

제약하의 예측조합 방법을 활용한 산업별 고용비중 예측

김정우

강릉원주대학교 경제학과 교수

Prediction of the employment ratio by industry using constrained forecast combination

Jeong-Woo Kim

Professor, Department of Economics, Gangneung Wonju National University

요약 본 연구는 우리나라 수출 분야의 산업별 고용비중을 다양한 머신러닝 기법을 활용하여 예측하고, 예측성능을 높이기 위하여 머신러닝 기법 예측값들에 예측조합 기법을 적용하였다. 특히, 본 연구에서는 각 머신러닝 기법 예측값들에 부여되는 가중치의 합을 1로 설정하는 제약하의 예측조합 기법을 사용하여 예측의 정확성과 안정성을 확보하고자 하였다. 또한, 본 연구는 산업별 고용비중에 영향을 주는 다양한 변수를 고려하기 위하여 재귀적특성제거 방법을 사용하여 주요 변수를 선별한 후, 머신러닝 기법에 적용함으로써 예측과정 상에서의 효율성을 높였다. 분석결과, 예측조합 방법에 따른 예측값은 머신러닝 기법의 예측값들보다 실제의 산업 고용비중에 근접한 것으로 나타났으며, 머신러닝 기법의 예측값들이 큰 변동성을 보이는 것과 달리 제약하의 예측조합 기법은 안정적인 예측값을 나타내었다.

주제어 : 머신러닝, 예측, 예측조합, 정규화, 재귀적특성제거

Abstract In this study, we predicted the employment ratio by the export industry using various machine learning methods and verified whether the prediction performance is improved by applying the constrained forecast combination method to these predicted values. In particular, the constrained forecast combination method is known to improve the prediction accuracy and stability by imposing the sum of predicted values' weights up to one. In addition, this study considered various variables affecting the employment ratio of each industry, and so we adopted recursive feature elimination method that allows efficient use of machine learning methods. As a result, the constrained forecast combination showed more accurate prediction performance than the predicted values of the machine learning methods, and in particular, the stability of the prediction performance of the constrained forecast combination was higher than that of other machine learning methods.

Key Words : Machine learning, Prediction, Forecasting combination, Regularization, Recursive feature elimination,

1. 서론

일반적으로 주어진 데이터를 특정 모델로 분석한다는 것은 특정 모델이 변수들 간에 관계를 얼마나 잘 드러내

는지를 의미한다. 예를 들어, 회귀분석에서의 결정계수(R-squared)와 같은 적합도 척도가 높으면 우리는 특정 모델이 주어진 데이터에 대해 가지는 설명력이 높다고 할 수 있다. 하지만, 변수의 개수가 늘어날수록 통계적 유

*Corresponding Author : Jeong-Woo Kim(kurtkim@gwnu.ac.kr)

Received August 24, 2020

Accepted November 20, 2020

Revised October 28, 2020

Published November 28, 2020

의성을 확보하기 위한 데이터의 개수는 많이 요구되므로 주어진 데이터만으로 모집단 수준에서의 변수들 간에 관계를 파악하는 것은 쉽지 않다. 이러한 경우, 분석의 실용적인 목적 중의 하나인 예측력이 감소될 수 있다. 또한, 변수들 간에 관계가 이론적으로 명확히 검증되어 있지 않은 경우에는 데이터의 개수를 늘리는 것이 모델의 예측력을 높인다고 하기는 어렵다. 이에 따라 특정 모델의 주어진 데이터에 대한 설명력과 별도로 모델의 예측력을 고려할 필요가 있다고 볼 수 있다. 주어진 데이터의 설명력을 측정하는 경우에는 주어진 샘플 전체를 활용하여 결정계수와 같은 적합도를 측정한다. 반면에 예측력과 관련해서는, Akaike Information Criterion(AIC), Bayesian Information Criterion(BIC), Mallows's Cp 등 상대적으로 다양한 평가 기준이 존재하며, 일반적인 모델에 대해서는 시험오류검증(Test error validation) 방법을 적용해볼 수 있다. 따라서 학습 모델의 예측력을 평가하는 기준은 다양하며, 하나의 학습 모델에 대한 예측력 평가 기준은 일관적이지 않을 수 있다.

또한, 모든 데이터에 대해 설명력과 예측력이 우수한 완벽한 모델은 존재하지 않으며, 이보다는 특정한 데이터에 적합한 학습 모델을 찾아서 적용하는 것이 주어진 데이터를 효율적으로 분석할 수 있는 접근방식이라고 볼 수 있다. 한편, 동일한 종류의 데이터라고 하더라도 사용하는 샘플의 크기 등에 의해서도 예측을 위한 최적의 모델은 영향을 받을 수 있다. 그러므로 하나의 모델만을 사용하여 주어진 데이터에 기반하여 예측하는 것은 예측의 일관성이 필요한 상황에서는 바람직하지 않은 접근법이라고 볼 수 있다.

예측조합 방법(Forecast combination)은 다수의 모델로부터 얻어진 예측값들을 적절히 조합하여, 하나의 모델에서 얻어진 예측값보다 더 정확하고 안정적인 예측을 하기 위해서 제시된 방법이다[1]. 즉, 예측조합은 다수의 예측값들을 토대로 새로운 예측값을 도출하여 이 값을 실제 값에 학습시킨 후 최종적인 예측값을 산출하는 방법으로, 다양한 예측력 평가 기준들과 샘플의 크기 등에 따라 달라지는 예측력을 고려할 시보다 일관적이고 정확한 예측을 가능하게 하는 방법이다. 또한, 예측조합 방법은 서로 이질적인 학습 모델들을 사용할수록 데이터의 변동에 따른 예측력 감소에 더욱 강건할 수 있는 장점이 있다. 그러므로, 본 연구에서는 다양한 예측 모델을 고려하기 위하여 다수의 머신러닝 기법들에 예측조합 방법을 적용하였다.

본 연구는 상기 방법론을 기반으로, 우리나라의 주요

수출분야인 반도체, 통신기기, 자동차, 조선 산업의 연도별 고용비중과 이와 관련된 경제변수(GDP, 환율 등 총 26개)들을 사용하여 머신러닝 기법과 예측조합 방법의 예측성능을 점검해보았다. 여기서 고용비중이란 전체 10인 이상의 광업제조업 기업들의 종업원 수 대비 해당 산업의 종업수를 의미한다. 각 산업의 종업원 수는 전체 인구 수 및 특성 등에 따라 영향을 받으므로, 광업제조업 내에서의 상대적인 수치인 고용비중은 특정 산업이 우리나라 광업제조업 내에서 고용의 기여 정도를 보다 정확하게 나타낸다고 볼 수 있다. 특히, 고용창출 및 안정성이 무엇보다 중요해진 현 시점에서 산업별 고용비중에 대한 예측연구는 우리나라의 수출산업의 경제 기여도를 진단하는데 유용한 시사점을 지닌다고 볼 수 있다.

