

국제경제연구
제28권 제4호
2022년 12월

머신러닝 방법을 활용한 한국의 수출입 증가율 예측 및 분석*

장나원** · 한희준***

본고는 수출입 관련 국내의 거시경제/금융 변수들을 폭넓게 사용하고 다양한 머신러닝 방법을 활용하여 한국의 수출입 증가율을 예측하고 분석한다. 2000년 1월에서 2021년 12월까지의 표본(총 264개월) 중 최근 10년(120개월)을 예측 기간으로 삼았으며, 1개월에서 12개월까지를 예측 대상기간(forecast horizon)으로 고려한다. 총 13개의 모형 중 모든 예측 대상기간에 있어 예측력이 우수한 모형이 존재하는 것으로 나타났는데, 이는 Kim and Han(2022)이 제안한 것으로 구체적으로 1) 보루타 알고리즘(Boruta algorism)을 통한 변수 중요도의 순위 도출, 2) 교차검증(cross-validation)을 통한 최적 변수의 수 선택, 3) 중요도가 높은 변수들을 선택된 수만큼만 사용하여 랜덤 포레스트(random forest)를 실시하는 것이다. Giacomini and White (2006)와 Hansen et al.(2009)의 검정을 통해 대부분 통계적으로 유의하게 예측력이 우수함을 확인하였고, 특히 에너지사용량이 수출과 수입 예측에 중요한 변수로 채택되는 것으로 나타났다.

핵심용어: 수출입, 예측, 머신러닝, 랜덤 포레스트, 보루타 알고리즘
경제학 문헌분류기호: C32, C50, F16

I. 서 론

2021년 우리나라는 사상 최대 수출실적인 6,444억 불을 달성하였고, 2022년에도 선박을 제외한 주력 품목들의 수출 호조세를 유지하며 2021년 3월부터 2022년 5월까지 두 자릿수 증가세를 보이다가 6월부터 증가세가 둔화되고 있다. 다만, 최근 러시아·우크라이나 전쟁 등으로 국제 유가가 큰 폭으로 상승하면서 에너지 수입이 증가함에

2022년 10월 26일 접수; 2022년 12월 17일 수정; 2022년 12월 20일 게재확정

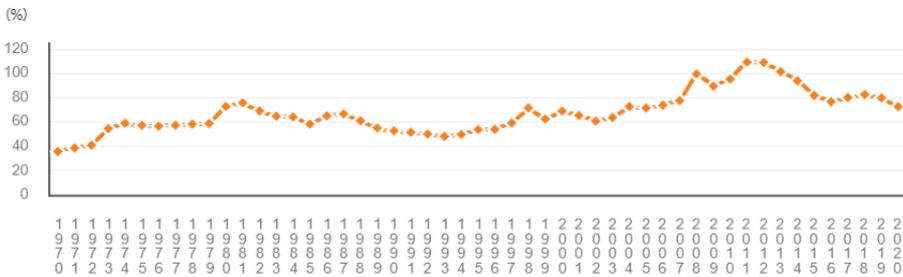
* 이 논문은 장나원의 성균관대학교 2022년 퀀트응용경제학과 연구논문을 수정·보완한 것임.

** 제1저자, 대한무역투자진흥공사(KOTRA), 사원, E-mail : jangnw@kotra.or.kr

*** 교신저자, 성균관대학교, 경제학과/퀀트응용경제학과 교수, 주소: 03063 서울시 종로구 성균관로 25-2, Tel: 02) 760-0428, E-mail: heejoonhan@skku.edu

따라 무역수지 적자가 발생하고 있다. 우리나라는 <Figure 1>과 같이 GDP 대비 수출입비율이 72.9%(20년)로 타 국가 대비 해외 의존도가 상대적으로 높은 경제구조를 보유하고 있다. 또한, World Bank에 따르면 2020년 수출의존도는 독일(43.4%), 멕시코(40.1%)에 이어 한국(36.4%) 순으로 높으며, 수입의존도는 멕시코(38.0%), 독일(37.7%), 그리고 한국(32.8%)으로 나타났다.

