeric(input[which(input$item=="고정자본형성"),3])
as.numeric(input[which(input$item=="정부소비지출"),3])
as.numeric(input[which(input$item=="민간소비지출"),3]) # 재고증감=최종수요-수출-고정
자본-정부소비지출-민간소비지출

#### 3. 비율표(명목) 민간 정부 고정 수출 수입
iotabler <- read.csv("iptable.csv", fileEncoding = "euc-kr")[-1]
## 민간고정 자본형성 = 고정자본형성 (정보고정 자본형성 더 해준것)
iotabler[3:384, which(iotabler[,2, ]=="민간고정Wn자본형성")] <- as.numeric(iotabler[3:384,
which(iotabler[,2, ]=="민간고정Wn자본형성"])] + as.numeric(iotabler[3:384,

```

```

which(iotabler[2, ]=="정부고정Wn자본형성"))

iotablerto <- sweep(matrix(as.numeric(unlist(iotabler[3:383, 3:400] )), 381, 398),
MARGIN = 2, (as.numeric(iotabler[384, 3:400])^(-1)), "*")
iotablerto[ is.na(iotablerto ) ] <- 0

### 4. 명목표(최종)
iotablefni <- sweep(iotablerto, MARGIN = 2,
as.numeric(iotable0[384, 3:400]), "*")
iotablefni[, 392] <- as.numeric(iotable0[3:383, 394]) # 총산출 조정
iotablefni[, 398] <- rowSums(iotablefni[, 392:393]) # 총공급계 조정
iotablefni[, 391] <- iotablefni[, 398] # 총수요계 조정
iotablefni[, 390] <- iotablefni[, 391] - iotablefni[, 382] # 최종수요계 조정
iotablefni[, 387] <- iotablefni[, 390] - iotablefni[, 389] - iotablefni[, 385] - iotablefni[,
384] - iotablefni[, 383] # 재고증감 = 최종수요계-수출-고정자본-정부소비-민간소비
iotablefni <- rbind(iotablefni, colSums(iotablefni ))

##### 해양수산 IO: 비율은 주어진 것으로
### 차수별 비율정리
rtoarr <- read_excel("2. 해양수산 중심 IO_(2차)해양수산업 중심
IO_2017_1216_200421.xlsx", sheet = 4) # 해양수산업 비율 추출

mrnotmr <- matrix(0, 381, 10)
colnames(mrnotmr) <- rtoarr[99, 4:13]
mrnotmr[,1] <- as.vector( unlist(rtoarr[100:480, 4]) )
mrnotmr[,2] <- as.vector(unlist(rtoarr[100:480, 5]))

for (i in 1:381){ # i <- 297 mar1 <- 2
  mar1 <- which(mrnotmr[i,1]==rtoarr[1:81, 4])
  if ( length(mar1)==0 ){mar1 <- 0
  } else {
    mar1 <- as.numeric(rtoarr[mar1, 8])
  }
}

```



```

mrnotmr[ i ,3] <- mar1
mar2 <- which(mrnotmr[ i ,1]==rtoarr[86:92, 4])
if ( length(mar2)==0 ){mar2 <- 0
} else {
  mar2 <- as.numeric(rtoarr[85+mar2, 8])
}
mrnotmr[ i ,4] <- mar2
mar3 <- which(mrnotmr[ i ,1]==rtoarr[95, 4])
if ( length(mar3)==0 ){mar3 <- 0
} else {
  mar3 <- as.numeric(rtoarr[95, 8])
}
mrnotmr[ i ,5] <- mar3
# mrnotmr[ i ,4] <- length(which(mrnotmr[ i ,1]==rtoarr[86:92, 4]))
# mrnotmr[ i ,5] <- length(which(mrnotmr[ i ,1]==rtoarr[95, 4]))
mrnotmr[ i ,6] <- 1- sum( as.numeric( mrnotmr[i ,3:5] ) )
}