2. 이론적 배경

산업별 고용비중은 산업구조 분석, 노동시장 연구 등의 분야에서 많이 다루어지는 자료이다. Yoo et al. (2010)은 제조업 세부산업별 고용비중을 통하여 부산의 산업구조 변화를 분석하였으며[2], Kim(2008)은 산업별 고용비중 자료를 토대로 업종별 고용창출력 분석 연구를 수행하였다[3]. 또한, Kim et al. (2019)은 글로벌 위기 전후로 제조업의 고용구조 변동 분석을 위하여 고용비중 자료를 사용하였고[4]. Jang(2017)은 노동이동성의 변동과 변동원인 분석의 기초자료로써 고용비중 자료를 사용하였다[5].

고용비중 또는 고용에 영향을 주는 변수들은 기본적인 생산요소인 자본, 노동 등의 공급측면의 요소와 정부지출, 민간소비 등의 수요측면 요소 그리고 개방경제를 고려하면 환율, 수출입 규모 등이 있다. Jung et al. (2018)은 자본투자가 산업별 고용성장에 어떠한 영향을 주는지 분석하였으며[6], Song (2009)은 자본투자가 산업별로 고용증대효과를 유도할 수 있다는 실증연구를 수행하였다[7]. 일반적으로 노동생산성은 공장 확충, 신규기계 도입 등의 자본투자에 의해 향상되므로 이에 따라 고용은 증대된다. 반면에 생산된 재화와 서비스가 적시에 소비되지 않는다면 생산요소에 대한 수요 또한 감소하고, 이에 따라 고용에 악영향을 끼칠 수 있다. Kim et al. (2013)은 인구고령화에 따른 민간 소비 감소가 취업에 부정적인 영향을 미친다는 연구결과를 보여주었다[8]. 또한, Kang et al. (2020)은 민간 소비 행태의 변화가 산업간 고용의 증가 또는 감소를 가져온다는 것을 보였는데, 이

러한 결과는 본 연구의 분석대상인 고용비중이 서로 다른 산업들 간에 영향을 받을 수 있다는 것을 의미한다고 볼 수 있다[9]. 정부 재정정책의 목적을 고려할 시 정부지출은 고용에 긍정적인 효과를 준다는 연구가 많이 수행되는 것으로 보인다. Lee et al. (2016)은 재정지출의 변화가 고용형태별로 어떠한 영향을 주는지 분석하였으며, 특히 자본에 대한 재정지출이 고용창출에 긍정적인 영향을 준다는 것을 보였다[10]. 또한, Lee et al. (2017)은 정부의 연구개발투자가 고용에 미치는 영향을 연구하였는데, 특히 연구개발투자는 공공부문에서 고용을 증가한다는 연구결과를 보여주었다[11]. 아울러, 본 연구는 우리나라의 수출산업을 고려하고 있으므로 수출산업의 고용에 영향을 주는 요소로 환율 등의 무역변수를 고려해 볼 수 있다. Park (2008)은 환율변동에 따른 고용변동을 산업별로 분석하였으며, 제조업의 경우 환율 상승이 고용을 증가한다는 연구결과를 내놓았다[12]. Rhee (2011)는 기업규모별로 환율상승이 고용에 끼치는 영향을 분석하였다. 특히, 이 연구에서는 규모가 큰 기업일수록 환율 상승이 고용을 증가시키는 것으로 나타나 우리나라의 수출 분야의 큰 비중을 차지하는 대기업에서의 환율과 고용비중의 관계에 대한 시사점을 보여주었다[13].

상기 연구들과 같이 고용 또는 고용비중은 산업구조 변화 및 원인 파악 등 관심 변수들의 관계 분석을 위한 중간 자료로써 사용되거나, 정책적 관점에서 다양한 경제 변수들이 고용에 영향을 끼치는 정도를 분석하는 연구들에서 많이 활용되었다. 반면에, 산업별 고용비중 자체를 예측 대상변수로 사용한 연구는 많지 않은 편이다. 이에 따라, 우리나라의 주요 수출 분야의 고용비중을 예측하는 작업은 향후 산업별 고용창출을 전망하고 각 산업별 필요한 지원시책을 선제적으로 고려하는데 도움이 될 수 있는 시도라고 볼 수 있다.

방법론적 측면에서, 머신러닝 기법들을 사용하여 산업별 고용비중을 직접 분석한 연구는 거의 없는 편이다. 다만, 머신러닝 기법을 사용하여 고용과 관련된 데이터를 분석한 연구는 상대적으로 많이 이루어진 것으로 보인다. Choi et al. (2019)은 미국의 고용수준을 100여개의 거시경제변수를 활용한 신경망(Neural network) 모형으로 예측하였다[14]. Lee et al. (2020)은 대학교를 졸업한 구직자들의 취업을 예측함에 있어서 신경망 모형, 랜덤포레스트(Random forest) 등의 머신러닝 기법을 적용하였다[15]. Alduayj et al. (2018)은 종업원 이탈을 예측하는 연구에서 서포트벡터머신(Support vector machine, SVM), k-최근접 이웃법(k-Nearest Neighbor,

kNN) 등의 머신러닝 기법을 활용하였다[16]. Adhikari (2009)는 종업원 이탈 요인을 분석하는 연구에서 주성분 분석(Principal component analysis, PCA) 기법을 사용하였다[17]. Mirko(2006)는 산업별 고용률 등의 다양한 변수들을 조합하여 새로운 변수를 생성하는 PCA 중간 과정을 활용하여 유럽 여러 나라들의 고용을 분석하였다[18].

Zhao et al. (2018)은 신경망 모형, SVM, kNN 등 다양한 머신러닝 기법으로 종업원들의 이직률을 예측하는 연구를 수행하였다[19]. 특히, 이 연구는 주어진 데이터 크기를 다양한 방식으로 조정하여 새롭게 생성된 데이터 집합들에 머신러닝 기법을 적용하였는데, 이를 통하여 데이터 개수와 독립 변수 개수의 다양한 조합을 통하여 분석결과의 일반화를 보여주었다. Hsu et al. (2008)은 벡터자기회귀모형(Vector autoregression)에 LASSO(Least Absolute Shrinkage and Selection Operator)를 결합한 기법으로 미국의 실업률에 대하여 거시경제변수가 지니는 예측력을 검증하였다[20]. 특히, 이 연구는 데이터의 개수가 많지 않은 상황에서 LASSO를 통한 변수 선택 효과가 예측력을 높일 수 있다는 것을 검증하였다. Kreiner(2019)는 미국의 실업률을 예측하는 연구에서 경제 관련 변수뿐만 아니라, 인구, 교육 등 다양한 성격의 변수들을 사용하였으며, 이 과정에서 신경망 모형과 LASSO를 적용하였다[21].