<Figure 1> Ratio of Exports and Imports to GDP in South Korea



Source: STATISTICS KOREA_K-indicator_https://www.index.go.kr/unify/idx-info.do?idxCd=4207.

상대적으로 작은 국내 시장을 가진 한국은 무역을 통해 지속적인 경제성장을 이루어 왔다. 하지만 수출입 의존도가 높다는 것은 경제가 대외환경 변화에 민감하다는 것을 의미하기도 한다. 특히, 코로나19, 우크라이나 전쟁 장기화 등 순수출에 부정적인 영향을 주는 국내외 경제 환경 변화에 따라 수출입변화를 예측하는 것은 정부나 기업들의 전략 수립에 중요한 과제일 것이다.

수출입이 한국 경제에 미치는 영향이 크기 때문에 수출액 및 수입액을 예측하거나 중요 요인을 탐색하는 선행연구들이 활발히 진행되었고, 관계기관에서도 매년 수출입 전망을 제시하고 있다. 2022년도 수출을 예측하는 연구기관들의 전망에 따르면 워드 코로나 전환에 따른 교역 활성화에도 불구하고, 공급망 병목 현상, 대외여건 악화 등의 이유로 수출이 둔화할 것이라는 전망이 우세한 상황이다. 한국경제연구원에서 실시한 '2022년 대기업 수출 전망조사'에 따르면 기업들은 원자재 가격 상승, 원·달러 환율 변동성 확대 등을 수출환경 위험 요인으로 응답했다. 또한, KDI국제정책대학원의 경제전망 보고서(정규철 외, 2022)에 따르면 주요국의 수출증가세 축소와 반도체 부품 수급 차질 등 품목별 감소 요인들을 수출증가율 둔화 요인으로 언급했다.

본 연구는 수출입에 영향을 주는 국내외 거시경제/금융 변수들을 폭넓게 고려하고 다양한 머신러닝 기법들을 활용하여 우리나라 수출입 증가율을 예측하고 모형들의 예측력을 평가하였다. 본 연구는 총 13개의 시계열 모형 및 머신러닝 모형들의 예측력을

비교/평가하였다. 수출입과 관련된 52개 국내의 거시경제/금융 변수와 이 52개 변수에서 추출한 4개의 주성분요소(principal component factor)들의 1차에서 4차까지의 시차변수들과 종속변수인 수출 또는 수입 증가율의 1차에서 4차까지의 시차변수 등을 개별 설명변수로 취급하여, 각 시계열 모형 및 머신러닝 모형에 사용되는 설명변수의 수는 총 228개가 되었다. 구간이동(rolling window) 방법을 활용하여 총 13개 모형의 예측치를 생성하고, 전체 표본 크기는 총 264개월(2000.1월~2021.12월)이며, 구간이동 방법을 적용할 때 추정 기간의 크기(window size)는 144개월(12년)로 설정하고 예측 기간은 총 120개월(10년)로 정하였다. 예측 대상기간은 1개월, 3개월, 6개월, 9개월, 12개월로 설정하였다.

분석 결과에 따르면 수출증가율과 수입증가율 예측에서 모든 예측 대상기간에 걸쳐 가장 낮은 예측오차를 나타낸 모형이 있는 것으로 나타났다. 이는 보루타 알고리즘을 통해 중요 설명변수들의 순위를 도출하고 이들 중 교차검증을 통해 정해진 최적 변수의 수만큼만 사용하여 랜덤 포레스트를 적용하는 모형이다. 이 모형은 Kim and Han(2022)이 소개하고 VIX 지수 예측에서 다단계 앞 예측(multi-step ahead forecasting)에 우수하다는 것을 보인 바 있다. Giacomini and White(2006) 검정과 Hansen et al.(2011)의 Model Confidence Set(MCS) 검정을 시행한 결과, 대부분의 예측 대상기간에서 이 모형의 예측력이 통계적으로 유의하게 우수한 것으로 나타났다. 또한 수출입 금액, 물량 및 물가지수의 과거값들을 제외하고 에너지사용량 변수의 과거값이 수출증가율과 수입증가율 예측에 중요한 역할을 하는 것으로 나타났다.