### 대각행렬(1524*1524)
d1 <- matrix(0, 1524, 381)
d2 <- matrix(0, 381, 1524)
for (i in 1:381){
  d1[(4*i-3):(4*i), i] <- as.numeric(mrnotmr[ i ,3:6])
  d2[i, (4*i-3):(4*i)] <- as.numeric(mrnotmr[ i ,3:6])
}
d3 <- d1%*%iotablefinl[1:381,1:381 ]%*%d2
## 부가가치계 등 추가(아래부분)
d4 <- matrix(as.numeric(unlist(iotable0[384:390, 3:383 ])), 7,381)
d5 <- rbind(d3, d4 %*% d2)
## 중간수요계 등 추가(오른쪽 부분)
int <- rbind(d1%*%iotablefinl[1:381, 382:398], matrix(0, 7, 17))
mat1524 <- cbind(d5, int )

### 대각행렬(1524*1524) 종대분류 열합
## 1524(381) to 80(분류 + 해양수산업) 은 주어진 것으로

```

---

```

sumrc      <-      read_excel("2.      해양수산      중심      IO_(2차)해양수산업      중심
IO_2017_1216_200421.xlsx", sheet = 7) # 80(분류+해양수산업)코드 매핑

mrsum <- matrix(0, 75,1542)
mrsum[1,] <- NA
ccde <- as.numeric(  unlist(sumrc[1541:1614,3]) )
mrsum[2:75, 1] <- ccde

mdcls <- as.numeric(unlist(sumrc[1540, 4:1527])) ## 1524  to 80 코드 매핑

mrsum[1, 2:1525] <- mdcls

mdcls1 <- c(mdcls, rep(NA, 7) )
int <- cbind(mdcls1, mat1524)
int1 <- rbind(rep(NA, NCOL(int)), int)
int1[1,2:1525] <- as.numeric(unlist(sumrc[1540, 4:1527]))
mat15241 <- int1

## 중분류 행함
for (i in 2:75){
  m <- which(mat15241[, 1]==mrsum[i,1])
  for (j in 2:1542){
    mrsum[i,j] <- sum(mat15241[m, j])
  }
}

## 중분류 열함
mcsum <- matrix(0, 82,92)
mcsum[1,] <- NA
mcsum[2:75,1] <- ccde
mcsum[1,2:75] <- ccde
for (j in 2:75){
  m <- which(mrsum[1, ]==mcsum[1,j])
  for (i in 2:75){
    mcsum[i,j] <- sum(mrsum[i, m])
  }
}

```

---

```

    }
    for (i in 76:82){
      mcsun[i,j] <- sum(mat15241[(1+1524+(i-75)    ), m])
    }
  }

# 중간수요계 등 삽입
mcsun[2:75, 76:92] <- mrsum[2:75, 1526:1542]
# 대분류 코드 추가
mcsun1 <- matrix(NA, 82, 93)
cde <- as.character(unlist(sumrc[1541:1614,2]))
mcsun1[2:75,1] <- cde
mcsun1[, -1] <- mcsun

## 대분류 행합
lrsum <- matrix(NA, 42, 92)
lrsum[2:42, 1] <- unique(cde)
lrsum[1, 2:75] <- cde
for (i in 2:42){
  m <- which(mcsun1[, 1]==lrsum[i,1])
  for (j in 2:92){
    lrsum[i,j] <- sum( as.numeric(mcsun1[m, (j +1) ] ) ) # + 1: 대분류 코드 추가
  }
}

## 대분류 열합
lcsun <- matrix(NA, 49, 59)
lcsun[2:42, 1] <- unique(cde)
lcsun[1, 2:42] <- unique(cde)
for (j in 2:42){
  m <- which(lrsum[1, ]==lcsun[1,j])
  for (i in 2:42){
    lcsun[i,j] <- sum( as.numeric(lrsum[i, m]) )
  }
}
for (i in 43:49){

```

```

lcsun[i,j] <- sum( as.numeric(mcsun1[ (which(mcsun1[,1]=="해양수산 관련 서비스
업") + i-42) , (m + 1) )) ) # m + 1: 대분류 + 취합코드
}
}
##### 중간수요계 등 삽입, 명목 GDP 최종 결과
lcsun[2:42, 43:59] <- lrsun[2:42, 76:92]