본 연구의 주요 자료인 고용과 관련한 변수들은 다양하며 그 변수들이 고용과 가지는 관계도 일관적으로 설명하기는 어렵다. 이러한 이유로 상기 관련 연구들은 전통적인 경제학적 모델보다 머신러닝 기법을 고용 관련 데이터에 적용하였다. 특히, 독립변수의 개수가 많은 경우에는 다중공선성, 과소결정(Underdetermined) 등의 문제가 발생하게 되는데, LASSO, PCA 등의 머신러닝 기법은 이러한 문제를 완화시키는 장점을 가지므로 다양한 독립변수를 다루는 연구에서 활용된 바 있다. 본 연구도 산업별 고용비중에 다양한 변수가 영향을 준다는 점을 고려, 머신러닝 기법을 사용하여 고용비중을 예측하였다.

또한, 상기 기존 연구들의 방법론적 접근방식은 주로 다양한 머신러닝 기법의 예측 결과들을 비교하거나 또는 예측 결과가 가장 정확한 머신러닝 기법을 설명하는 내용들이 대부분이었다. 반면에, 본 연구는 다양한 머신러닝 기법의 예측 결과를 비교할 뿐만 아니라, 해당 결과들에 예측조합 방법을 적용하여 예측력이 향상될 수 있는지 살펴보고자 하였다.

즉, 본 연구는 기존의 고용 관련 연구들이 취업자 수를

자료로 사용하거나, 산업구조를 분석하는 보조자료로서 고용 자료를 사용한 것과는 달리 산업별 고용비중을 자료로 사용함으로써 우리나라의 수출 산업의 경쟁력을 가늠해보고자 하였다. 방법론적인 측면에서는 본 연구는 다양한 머신러닝 기법을 사용할 뿐만 아니라, 각 머신러닝 기법을 종합적으로 고려하는 예측조합 방법을 적용한 점에서 차별성을 지닌다고 하겠다.

이후 장에서부터는 본 연구에서 사용된 머신러닝 기법과 예측조합 방법에 살펴보고, 자료 전처리 및 산업별 고용비중에 대한 머신러닝과 예측조합 방법의 예측 결과를 비교 분석한다. 결론 부분에서는 본 연구의 의의, 향후 필요한 연구 방향 등에 대하여 논의할 것이다.

3. 연구방법

3.1 머신러닝 기법

주어진 데이터를 모두 활용하고 모델의 설명력에만 초점을 맞추어 데이터를 분석하는 경우에는 과적합(Overfitting) 문제가 초래될 수 있다. 과적합 문제가 발생하면 사용된 학습 모델이 주어진 자료는 잘 설명하는 반면, 주어진 자료를 벗어나는 예측값들에 대한 예측력은 저하될 수 있다. 과적합 문제는 학습 모델 복잡도(복잡한 함수 형태 등)가 증가할수록 커지는 경향을 나타내는 것으로 알려져 있다[22].

3.1.1 모델 복잡도 감소 방법

학습 모델의 예측력을 높이기 위하여 모델 복잡도의 문제를 완화하는 방법 중 대표적인 것이 정규화(Regularization) 방법이다. 일반적으로 정규화는 회귀 계수 등의 추정된 파라미터의 크기를 감소시키는 방향으로 학습 모델을 최적화하여 모델의 예측 성능을 높게 된다. 대표적인 모델 정규화 기법으로는 Ridge regression, LASSO 등과 같은 기법이 있다. 특히, LASSO는 회귀분석 과정에서 종속변수에 영향력이 적은 변수의 회귀계수를 0으로 추정할 수 있는 방법이다. 이에 따라, LASSO는 관련 변수의 개수를 줄여주는 방식으로 모델의 복잡도를 감소시킬 수 있으므로, 관측치 개수보다 많은 독립 변수들을 사용한 본 연구에서 적합한 머신러닝 기법 중 하나라고 할 수 있다[23].

LASSO가 기본적으로 하나의 회귀 모형을 설정하고 추정하는데 반면, SVM은 주어진 데이터의 종속변수를 가장 잘 구분하도록 독립변수들을 활용하는 방법이다.

즉, 각 종속변수에 대응되는 독립변수들을 구분하는 함수식을 설정하고, 이 함수식을 조정함에 따라 종속변수가 얼마나 잘 구분되는지를 측정하는 방법이다. 함수식은 보통 독립변수들 간의 거리를 사용하므로 복잡한 함수식이 요구되지 않으며, 이에 따라 SVM에 기반한 서포트벡터 회귀(Support vector regression, SVR) 또한 모델 복잡도를 낮추는 방법으로 사용될 수 있다[24].

3.1.2 부분 데이터 활용 방법

모델 복잡도 문제를 완화하고 예측력을 제고시킬 수 있는 또 다른 방법 중 한가지로써 주어진 데이터 중 일부를 사용하는 방법이 있다. 이러한 접근법은 주어진 데이터를 모두 사용하지 않으므로, 전체 데이터 적합(Fitting)을 위해 필요한 복잡한 함수형태가 필요하지 않으므로 모델 복잡도로 인한 문제를 완화하는데 도움이 될 수 있다. 데이터의 일부분만을 사용하는 방법에는 대표적으로 kNN과 k-평균 군집화(k-means clustering) 등이 있다. kNN 기법은 예측하고자 하는 값에 근접한 자료들만을 추출하여 이 샘플들에 기반한 추정값을 예측값으로 간주하는 비교적 단순한 머신러닝 기법이다. 한편, kNN 기법이 목표 예측값 근방의 자료들만을 활용하는 방법이라면, k-평균 군집화 기법은 전체 샘플을 하위 샘플들로 구분한 후 예측값에 근접한 하위 샘플을 이용하여 예측값을 도출하는 기법이다. 이에 따라, k-평균 군집화 기법은 전체 샘플의 특성을 고려할 수 있으며, 비교적 샘플의 크기가 큰 경우에도 활용이 가능하다. 이 두 가지 기법들은 추출된 부분 데이터에 비교적 단순한 형태의 함수(평균값, 선형함수 등)를 적용함으로써 모델 복잡도를 낮추고 예측력을 높이는 데에 유용한 방법으로 쓰일 수 있다.

본 연구에서는 비교적 많은 수의 독립변수가 사용되므로 다중공선성 문제가 발생할 수 있다. PCA는 데이터의 각 변수에 가중치를 부여하여 가중치 평균 형태의 새로운 변수들을 생성한 후, 이 새로운 변수들을 직교변환하는 기법이다. 이에 따라, 새로운 변수들은 서로 직교하므로 다중공선성 문제가 제거될 수 있다. 주성분회귀(Principal component regression, PCR)는 이 새로운 변수들을 독립변수로 간주하여 회귀분석에 적용한 것으로서 다중회귀분석에서 흔히 발생하는 다중공선성 문제를 감소시키는 회귀분석 중 하나라고 볼 수 있다.