본 논문의 구성은 다음과 같다. 제Ⅱ장에서는 수출입 예측 및 분석 관련 기존 연구를 서술한다. 제Ⅲ장에서는 본 연구에서 고려한 시계열 모형 및 머신러닝 방법들과 연구에 사용한 프로그램과 코드를 설명한다. 제Ⅳ장에서는 데이터, 예측 절차, 예측력 평가 방법 등을 설명하고, 제Ⅴ장에서는 수출증가율과 수입증가율의 예측 결과를 분석한다. 제Ⅵ장에서 연구의 결론을 제시한다.

Ⅱ. 문헌 연구

기존 관련 문헌 연구를 검토해보면, 우선 수출입 함수를 추정하고 수출입 결정요인을 살펴보는 연구들이 있다. 분석 시기나 대상(국가), 산업·품목 그리고 수출입 함수에 포함된 설명변수가 연구자마다 상이하나, 가장 기본적으로 수출의 수요 및 공급 관점에서 소득이나 가격요인을 이용하여 수출입 함수를 추정한 연구로 Kreinin(1967), Goldstein and Khan(1978), 유정호(1984), 이환호(1995), 이재열·한희준(2001), 변창욱(2009)의 연구가 있다. 제품의 차별성 및 경쟁력을 고려한 연구로 Helkie and Hooper(1988),

Muscatelli et al.(1995), 김치호(2000) 등의 연구가 있다. 이외에도 다국적 기업의 활동을 설명변수로 추가하여 진행한 김치호(2000), 김종성·김별화(2005) 등이 있다.¹⁾

다음으로 수출입 결정요인을 활용하여 수출 중·장기 전망모형을 구축한 선행연구 논문들이 있다. 홍성욱(2017)은 실질환율, 실질GDP, 국제유가, 글로벌 금융위기로 인한 구조변화, FTA 등을 고려한 더미변수를 사용하여 중기전망 모형과 벤치마크 모형을 활용하여 구간확장(expanding window) 방식으로 예측을 진행하였다. 송민기(2022)는 일별 국제유가 및 수출금액 속도치를 기반으로 수출물량 급변동에 대한 예측모형 개발하였다. 임의보행 모형과 ARIMA 모형을 사용하여 구간이동 방식으로 1기 후 예측(one-period-ahead forecast) 수행하였으며, 공식 통계공표일보다 98일 빠르게 위기 징후를 포착했다는 유의미한 결과를 제시하였다. 국가 전체의 수출입을 전망한 모형뿐만 아니라 산업별 중장기 전망을 진행한 연구도 존재한다. 정보통신정책연구원(고동환 외 12인)의 보고서에서는 ICT 생산 및 수출의 시계열 전망을 1997년부터 2020년 10월까지의 월별 데이터를 사용하여 Seasonal ARIMA 모형을 기반으로 2025년 12월까지 전망한 바 있다.

거시경제 변수 및 시계열 모형을 사용한 연구논문 외에도 설문조사 방식을 활용하여 경기를 예측하기도 한다. 대표적으로 한국무역협회(KITA)에서는 수출산업경기전망지수(EBSI)를 분기마다 발표한다. 동 지표는 수출실적이 50만 달러 이상인 무역협회 회원사를 대상으로 온라인 설문을 시행하여 수출산업 경기 동향과 연관된 수출상담, 제조원가, 채산성 등에 대한 기초자료를 조사하여 산출하고 있다. 중소기업부과 중소기업중앙회는 중소기업 총 3,150개사를 대상으로 설문조사를 진행하여 중소기업 경기전망지수를 발표하고 있다. 대한상공회의소에서도 '21년 전국의 2,400여 개의 제조업체를 대상으로 경기전망지수(BSI: Business Survey Index)를 조사하여 전망치를 발표하였다.