##### 실질 GDP 만들기

## 116(65.소수점) 코드 주어진것으로
mt1524r <- read_excel("3. (종합) 해양수산업 IO_2015_2017_실질표.xlsx", sheet = 6)
cde116 <- mt1524r[1541:1656, 7]
# int <- mt1524r[3:1536, 5:(7+1524+17)]
tmp0 <- mt1524r[3:1536, 5:(7+1524+17)]
## 명목결과 1524 행렬 활용
tmp <- matrix(NA, (1524+7+3), (1524+17+3))
tmp[4:(1524+7+3) , 4:(1524 + 17 + 3)] <- mat1524 # 값만 있는 명목 1524 행렬
tmp[1:3, ] <- unlist(tmp0[1:3, ])
tmp[, 1:3] <- unlist(tmp0[, 1:3])
int <- tmp

idx <- list()
for (i in 1:116){
  idx[[i]] <- which(as.character(int[1,]) == as.character( cde116[i,1]))
}
mt116r <- matrix(NA, (117 + 7) , (117 + 17) )

## 디플레이터 116 코드
for (i in 1:116){
  for (j in 1:116){
    mt116r[ (i+1),(j+1) ] <- sum(as.numeric(unlist(int[ idx[[i]], idx[[j]] ])))
  }
}

```

```

## 중간투입계 등
int1 <- which(int[, 3]=="중간투입계")
for (i in 1:7){
  for (j in 1:116){
    mt116r[ (i+17),(j+1) ] <- sum(as.numeric(unlist(int[ (int1 - 1 + i) , idx[[j]] ])))
  }
}

## 중간수요계 등
int2 <- which(int[3, ]=="중간수요계")
for (i in 1:116){
  for (j in 1:17){
    mt116r[ (i+1),(j+117) ] <- sum(as.numeric(unlist(int[ idx[[i]], (int2 - 1 + j) ])))
  }
}

## 디플레이터 116
def <- read.csv("deflator.csv") # 디플레이터
mt116df <- matrix(NA, (117 + 7) , (117 + 17) )
mt116df[2:117, ] <- mt116r[2:117, ]/def[,3]
mt116df[118:(117+6),2:117] <- t(t(mt116r[118:(117+6), 2:117])/def[, 2])
mt116df[(117+7) ,2:117] <- t(t(mt116r[(117+7), 2:117])/def[, 3])

## 부가가치
code65 <- as.numeric(unlist(cde116 ) )
cde65 <- floor( as.numeric(unlist(cde116 ) ) )
mint <- mt116df[(117+6), ][2:117]
mint1 <- data.frame(cbind(cde65, mint))
mint2 <- aggregate(mint1[2], by= mint1[1], sum)
mintrt <- c()
for (i in 1:NROW(cde65) ){
  mintrt[i] <- mint1[i,2]/mint2[mint2[,1]%in%mint1[i ,1],2]
}
mintrt1 <- cbind(code65, mintrt)
adx <- which(mt1524r[,6]=="부가가치")[2]
adv1 <- cbind(as.numeric( unlist( mt1524r[(adx+2):(adx+65), 5] ) ),

```

```

as.numeric( unlist( mt1524r[(adx+2):(adx+65), 6] ) ) )
advl1 <- c()
for (i in 1:NROW(cde65) ){
advl1 [i] <- mintrt1[i,2]*advl[advl[,1]%in%cde65[i],2]
}
advl2 <- cbind(code65, advl1)
## 총투입
tint <- mt116df[(117+7), ] [2:117]
tint1 <- data.frame(cbind(cde65, tint))
tint2 <- aggregate(tint[2], by= tint1[1], sum)
tintrt <- c()
for (i in 1:NROW(cde65) ){
tintrt[i] <- tint1[i,2]/tint2[tint2[,1]%in%tint1[i ,1],2]
}
tintrt1 <- cbind(code65, tintrt)
tipt <- cbind(as.numeric( unlist( mt1524r[(adx+2):(adx+65), 5] ) ),
as.numeric( unlist( mt1524r[(adx+2):(adx+65), 7] ) ) )
tipt1 <- c()
for (i in 1:NROW(cde65) ){
tipt1 [i] <- tintrt1[i,2]*tipt[tipt[,1]%in%cde65[i],2]
}
tipt2 <- cbind(code65, tipt1)