상기 머신러닝 기법들이 기본적으로 데이터의 특성을 고려하여 개발된 기법이라면 신경망 모형은 예측성능만을 높이기 위하여 독립변수들을 활용하는 방법이라고 할

수 있다. 우선, 신경망 모형은 독립 변수들 간의 가중치를 다르게 설정하여 새로운 변수들을 생성하고, 이 새로운 변수들을 종속변수에 적합하는 과정을 예측력이 가장 높아지는 시점까지 반복하는 기법이다. 또한, 신경망 모형은 새로운 독립변수를 종속변수에 적합시키는 과정에서 발생가능한 과적합 문제를 역전파(Backpropagation)라는 과정을 통하여 완화시킨다. 역전파는 예측값과 실제값의 차이를 변수의 가중치에 반영하는 방법으로 신경망 모형이 높은 예측력을 가지게 되는데 중요한 역할을 한다고 볼 수 있다.

3.2 예측조합 방법

본 연구에서 쓰인 관측치 개수는 2005년부터 2018년까지 총 14개로 독립변수의 개수(26개)보다 적으므로 예측값들의 분산이 클 수가 있다. 예측조합 기법을 사용하면 다수의 예측값들을 가중평균하여 최종 예측값을 도출함으로써 이러한 문제를 완화할 수 있다. 예측조합 기법의 기본적인 수식은 식 (1)과 같다.

$$f_c = \sum_{i=1}^m \omega_i f_i \tag{1}$$

여기서, f_c 는 예측조합에 따른 예측값, m 은 예측 모델의 개수, w_i 는 각 예측 모델의 예측값에 부여되는 가중치, f_i 는 각 예측 모델의 예측값임.

가중치 w_i 를 정하는 방식에 따라 다양한 예측조합 기법이 존재한다. 각 예측값에 동일하게 가중치를 부여하는 방법은 간단하지만 실용적인 측면에서 자주 활용되는 예측조합 기법이다[25]. 또한, 예측값들의 평균제곱오차(Mean squared error)에 반비례하는 가중치를 부여하는 방법[1], 각 예측 값들을 실제 값에 회귀하여 가중치를 결정하는 방법[26] 등이 있다.

이와 같은 예측조합 방법 중 회귀계수의 가중치를 사용하는 예측조합 방법은 최소자승법에 기반을 두는 직관적인 방법이나 과적합 문제가 발생할 수 있다. 이러한 문제를 완화하기 위하여 제안된 예측조합 기법이 제약하의 예측조합 기법이다[27]. 제약하의 예측조합 기법은 회귀계수들에 제약(회귀계수의 합을 1로 제약)을 부여하는 정규화 방식을 따르고 있다. 제약하의 예측조합 방법에서의 가중치는 식 (2)와 같이 구해진다.

$$\omega_i = \underset{j}{\operatorname{argmin}} \sum_{j=n}^l (y_j - f_{j_c})^2 \tag{2}$$

$$\text{s.t. } \sum_{i=1}^m \omega_i = 1$$

여기서, f_{j_c} 는 j 번째 실제값에 대응하는 예측조합 기법에 따른 예측값임.

식 (2)와 같이 가중치에 대한 제약은 특정 예측기법에 과도한 과중치를 부여하게 되는 문제점을 방지하여 과적합의 문제가 완화할 수 있으므로 안정적인 예측값을 얻는데 도움이 될 수 있다.

3.3 자료 전처리

머신러닝 기법과 예측조합 방법을 사용한 산업별 고용비중 예측결과를 알아보기 전에, 본 연구에서 사용된 데이터의 특성과 이에 따른 전처리 과정을 살펴보고자 한다. 본 연구에서는 산업별 고용비중을 예측하기 위하여 26개의 독립변수가 사용되었다. 독립변수들은 우리나라의 자본, 노동, 임금, 수출, 수입 등 국내 경제변수와 유가, 전 세계 GDP 등 국제 경제변수로 구성된다.

총 관측치 수는 14개(2005년 ~ 2018년)로 사용하는 독립변수의 개수보다 적다. 이에 따라, 주어진 데이터에 회귀분석을 하게 되면 과소결정 문제가 발생하게 된다. 또한, 서로 관련된 많은 독립변수를 사용하는 경우 다중

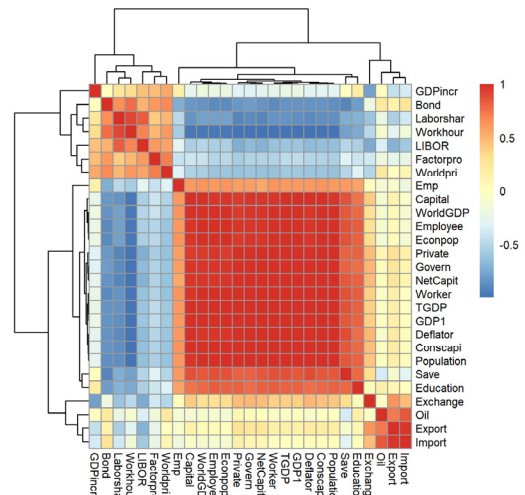


Fig. 1. Heatmap of correlation coefficient (semiconductor)

공선성 문제가 발생하는 경우가 있어 학습 모델의 예측 성능이 저하될 수 있다. Fig. 1은 본 연구에서 사용된 변수들 간의 상관계수를 Heatmap으로 나타낸 것이며, 모눈의 색이 진할수록 높은 양(+) 또는 음(-)의 상관관계를 나타내고 있다. 대부분의 변수들이 짙은 색의 모눈을 나타내고 있으므로 본 연구에서 사용되는 독립변수들은 밀접하게 연관된 경우가 많은 것으로 파악된다. 이에 따라, 본 연구에서 머신러닝 모델을 데이터에 적용하기 이전 단계에서, 독립변수들을 선별하는 작업이 필요할 것으로 판단된다.