예측에 있어 전통적인 시계열 통계기법 또는 설문조사를 활용하는 연구도 있으나, 최근 머신러닝 및 딥러닝을 활용한 연구가 증가하고 있다. 딥러닝 모델은 다양한 변수를 활용하여 전통적인 시계열 분석 방법보다 성능이 우수하다는 평가를 받고 있다. 이동원(2019) 연구에서는 수출상황지수(ET-COIN)를 개발하였다. 세계 경제, 대외 거래, 수출입가격, 국내 산업활동, 환율 및 금융 등 총 123개의 월별 지표를 활용하고, 신축망 정규화(Elastic Net regularization) 회귀모형 추정법을 사용하였다. 수출입 예측뿐만 아니라 다양한 타겟 변수들을 넣고 예측 연구를 진행하기도 한다. 한국콘텐츠학회 김정우(2020) 연구는 LASSO, PCR, KNN 등의 머신러닝 기법을 기반으로

1) 홍성욱 외(2017), 63~72쪽에 보다 자세한 설명이 있음.

산업부가가치율을 예측하였다. 조재영(2021)은 무역보험공사의 수출신용보증 기업 데이터를 활용하여 사고예측을 진행하였다. 총 30개의 변수(재무변수 25개, 비재무변수 5개)를 선택하여 로지스틱 회귀, 랜덤 포레스트, XgBoost, LightGBM 그리고 심층신경망 모형을 사용하여 분석을 진행하였다.

Ⅲ. 예측모형

본 연구에서는 다음과 같은 총 13개의 시계열 모형과 머신러닝 방법들을 고려하였다. 자기회귀(AR model with lag 4), Ridge regression, LASSO(least absolute shrinkage and selection operator), adaptive LASSO, elastic net, adaptive elastic net, target factor, CSR(complete subset regression), 랜덤 포레스트, AdaLasso/RF (adaptive LASSO를 통해 선택된 설명변수들만을 이용하여 랜덤 포레스트 적용), 인공신경망 (Neural Network with two hidden layers and 32 and 16 nodes in each hidden layer), XgBoost, Boruta_selected_RF(보루타 알고리즘을 통해 중요 설명변수들의 순위를 도출하고 이들 중 교차검증을 통해 정해진 최적 변수의 수만큼만 사용하여 랜덤 포레스트를 적용, 아래 설명 참고) 등이다. Medeiros et al.(2019)이 벤치마크(benchmark) 모형 중 하나로 포함한 임의보행(random walk) 모형도 고려해 보았으나 예측력이 매우 낮은 것으로 나타나 이를 생략하였다. 수출입 증가율은 지속성(persistence)이 낮고 정상성이 뚜렷하여서 임의보행의 예측력이 낮게 나오는 것이 당연한 것으로 판단된다.

위 모형들은 머신러닝 방법을 활용한 경제/금융 시계열 예측에서 널리 사용되는 방법이며, 각 모형에 대한 설명은 Medeiros et al.(2019)와 Kim and Han(2022)을 참고할 수 있다. Boruta_selected_RF 모형은 Kim and Han(2022)이 제안하였고 본고에서도 동일한 방법으로 적용되었다. Kim and Han(2022)은 금융 변수 위주의 298개 일별(daily) 변수들을 사용하여 미국의 일별 VIX 지수를 예측하였고, 본고는 한국 수출입과 관련된 월별(monthly) 52개의 거시경제/금융 변수들을 사용하여 한국의 월별 수출입증가율을 예측한다는 차이점이 있다. Boruta_selected_RF 모형은 보루타 알고리즘과 교차검증 그리고 랜덤 포레스트를 통해 예측치를 도출하는 방법이다. 이를 간단히 설명하면, i) 보루타 알고리즘을 통해 중요한 변수로 구분된 변수들을 중요도가 높은 순서에 따라 정렬하고 ii) 교차검증을 통해 랜덤 포레스트에 사용할 최적 변수의 수를 결정하고 iii) 중요도가 높은 변수 중 결정된 최적의 수만큼만 랜덤 포레스트에 적용하여 예측치를 생성하는 것이다.

Kursa and Rudnicki(2010)가 소개한 보루타 알고리즘은 랜덤 포레스트를 기반으로

중요 변수들을 안정적으로 선택하는 기법으로, 각 변수에 대해 순열 중요도(permutation importance measure)를 그 표준편차로 나누어 계산하는 Z-score²⁾를 변수 선택의 척도로 사용한다. Boruta Algorithm은 다음과 같은 단계로 실행된다.