## 한국은행*비율
mt116rt <- matrix(NA, (117 + 7) , (117 + 17) )
mt116rt[(117 + 6), 2:117] <- advl2[,2]
mt116rt[(117 + 7), 2:117] <- tipt2[,2]
mt116rt[(117 + 1), 2:117] <- tipt2[,2] - advl2[,2]
rt1 <- sweep(mt116df[2:117, 2:117], MARGIN = 2, colSums(mt116df[2:117, 2:117]), "/")
rt2 <- sweep(mt116df[(117 + 2):(117 + 5), 2:117], MARGIN = 2, colSums(mt116df[(117
+ 2):(117 + 5), 2:117]), "/")
rt3 <- t(t(rt1)*mt116rt[(117 + 1), 2:117])
rt4 <- t(t(rt2)*mt116rt[(117 + 6), 2:117])
mt116rt[2:117, 2:117] <- rt3
mt116rt[(117 + 2):(117 + 5), 2:117] <- rt4

```

```

inputr <- read_excel("inputr.xlsx", sheet = 1 ) # 총산출 등 실질 국민계정 항목
mt116rt[(117 + 1), (117+c(2,3,4,8,12) )] <- as.numeric(unlist(inputr[10:14, 3])) # 실질
민간, 수출 입력
mt116rt[ which(mt116rt=="NaN", arr.ind = T)[,1], which(mt116rt=="NaN", arr.ind = T)[,2]
] <- 0
## 중간수요계 등, 생산물세 등 0항목은 처리 안함
mt116rt[2:117, (117+1)] <- rowSums(mt116rt[2:117,2:117 ])
for (i in (117+c(2,3,4,8,12)) ){
  rt5 <- mt116df[2:117, i]/sum(mt116df[2:117, i])
  mt116rt[2:117, i] <- t(t(rt5)*mt116rt[(117 + 1), i])
}
mt116rt[2:117, (117+11)] <- mt116rt[(117 + 7), 2:117]
mt116rt[2:117, (117+17)] <- mt116rt[2:117, (117+11)] + mt116rt[2:117, (117+12)]
mt116rt[2:117, (117+10)] <- mt116rt[2:117, (117+17)]
mt116rt[2:117, (117+9)] <- mt116rt[2:117, (117+10)] - mt116rt[2:117, (117+1)]
mt116rt[2:117, (117+6)] <- mt116rt[2:117, (117+9)] - rowSums(mt116rt[2:117,
(117+c(2,3,4,8)) ])

### 116 to 40코드

cde40 <- unlist(mt1524r[2014, 8:(116+8-1)])
int1 <- mt116rt
int1[1,2:117] <- cde40
cde40uq <- sort(unique(cde40))
cde40uq[10] <- cde40uq[2]
cde40uq[2:9] <- c(cde40uq[3:9], "9")
idx <- list()
for (i in 1:41){
  idx[[i]] <- which(as.character(int1[1,i]) == as.character( cde40uq[i]))
}
mt40 <- matrix(NA, (42 + 7) , (42 + 17) )
## 116 코드
for (i in 1:41){
  for (j in 1:41){
    mt40[ (i+1),(j+1) ] <- sum(as.numeric(unlist(int1[ idx[[i]], idx[[j]] ])))
  }
}

```

```

    }
  }
  ## 중간투입계 등
  for (i in 1:7){
    for (j in 1:41){
      mt40[ (i+42),(j+1) ] <- sum(as.numeric(unlist(int1[ (118 -1 + i) , idx[[j]] ])))
    }
  }
  ## 중간수요계 등
  for (i in 1:41){
    for (j in 1:17){
      mt40[ (i+1),(j+42) ] <- sum(as.numeric(unlist(int1[ idx[[i]] , (118 -1 + j) ])))
    }
  }
  ### 명목 최종 과
  mt40