재귀적특성제거(Recursive feature elimination, RFE)는 이러한 독립변수 선별에 사용되는 기법들 중 하나로써, 특히 샘플의 개수가 적은 경우에도 적용가능한 방법이다[28]. RFE는 우선 모든 독립변수를 종속변수에 적합시킨 후 종속변수의 설명력이 가장 높아지는 시점까지 독립변수들을 제거해나가는 후진제거법(Backward selection) 방식을 취하고 있다. 본 연구에서는 데이터 개수보다 많은 독립변수의 개수를 고려하여 머신러닝 방법 중 하나인 랜덤포레스트를 독립변수를 종속변수에 적합한 과정에 사용하였다. RFE의 방법의 알고리즘은 Fig. 2와 같다[29].

```

Algorithm: Recursive feature elimination
1  Tune/train the model on the training set using all predictors
2  Calculate model performance
3  Calculate variable importance or rankings
4  for Each subset size  $S_j$ ,  $j = 1 \dots S$  do
5      Keep the  $S_j$  most important variables
6      [Optional] Pre-process the data
7      Tune/train the model on the training set using  $S_j$  predictors
8      Calculate model performance
9      [Optional] Recalculate the rankings for each predictor
10 end
11 Calculate the performance profile over the  $S_j$ 
12 Determine the appropriate number of predictors
13 Use the model corresponding to the optimal  $S_j$ 
    
```

Fig. 2. RFE Algorithm

4. 연구결과

본 연구에서는 주어진 데이터에 RFE로 자료 전처리 후, 6개의 머신러닝 기법과 예측조합 기법을 적용하여 각 산업별 고용비중을 예측하고 각 예측기법들의 예측 성능을 비교해보았다. 2005년부터 2017년까지의 데이터를

토대로 2013년부터 2018년까지의 예측 값들을 구하여 각 머신러닝 기법과 예측조합 기법의 예측결과들을 산출하였다.

Fig. 3과 같이 우리나라의 주요 수출 품목 중 고용비중이 높은 산업은 자동차(Car)이며 조선업(Ship)과 반도체(Semi)가 그 뒤를 잇고 있다. 특히, 자동차와 조선 분야는 대표적인 노동집약적인 산업으로 상대적으로 높은 고용비중이 유지되고 있는 것으로 보인다. 통신기기(Comm)의 경우 중국 등의 후발업체들의 휴대폰 시장 확대로 고용비중이 점점 낮아지고 있는 것으로 파악된다.

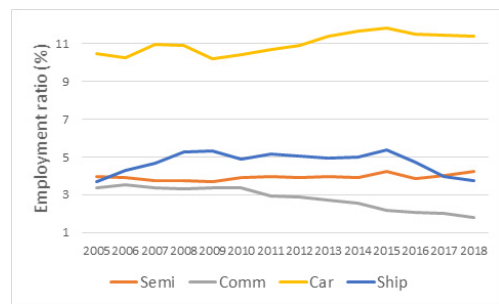


Fig. 3. Employment ratio by industry

또한, 기술통계량에서도 통신기기의 고용비중은 평균 값과 중간값이 가장 낮은 수준이나, 표준편차가 상당히 높은 편으로 나타나 통신기기의 고용비중은 시장 환경에 따라 변동성이 큰 것으로 파악된다. 이러한 변동성이 큰 데이터 특성은 예측 결과에도 영향을 줄 수 있다고 볼 수 있다.

Table 1. Descriptive statistics (unit: %)

	MEAN	STD	MIN	0.25	0.5	0.75	MAX
Semi	3.94	0.16	3.71	3.87	3.92	3.99	4.24
Comm	2.81	0.60	1.82	2.28	2.89	3.35	3.51
Car	11.01	0.53	10.21	10.53	10.95	11.43	11.80
Ship	4.71	0.57	3.70	4.39	4.90	5.11	5.35

* STD: standard deviation

예측 결과를 살펴보기 전에, RFE 방법에 따라 선택된 변수들에 대하여 알아보하고자 한다. Table 2에서와 같이 각 연도별로 선택된 변수들은 차이를 보이고 있다. 반면에 특정 변수들은 여러 연도에 걸쳐 중요성이 높은 변수로 RFE에 의해 선별된 것이 발견된다. 예를 들어, 반도체 산업에서는 회사채율(Bond)이 대부분의 연도에 걸쳐 중요한 변수로 선별되었다. 이는 우리나라의 반도체 산업의

발전에 따라 반도체 회사에 많은 투자가 이루어져 회사 채용도 함께 상승한 것에 기인한다고 볼 수 있다. 한편, 이와 같은 변수 관계에 대한 이론적 설명은 충분한 양의 데이터 개수가 확보되어야만 통계적 유의성에 의해 뒷받침될 수 있다. 하지만, 본 연구는 변수간의 이론적인 관계를 규명하기보다는 주어진 데이터를 기반으로 예측만을 목적으로 하고 있으므로 RFE에 의해 선별된 공통된 변수들을 머신러닝 기법에 우선 사용하기로 한다.

Table 2. Selected variables by RFE

	Semi	Comm	Car	Ship
Total	TGDP, GDP1, Bond, Deflator, Private, Govern, Export, Import, Employee, Worker, Save, NetCapit, Laborshar, Education, Factorpro, Capital, Conscapi, Population, Workhour, Econpop, Exchange, WorldGDP, Worldpri, GDPIncr, LIBOR, Oil			
2014	Save, Bond, Employee, Econpop, Private, etc (Total 8)	Govern, Bond, Workhour, Worker, Population, etc (Total 16)	Oil, Education, GDPIncr, Private, Workhour, etc (Total 26)	NetCapit, Govern, Deflator, GDPIncr
2015	Bond, Save, Factorpro, Workhour, Worker, etc (Total 16)	Conscapi, Private, Workhour, Govern, NetCapit, etc (Total 8)	Deflator, Education, Private, WorldGDP, Bond, etc (Total 8)	GDPIncr, World GDP, Import, Exchange
2016	Bond, LIBOR, Save, Factorpro, Laborshar, etc (Total 16)	GDP1, Workhour, Bond, Private, Population, etc (Total 16)	Population, Education, Worker, WorldGDP	Import, Deflator, Employee, Export, Conscapi, etc (Total 16)
2017	Bond, Save, Laborshar, Factorpro, LIBOR, etc (Total 16)	Bond, Workhour, TGP, Govern, WorldGDP, etc (Total 16)	Private, Workhour, WorldGDP, Education	Import, Export, GDPIncr, LIBOR, Laborshar, etc (Total 16)

* Variables are listed by importance

Fig. 4는 각 산업별로 머신러닝 기법과 예측조합 기법에 의한 예측결과를 보여주고 있다. 붉은색 쇠선이 실제 산업별 고용비중(Emp)이며 검은색 쇠선이 제약하의 예측조합 기법에 따른 예측값(ForeComb)이다. 제약하의 예측조합 기법에 따른 예측값은 이전의 머신러닝 기법 예측값들을 기반으로 얻을 수 있으므로, 2016년부터 2018년까지의 예측값만 제시되었다.