1. 모든 변수의 복사본을 추가하여 데이터 세트(dataset)를 확장한다.
2. 추가된 변수들을 섞어서(shuffle) 종속변수와의 상관관계를 제거한다(이를 shadow attributes라 칭함).
3. 확장된 데이터 세트를 이용하여 랜덤 포레스트를 실행하고 각 변수들의 Z-score를 수집한다.
4. shadow attributes의 Z-score 중 최댓값(maximum Z-score among shadow attributes, 이하 'MZSA')을 구한 후, MZSA보다 Z-score가 큰 모든 변수에 히트를 할당(assign a hit)한다.
5. 4에 해당하지 않는 변수의 경우, MSZA에 대한 양측 동등성 검정(two-sided equality test)을 수행한다.
6. MZSA보다 Z-score가 유의하게 작은 변수를 중요하지 않은 변수로 간주하고 데이터 세트에서 모두 제거한다.
7. MZSA보다 Z-score가 유의하게 큰 변수를 중요한 변수로 간주한다.
8. 모든 shadow attributes를 제거한다.
9. 모든 변수에 중요도가 할당되거나 또는 랜덤 포레스트의 최대 실행 횟수에 도달할 때까지 위 절차를 반복한다.

데이터 세트에 포함된 변수의 수가 많으면 보루타 알고리즘을 실행할 때 모든 변수들이 중요하거나 중요하지 않은 두 종류로 분명히 구분되기보다는 중요한 변수(confirmed important), 중요하지 않은 변수(confirmed unimportant), 결정하기 힘든 변수(tentative) 등 세 가지로 구분되는 경향이 있다. 보루타 알고리즘의 R 패키지는 랜덤 포레스트 최대 실행 횟수의 기본값(default)을 100으로 정하고 있는데, 이 경우 랜덤 포레스트를 100번 실행하여 각 변수의 Z-score의 평균, 중위값 등을 구할 수 있다.

Kim and Han(2022)은 보루타 알고리즘이 중요한 변수로 구분한 변수들의 Z-score 평균을 내림차순으로 정렬하고 이를 변수 중요도 순위로 취급하였다. 그리고 다음과 같은 교차검증을 통해 최적 변수의 수를 선택하였다. 우선 변수 중요도가 가장 높은

2) Out-of-bag 샘플은 랜덤 포레스트 실행시 부트스트랩(bootstrap) 샘플링에서 제외된 샘플을 의미한다. 랜덤 포레스트에서 각 변수의 중요도는 out-of-bag 샘플에서 계산한 OOB MSE(out-of-bag mean squared error)와 해당 변수를 무작위 치환(random permutation)하고 계산한 OOB MSE의 차이를 이용하여 측정하고, 각 변수의 Z-score는 이 차이의 평균(Kim and Han(2022, equation (5) 참조)을 표준편차로 나누어 계산한다.

변수와 두 번째로 높은 변수들만을 이용하여 랜덤 포레스트를 적용하고, in-sample OOB(out-of-bag) MSE(mean squared error)를 계산한다. 변수 중요도 순위에 따라 그 다음 순위의 변수를 하나씩 추가하면서 각각의 경우 in-sample OOB MSE를 계산한다. 이때 보루타 알고리즘이 중요하지 않은 변수 또는 결정하기 힘든 변수로 구분한 변수들은 고려하지 않는다. 이 중 MSE가 가장 낮은 변수의 수를 최적 변수의 수로 정한다(<Figure 3>과 <Figure 4> 참고). 최종적으로 변수 중요도가 높은 변수들을 최적 변수의 수까지만 사용하여 랜덤 포레스트에 적용하고 예측치를 도출하는데, 이때 선택되지 않은 나머지 설명변수들은 모두 제거하고 랜덤 포레스트를 실행한다.

본 연구에서 실행한 모든 예측 및 예측 평가 등은 R 프로그램을 이용하였다. 보루타 알고리즘을 포함하여 모두 공개된 패키지(package)들을 사용하였고, 특히 Medeiros et al. (2019)의 코드(<https://github.com/gabrielrvsc/>)와 Kim and Han(2022)의 코드(<https://github.com/bryank2242/Forecasting-VIX>)를 주로 참고하였다. 논문에서 사용한 머신러닝 방법들의 예측에 채택한 하이퍼 파라미터(hyper parameter)들은 모두 Medeiros et al.(2019)와 같은 방법으로 설정하였고³⁾, 보루타 알고리즘과 교차검증을 통해 최적 변수의 수를 정하는 방법은 Kim and Han(2022)과 동일하게 적용하였다. 보루타 알고리즘을 실시할 때 랜덤 포레스트 최대 실행 횟수는 100으로 정하였다.