  ##### 취업자 피용자
  # mt165 <- read_excel("3.(고용표)_해양수산업 IO_2015_2017_1213.xlsx", sheet = 3)
  mtchcd <- read_excel("3.(고용표)_해양수산업 IO_2015_2017_1213.xlsx", sheet = 5) # 취업
  자피용자 코드 매핑
  empstat <- read.csv("취업피용.csv") # 취업자 피용자 데이터
  int165 <- int
  # cde165 <- mt165[5:(1524+4), 3]
  cdtmp <- floor(as.numeric( unlist(int[4:(1524+3), 1] ) ))
  cdtmpdec <- as.numeric( unlist(int[4:(1524+3), 1] ) ) - cdtmp
  nmtmp <- as.character( unlist(int[4:(1524+3), 3] ) )

  ## 165 코드로 변경
  nmcd <- mtchcd[3:383, 6:7]
  nmcd1 <- as.character( unlist(nmcd[,1]) )
  nmcd2 <- as.character( unlist(nmcd[,2]) ) # 문제 : "천연가스(LNG) " 띄워쓰기
  for (i in 1:NROW(nmcd)){
    int165[ (intersect(which(cdtmp!=0), which(nmtmp==nmcd1[i]) ) + 3) ,1] <- nmcd2[i]
  }

```



```

int165tmp <- as.character( as.numeric( unlist(int165[4:(1524+3),1]) ) + cdtmpdec )

int165[4:(1524+3),1] <- int165tmp
int165[1, 4:(1524+3)] <- t(int165tmp)

slry <- int165[ which(int165[,3]=="피용자보수"), 4:(1524+3) ]
slry1 <- data.frame( X1=(int165[1, 4:(1524+3)]), X2=(slry) )
slry2 <- aggregate( as.numeric(unlist(slry1[2]) ), by = slry1[1], sum)
slry21 <- slry2[(as.numeric(slry2$X1)-floor(as.numeric(slry2$X1)))!=0, ]

slry211 <- floor(as.numeric(unlist(slry21[1])))

slry3 <- data.frame(cbind( (floor( as.numeric(unlist(slry2[1] )))), (slry2[2]) ))

slry4 <- aggregate( as.numeric(unlist(slry3[2]) ), by = slry3[1], sum)

slry41 <- cbind(slry4[as.numeric(unlist(slry4[1]) ) %in%slry211, 1],
slry4[as.numeric(unlist(slry4[1]) ) %in%slry211, 2])
slryrt <- c()
for (i in 1:NROW(slry211)){
  slryrt[i] <- slry21[i, 2]/slry41[which(slry41[,1]==slry211[i]),2]
}
slryrt1 <- cbind(slry21[1], slryrt)
slryrt2 <- slryrt1[order(as.numeric(unlist(slryrt1[,1])) ), ]

wrk1 <- c();wrk2 <- c()
for (i in 1:NROW(slry211)){
  wrk1[i] <- empstat[which(substring(empstat[,1], 5)==slry211[i]),3]
  wrk2[i] <- empstat[which(substring(empstat[,1], 5)==slry211[i]),4]
}
wrk <- cbind(slry211, wrk1, wrk2)
wrk3 <- wrk[order(wrk[,1]), ]
wrk4 <- cbind( signif(as.numeric(unlist(slryrt2[1])) - sort(slry211), digits = 2)*10,
wrk3[,2:3]*slryrt2[,2])
wrk4 <- data.frame(wrk4)
wrk5 <- aggregate( as.numeric(unlist(wrk4[2]) ), by = wrk4[1], sum)

```

```

wrk6 <- aggregate( as.numeric(unlist(wrk4[3]) ), by = wrk4[1], sum)
# 1.1 ->11
for (i in 1:NROW(wrk5)){
  if (substring(wrk5[i,1], 3)=="1"){
    dgx <- i
  }
}
wrk5[dgx,1] <- wrk5[dgx,1]*10
wrk6[dgx,1] <- wrk6[dgx,1]*10
empfnl <- cbind(wrk5[order(wrk5[,1]), ], wrk6[order(wrk6[,1]), 2 ])

##### 자본스톡(피용자량 같은 방식 코드)
# mt47 <- read_excel("4.(자본스톡)_해양수산업 IO_2015_2017_1212.xlsx", sheet = 4)
mtchcdct <- read_excel("3.1.(자본스톡)_해양중심_취합_연계표_2015.xlsx", sheet = 2) # 자본스톡 코드 매핑
catstat <- read.csv("자본스톡.csv") # 자본스톡 데이터
int47 <- int