Fig. 4와 같이 산업별 및 연도별로 우수한 예측 성능을 일관되게 보이는 머신러닝 기법은 존재하지 않는다. 반면에, 제약하의 예측조합에 의한 예측값은 대부분 실제 고용비중에 근접한 모습을 보여주고 있으며, 특히 분산이 적은 것으로 관찰된다. 이것은 제약하의 예측조합 기법이 다양한 머신러닝 기법의 가중치 평균이라는 점에 기인하는 것으로, 안정적인 예측값을 얻는데 있어서 예측조합 기법을 사용하는 것이 유리하다는 것을 보여준다고 할 수 있다.

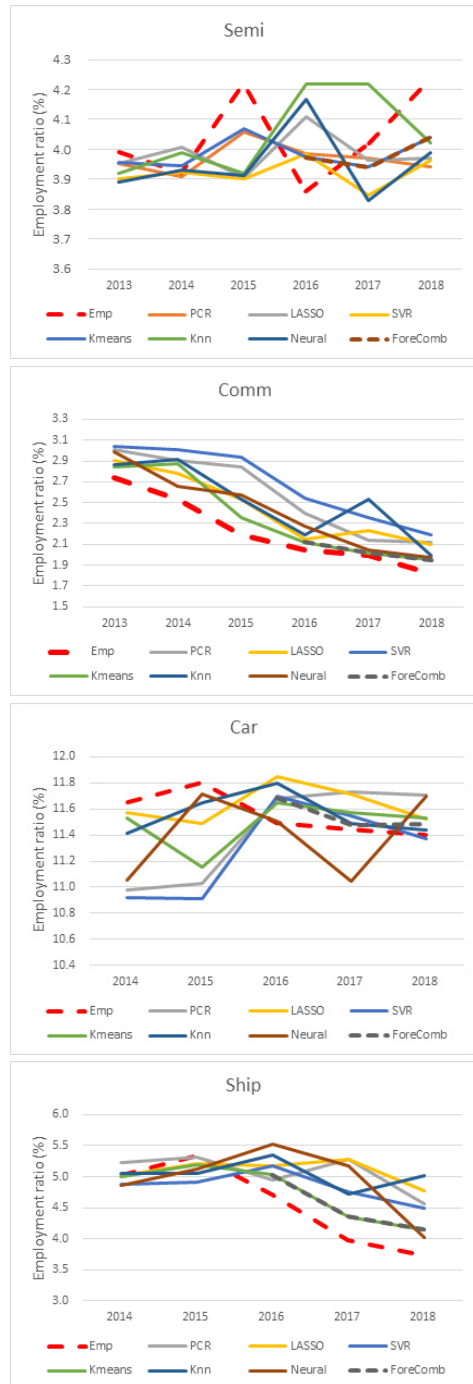


Fig. 4. Comparison of prediction results

아울러, 각 기법들의 예측성능을 보다 정확하게 비교하기 위하여 예측오차의 평균절대비오차(Mean Absolute Percentage Error, MAPE)를 구하였다. MAPE의 수식

은 식 (3)과 같다.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - f_i|}{y_i} \times 100 \quad (3)$$

여기서, N 은 데이터 개수, y_i 는 실제 GDP값, f_i 는 예측값임.

각 머신러닝 기법은 산업별로 상이한 예측성능을 나타내고 있다. 예를 들어, LASSO의 경우에는 반도체와 조선업에서는 연도별로 높은 MAPE를 나타내고 있어 예측성능이 높지 않으나, 타 산업에서는 타 예측 기법과 유사한 MAPE 값들을 나타내고 있다. 아울러, 통신기기의 경우

선형모형에 바탕을 두고 있는 PCR, LASSO, SVR의 MAPE 값들이 나머지 머신러닝 기법의 MAPE 값들보다 높게 나타나고 있는데, 이것은 앞서 기술통계량에서 언급된 바와 같이 통신기기 고용비중의 높은 변동성에 영향에 기인한다고 볼 수 있다.

한편, 제약하의 예측조합 기법의 MAPE는 다른 머신러닝 기법들의 MAPE에 비해 전반적으로 낮은 수준으로 나타났다. 자동차 산업의 경우에서 제약하의 예측조합 기법의 예측성능이 가장 좋은 것으로 나타나고 있는데, 이것은 자동차 산업에서 비교적 낮은 MAPE를 보이는 SVR과 kNN 기법이 제약하의 예측조합 기법에 영향을 준 것으로 볼 수 있다. 또한, 각 머신러닝 기법의 연도별 MAPE 값의 표준편차에 비해 제약하의 예측조합 기법이 MAPE 값의 표준편차가 대체적으로 적은 것으로 나타나 안정적인 예측성능을 보여준다고 볼 수 있다. 특히, 2017년과 2018년의 통신기기와 조선 분야에서 일부 머신러닝 기법의 MAPE가 상당히 높게 나타났지만, 제약하의 예측조합 기법은 안정적인 예측성능을 보였다. 그러므로 위험기피적인 예측을 목적으로 하는 연구 또는 관련 작업에서는 제약하의 예측조합 기법이 유용성이 크다고 볼 수 있다.

Table 3. Comparison of MAPE (unit: %)

Year	PCR	LASSO	SVR	Kmeans	Knn	Neural	Fore Comb
Semi							
2013	0.90	0.91	2.22	0.88	1.75	2.53	
2014	0.29	2.20	0.05	0.66	1.79	0.31	
2015	3.84	7.46	7.53	3.55	7.11	7.30	
2016	3.29	6.47	3.29	2.92	9.33	7.94	2.92
2017	1.18	1.44	4.27	1.99	4.98	4.81	1.99
2018	7.00	6.35	6.54	4.72	5.19	5.96	4.72
MEAN	2.75	4.14	3.98	2.45	5.02	4.81	3.21
STD	2.51	2.93	2.77	1.58	2.97	2.92	1.39
Comm							
2013	9.79	5.97	10.82	3.65	4.38	8.94	
2014	14.92	10.05	19.13	13.75	15.42	5.01	
2015	29.95	15.74	34.21	7.76	15.53	17.60	
2016	17.41	5.25	24.58	3.68	7.35	11.60	3.84
2017	7.48	11.93	18.42	1.26	27.14	2.87	1.26
2018	16.26	15.29	20.25	7.14	9.34	8.58	7.14
MEAN	15.97	10.70	21.23	6.21	13.19	9.10	4.08
STD	7.86	4.48	7.76	4.42	8.14	5.18	2.95
Car							
2013	6.67	6.99	5.10	3.51	4.21	1.38	
2014	5.75	0.66	6.23	1.03	2.06	5.14	
2015	6.57	2.66	7.50	5.49	1.27	0.72	
2016	1.63	3.09	1.80	1.35	2.70	0.13	1.71
2017	2.54	2.39	0.93	1.19	0.44	3.45	0.28
2018	2.70	1.09	0.22	1.16	0.35	2.60	0.78
MEAN	4.31	2.81	3.63	2.29	1.84	2.23	0.92
STD	2.26	2.25	3.04	1.83	1.48	1.87	0.72
Ship							
2013	8.32	10.81	2.52	1.52	2.85	0.74	
2014	4.29	0.00	2.67	0.40	0.80	3.31	
2015	0.64	2.67	8.10	3.08	5.42	4.37	
2016	4.72	9.69	9.58	6.57	13.35	17.24	6.56
2017	32.51	32.90	19.69	9.30	18.59	30.04	9.52
2018	22.11	27.79	20.65	11.08	34.58	7.82	11.08
MEAN	12.10	13.98	10.53	5.33	12.60	10.59	9.05
STD	12.47	13.42	7.99	4.34	12.69	11.13	2.30

5. 결론

본 연구에서는 다양한 머신러닝 기법으로 우리나라의 주요 수출분야인 반도체, 통신기기, 자동차, 조선업의 고용비중을 예측하여 각 기법들의 예측 성능을 비교해보았다. 또한, 제약하의 예측조합 기법을 머신러닝 기법에 따른 예측값들에 적용하여 예측 성능의 개선 여부를 검증해보았다. 아울러, 많은 실용적 문제에서 직면하는 독립변수 개수보다 적은 관측치가 주어지는 상황을 고려하여 재귀적특성제거 기법으로 자료 전처리를 수행하였다. 이에 따라 각 산업별 또는 연도별로 고용비중에 영향을 주는 중요한 변수들을 선별하여 각 머신러닝 기법에 적용하였다. 한편, 이러한 머신러닝 기반의 고용데이터 분석 연구들은 기존에도 이루어진 바 있다. 하지만, 본 연구는 수출산업 특성상 변동이 심한 데이터를 안정적으로 예측하기 위하여 제약하의 예측조합 기법을 사용하였고, 또한 고용규모가 아닌 고용비중을 사용하여 특정 산업의 상대적 경쟁력을 진단하고자 했다는 점에서 연구의 차별성을 찾고자 하였다.

산업별 고용비중의 예측결과를 살펴보면 반도체 산업

을 제외한 다른 산업분야에서는 실제 고용비중이 감소하는 추세에 있는 것으로 나타났다. 또한, 제약하의 예측조합 기법의 예측값도 동일한 감소추세를 보여주었다. 이러한 경향은 실제 우리나라의 수출 산업 현황을 반영하고 있다고 볼 수 있다. 이는 기술력이 가장 증시되는 반도체 산업의 경우 타 경쟁국가 대비 기술우위에 있어 생산 및 고용에서 안정된 증가 추세를 유지하고 있으나, 조립, 가공 공정이 많이 개입되는 조선 등의 산업의 경우에는 다른 나라와의 경쟁이 심화되고 있기 때문이다. 특히, 휴대전화, 자동차 산업의 경우 국내 기업들의 국외 생산 비중이 증가한 것이 고용비중의 감소 추세에 영향을 준 것으로 보인다. 예측결과와 같이 제약하의 예측조합 기법은 실제 고용비중과 가장 근접한 예측값을 나타내었다. 특히, 예측값의 변동성이 적다는 점은 기업 또는 정부 등의 기관에서 이루어지는 의사결정 과정에서 예측조합 기법이 실용적으로 활용될 수 있다. 예를 들어, 본 연구에서 예측조합 기법은 비교적 적은 수의 예측치들을 토대로 안정된 예측성능을 보여주었다. 이러한 점은 정부 또는 기업 등에서 적은 양의 예측자료들로부터 안정된 예측결과를 도출하여 실무 현장에서 활용할 수 있다는 점을 시사한다고 볼 수 있다.

본 연구에서 사용된 예측 조합 기법은 기본적으로 가중치의 합이 1이 되는 제약하의 가중치 평균이므로 예측조합의 기초가 되는 머신러닝 기법들의 예측 성능을 언제나 앞선다고는 할 수 없다. 하지만, 예측이라는 것의 중요한 실용적 특징을 예측의 안정성이라고 본다면 제약하의 예측조합 기법은 가중치 평균을 통하여 예측값의 분산을 감소시킬 수 있고, 이에 따라 안정적이고 정확한 예측을 도모할 수 있다. 실제로 본 연구의 MAPE 결과에서도 제약하의 예측조합 기법은 다른 머신러닝 기법들의 보다 대부분 적은 MAPE 값을 보였으며, 특히 MAPE의 표준편차는 상당히 적은 것으로 나타났다.

본 연구에서는 머신러닝 기법 중 다중공선성 문제를 고려한 PCR 기법을 사용하였다. PCR 기법은 기존 변수들에 서로 다른 가중치를 부여하여 새로운 변수들을 생성해내고 이 새로운 변수들의 직교성에 기반하여 다중공선성 문제를 완화하는 기법이다. 그러므로 이 가중치는 독립변수와 종속변수 간의 관계에서 중요한 의미를 가질 수 있다. 하지만, 본 연구에서는 독립변수 대비 부족한 관측치 규모로 인하여 독립변수들의 가중치들이 연도별, 산업별로 일관적으로 추정되지 않았다. 이러한 문제는 향후 충분한 양의 관측치를 확보함으로써 해결될 수 있을 것으로 보이며, 또한 자료 전처리 과정에서 다양한 적합 기

법을 적용하여 관측치보다 적은 변수를 선별하는 적합 기법을 사용하는 것도 향후에는 고려해볼 수 있다.

아울러, 본 연구는 다양한 예측기법들의 예측성능을 비교하였으므로 사후검정 또는 다중비교를 통하여 예측성능을 검증하는 것도 의미가 있다고 볼 수 있다. 반면에 본 연구는 Table 1에서와 같이 예측값의 개수가 3개 연도에 한정되어 사후검정이나 다중비교를 수행하기에는 애로가 있다. 하지만, 본 연구는 다양한 예측기법들의 예측성능을 비교하기보다는 제약하의 예측조합 기법을 통하여 예측의 정확성과 안정성 개선 여부에 의미를 두고 있으므로, 충분한 관측치와 예측값을 활용한 사후검정 또는 다중비교 분석은 향후 연구과제로 남겨두기로 한다.

본 연구에서 사용된 제약하의 예측조합 기법은 이전의 머신러닝 예측값들을 제약조건 하에서 실제값에 적합시켜 가중치를 계산하고, 이를 통하여 새로운 예측값을 도출하는 방식을 따르고 있다. 여기서, 얼마나 많은 이전의 머신러닝 예측값들을 예측조합 기법에서 고려해야하는가에 대한 문제를 생각해볼 수 있다. 데이터의 종속변수가 예측 시점주변에서 불규칙한 양상을 보인다면, 비교적 최근의 예측값들만을 예측조합 기법에서 고려하는 것이 유리할 수 있다. 하지만, 이러한 경우 예측 결과의 분산이 커져서 불안정한 예측이 될 수 있는 단점이 있을 수 있다. 즉, 제약하의 예측조합 기법에서도 편향-분산 트레이드 오프(Bias-variance tradeoff) 문제가 개입될 수 있는 것이다. 이에 따라, 예측조합 기법에 포함되는 이전 예측값들의 개수를 데이터의 종속변수의 양상에 따라 결정하는 새로운 방식의 예측조합 기법의 개발이 향후에는 필요할 것으로 보인다.