## 47 코드로 변경
nmcdct <- mtchcdct[3:383, c(6,11)]
nmcdct1 <- nmcd1 # as.character( unlist(nmcdct[,1]) ) # 문제 : "천연가스(LNG)" 띄워쓰기가 안고쳐짐
nmcdct2 <- as.character( unlist(nmcdct[,2]) )
for (i in 1:NROW(nmcdct)){
  int47[ (intersect(which(cdtmp!=0), which(nmtmp==nmcdct1[i]) ) + 3) ,1] <- nmcdct2[i] # nmtmp=="천연가스(LNG)" i <- 29
}
int47tmp <- as.character( as.numeric( unlist(int47[4:(1524+3),1]) ) + cdtmpdec )
int47[4:(1524+3),1] <- int47tmp
int47[1, 4:(1524+3)] <- t(int47tmp)
cat1 <- int47[ which(int47[,3]=="고정자본소모"), 4:(1524+3) ]
cat1 <- data.frame( X1=int47[1, 4:(1524+3)], E2=cat1 )
cat12 <- aggregate( as.numeric(unlist(cat1[2]) ), by = cat1[1], sum)
cat121 <- cat12[(as.numeric(cat12$X1)-floor(as.numeric(cat12$X1)))!=0, ]

```

```

catl211 <- floor(as.numeric(unlist(catl21[1])))

catl3 <- data.frame(cbind( (floor( as.numeric(unlist(catl2[1] )))), (catl2[2]) ))

catl4 <- aggregate( as.numeric(unlist(catl3[2]) ), by = catl3[1], sum)

catl41 <- cbind(catl4[as.numeric(unlist(catl4[1])) ] %in%catl211, 1],
catl4[as.numeric(unlist(catl4[1])) ] %in%catl211, 2])
catlrt <- c()
for (i in 1:NROW(catl211)){
  catlrt[i] <- catl21[i, 2]/catl41[which(catl41[,1]==catl211[i]),2]
}
catlrt1 <- cbind(catl21[1], catlrt)
catlrt2 <- catlrt1[order(as.numeric(unlist(catlrt1[,1])) ), ]

fxcatl1 <- c()
for (i in 1:NROW(catl211)){
  fxcatl1[i] <- catstat[which(substring(catstat[,1], 5)==catl211[i]),3]
}
fxcatl <- cbind(catl211, fxcatl1)
fxcatl2 <- fxcatl[order(fxcatl[,1]), ]
fxcatl3 <- cbind( signif(as.numeric(unlist(catlrt2[1])) - sort(catl211), digits = 2)*10,
fxcatl2[,2]*catlrt2[,2])
fxcatl4 <- data.frame(fxcatl3)
fxcatl5 <- aggregate( as.numeric(unlist(fxcatl4[2]) ), by = fxcatl4[1], sum)

# 1.1 ->11
for (i in 1:NROW(fxcatl5)){
  if (substring(fxcatl5[i,1], 3)=="1"){
    dgx <- i
  }
}
fxcatl5[dgx,1] <- fxcatl5[dgx,1]*10
catlfnl <- cbind(fxcatl5[order(fxcatl5[,1]), ])

```

```
##### 투자
ivtmt <- read_excel("5.고정자본형성_전환행렬_2005_2010_1230.xlsx", sheet = 3)
ivt1 <- sum(mt40[2:41, 46])
ivt2 <- mt40[(1+19), 46]
ivt3 <- ivt1-ivt2
ivtrt <- mt40[2:41, 46]/ivt3
ivt4 <- ivtrt*ivt3
ivt5 <- ivtrt*ivt2
ivt6 <- ivt4+ivt5
invst <- colSums(ivtmt[46:85, 3:42]*ivt6)[1:10]