REFERENCES

- [1] J. M. Bates & C. W. J. Granger. (1969). The Combination of Forecasts. *Journal of the Operational Research Society*, 20(4), 451-468. DOI : 10.1057/jors.1969.103
- [2] Y. M. Yoo, H. B. Kim & S. H. Joo. (2010). An Analysis on the Competitiveness based on Industrial Structure Changes in Busan. *Journal of Local Government Studis*, 14(3), 295-313.
- [3] B. S. Kim. (2008). Job Creation of Service Industry and Changes in Employment Structure in Terms of the Quality of Job. *Monthly Labor Review*, 40, 23-35.
- [4] S. J. Kim & B. H. Choi. Input-Output Structural Decomposition Analysis on the Structure of Employment Change in Korean Manufacturing

- Industry. *Journal of Industrial Economics and Business*, 32(1), 375-403.
DOI : 10.22558/jieb.2019.02.32.1.375.
- [5] I. S. Jang. (2017). *The effect of labor mobility on the relationship between technology innovation and employment*. Sejong : KLI.
- [6] L. K. Chung & J. B. Hong. (2018). An Empirical Study on the Relationship between Investment and Employment Growth : Focusing on the Differences between Industries. *Korean Journal of Business Administration*, 31(7), 1363-1382.
DOI : 10.18032/kaaba.2018.31.7.1363.
- [7] I. H. Song. (2009). An Economic Analysis on the Effect of Facility Investment on Productivity and Employment. *Productivity Review*, 23(3), 259-278.
- [8] D. K. Kim & S. Y. Park. (2013). Impact on Production and Employment of Consumption Structure Changes by Increasing Proportion of Old Population. *Journal of Industrial Economics and Business*, 26(6), 2519-2546.
- [9] D. W. Kang. (2020). *Effect of change in consumption patterns on employment*. Sejong : KLI.
- [10] J. H. Lee & J. Y. Hwang. (2016). The Employment Creation Effects of Exogenous Fiscal Shocks in Korea. *Ordo Economics Journal*, 19(1), 19-40.
DOI : 10.20436/OEJ.2016.19.1.019.
- [11] J. H. Lee & E. S. Lim. (2017). R&D Investment and Employment: Evidence from 15 Regions. *The Korea Journal of Local Public Finance*, 22(2), 167-190.
- [12] M. S. Park. (2008). *Analysis of dynamic relationship between exchange rate fluctuations and employment*. Sejong : KIET.
- [13] C. O. Rhee. (2011). Special Section Papers : Business Cycle and SME's Government-Supported Financing. *Asia Pacific Journal of Small Business*, 33(1), 17-32.
- [14] J. E. Choi & D. W. Shin. (2019). The roles of differencing and dimension reduction in machine learning forecasting of employment level using the FRED big data. *Communications for Statistical Applications and Methods*, 26(5), 497-506.
DOI : 10.29220/CSAM.2019.26.5.497
- [15] D. H. Lee & T. H. Kim. (2020). Study on the Prediction Model for Employment of University Graduates Using Machine Learning Classification. *The Journal of Information Systems*, 29(2), 287-306.
DOI : 10.5859/KAIS.2020.29.2.287
- [16] S. S. Alduayj & K. Rajpoot. (2018, 18-19 Nov. 2018). Predicting Employee Attrition using Machine Learning. *Paper presented at the 2018 International Conference on Innovations in Information Technology (IIT)*.
DOI : 10.1109/INNOVATIONS.2018.8605976
- [17] A. Adhikari. (2009). Factors affecting employee attrition: a multiple regression approach. *IUP Journal of Management Research*, 8(5), 38.
- [18] M. Savić. (2006). Principal components analysis of employment in Eastern Europe. *Panoeconomicus*, 53(4), 427-437.
DOI : 10.2298/PAN0604427S
- [19] Y. Zhao, M. K. Hryniewicki, F. Cheng, B. Fu & X. Zhu. (2018). Employee turnover prediction with machine learning: A reliable approach. *Paper presented at the Proceedings of SAI intelligent systems conference*.
DOI : 10.1007/978-3-030-01057-7
- [20] N. J. Hsu, H. L. Hung & Y. M. Chang. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, 52(7), 3645-3657.
DOI : 10.1016/j.csda.2007.12.004
- [21] A. Kreiner & J. V. Duca. (2019). Can machine learning on economic data better forecast the unemployment rate? *Applied Economics Letters*, 27(17), 1434-1437.
DOI : 10.1080/13504851.2019.1688237
- [22] T. Hastie, R. Tibshirani & J. Friedman. (2009). *The elements of statistical learning: data mining, inference, and prediction*, Springer Science & Business Media.
DOI : 10.1111/j.1751-5823.2009.00095_18.x
- [23] R. Tibshirani. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
DOI : 10.1111/j.1467-9868.2011.00771.x
- [24] V. Cherkassky & Y. Ma. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 17(1), 113-126.
DOI : 10.1016/S0893-6080(03)00169-2
- [25] R. T. Clemen. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559-583.
DOI : 10.1016/0169-2070(89)90012-5
- [26] C. W. Granger & R. Ramanathan. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2), 197-204.
DOI : 10.1002/for.3980030207
- [27] C. Aksu & S. I. Gunter. (1992). An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. *International Journal of Forecasting*, 8(1), 27-43.
DOI : https://doi.org/10.1016/0169-2070(92)90005-T
- [28] I. Guyon, J. Weston, S. Barnhill & V. Vapnik (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3), 389-422.
DOI : 10.1023/A:1012487302797
- [29] K. E. Rao & G. A. (2020). Rao, Ensemble learning with recursive feature elimination integrated software effort estimation: a novel approach. *Evolutionary Intelligence*, 1-12.
DOI : 10.1007/s12065-020-00360-5

김 정 우(Jeong-Woo Kim)

[정회원]



- 2005년 8월 : 고려대학교 경제학(경제학사)
- 2012년 8월 : 연세대학교 경제학(경제학석사)
- 2018년 2월 : 연세대학교 경제학(경제학박사)
- 2020년 3월 ~ 현재 : 강릉원주대학교

경제학과 교수

- 관심분야 : 계량경제학, 머신러닝
- E-Mail : kurtkim@gwnu.ac.kr