### 최종결과(실질)
cname <- read_excel("1.전망모형DB_산업별자료_명목_실질_18년포함__1230.xlsx", sheet =
2) # 최종결과 항목들
year <- rep(substr(colnames(iotable)[1]), 12,15), NROW(fxcatl5))
RES <- cbind(year, mt40[(42+6),2:11], catlfnl$x, empfnl[,2], invst, mt40[2:11, 46],
mt40[2:11, 50], mt40[2:11, 54], empfnl[,3]
, mt40[44,2:11], mt40[43,2:11], t(mt40[45:47,2:11]), mt40[2:11, c(43,44,45,48,51,
53,53, 59,59)] )
RES1 <- cbind(c(1:10), matrix( as.numeric(RES), dim(RES)))
RES1[, c(2,1)] <- RES1[, c(1,2)]
colnames(RES1) <- as.character( cname[1, -2])
write.csv(RES1, "RES1.csv")

### 최종결과(명목) 투자는 없으나 실질 항목들에 맞춤
RESn <- cbind(year, lcsun[(42+6),33:42], catlfnl$x, empfnl[,2], invst, lcsun[33:42, 46],
lcsun[33:42, 50], lcsun[33:42, 54], empfnl[,3], lcsun[44,33:42], lcsun[43,33:42],
t(lcsun[45:47,33:42]), lcsun[33:42, c(43,44,45,48,51, 53,53, 59,59)] )
RESn1 <- cbind(c(1:10), matrix( as.numeric(RESn), dim(RESn)))
RESn1[, c(2,1)] <- RESn1[, c(1,2)]
colnames(RESn1) <- as.character( cname[1, -2])
write.csv(RESn1, "RESn1.csv")
```

## 2. 머신러닝 기법 독립변수

분야	분야별 변수	공통 변수
해운업	컨테이너물동량, 건화물선물동량, SCFI, 상해.유럽.달러, 상해.미.서부.달러, 상해.싱가포르.달러, BDI.종합, BPI, 연료유가, LIBOR.Interest.Rates, USA.Interest.Rates, World.Seaborne.Trade,	(산업연관표) 취업자수, 상용근로자, 실질순자본소득, 투자, 수출, 수입, 상용근로자임금, 중간투입계, 영업잉여, 고정자본소모, 생산세, 중간수요계, 민간소비지출, 재고증가, 최종수요계, 총산출액, 총공급 ※ 총 17개
항만업	World.Seaborne.Container.Trade, 국적선사.톤수, World.Fleet.GT, GDP성장률, Weighted.Average.Earnings.All.Bulkers, CCFI.Composite.Index, Total.Contracting..GT, Total.Containership.Sales,	(거시경제변수) 총GDP, 1인당GDP, 회사채수익률, GDP디플레이터, 민간고정자본형성, 정부고정자본형성, 총고정자본형성, 재화와서비스의수출, 재화와서비스의수입, 총취업자수, 총상용근로자, 총저축률, 총순자본소득, 노동소득분배율, 노동소득분배율1, 노동소득분배율2, 노동소득분배율3, 평균교육기간, 다요소생산성, 총고정자본형성, 고정자본소모, 총인구, 1인당연간노동시간, 경제활동인구, 평균원화환율, 세계GDP, 세계물가 ※ 총 26개
조선업	Total.Containership.Sales.1, Total.Containership.Sales.2, Total.Containership.Sales.3 ※ 총 23개	
수산업	어획노력량, 면세유가격, 수산물총생산량, 수산물총생산액, 수산물평균가격, 수산물수출액, 수산물수입액, 어업부가가치, 어가수, 어가인구, 어가소득, 1인당수산물소비량 ※ 총 12개	
해양관광업	크루즈관광객, 해수욕장방문객, 연간강수일수, 연간열대야일수, 연간폭염일수, 연간황사일수, 도로길이, 고속도로길이, 연안여객선이용객, 동력수상레저기구, 세계관광객 ※ 총 11개	



일반사업 2021-11-2

## 해양수산업 전망모형 개발 연구(Ⅲ)

---

인쇄 2022년 2월 26일

발행 2022년 2월 28일

발행인 김 종 덕

발행처 한국해양수산개발원

주소 49111 부산시 영도구 해양로 301번길 26(동삼동)

연락처 051-797-4800 (FAX 051-797-4810)

등록 1984년 8월 6일 제313-1984-1호

조판·인쇄 등록업체명 (000-000-0000)

---

비매품