IV. 데이터 및 예측 절차

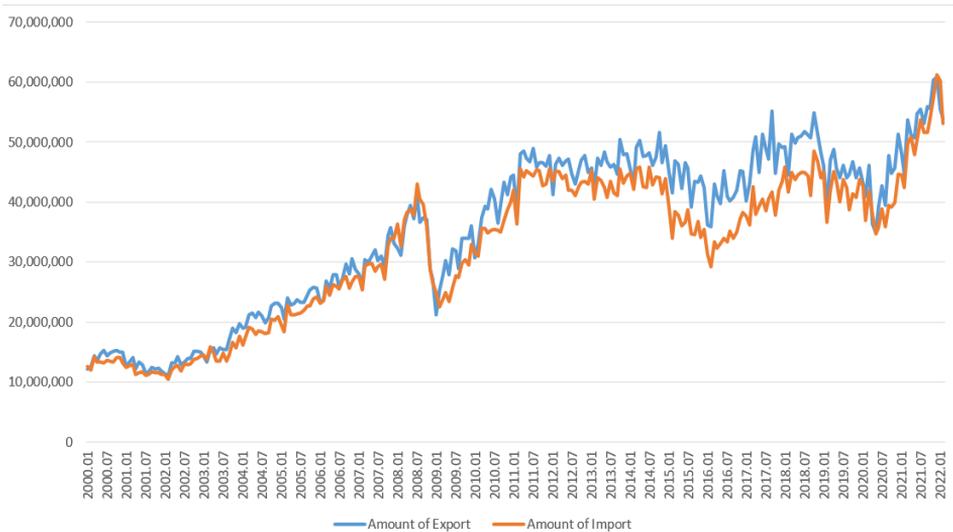
1. 데이터 개요

본 연구에서는 2000년 1월부터 2021년 12월까지의 월별 수출금액 및 수입금액의 증가율(수출입금액의 로그차분)을 종속변수(dependent variable 또는 target variable)로 사용하였다. 한국의 수출입은 <Figure 2>와 같이 2000년 이래로 증가추세를 보이고 있으며 글로벌 금융위기(2009), 유럽재정위기(2011), 저유가(2015, 2016), 코로나(2020)의 원인으로 수출입 금액이 크게 하락하는 양상을 보인다.

수출입 증가율을 예측하기 위하여 52개의 거시경제/금융 변수를 활용하였다. 한국은행 수출상황지수(ET-COIN) 개발에 사용된 변수 구분을 참조하여 <Table 1>과 같이 대외거래, 국제상품가격, 주가, 환율, 국내경기 그리고 세계경기를 나타내는 변수들을 포함하였다.

3) 예를 들어 Medeiros et al.(2019)은 계산 시간을 줄이기 위해 교차검증 대신 BIC를 기준으로 LASSO 모형의 하이퍼 파라미터 λ 를 결정하였다. 본 논문의 데이터를 이용할 때, 교차검증을 통해 λ 를 결정할 때도 비슷한 예측치가 생성되는 것으로 확인하였다.

〈Figure 2〉 Exports, Imports, and Their Growth Rates in South Korea



〈Table 1〉 List of Variables Used for Exports and Imports Forecasting

	Category	Variables	Variable(sum)	Unit	Frequency	Institution
-	-	Amount of Export/Import	EX/IM_TOT	1,000\$	month	Customs Service
1	Trade	Export Quantity Index	EX_INDEX_Q	Index	month	ECOS
2		Import Quantity Index	IM_INDEX_Q	Index	month	ECOS
3		Export Price Index	EX_INDEX_P	Index	month	ECOS
4		Import Price Index	IM_INDEX_P	Index	month	ECOS
5		Current account balance	CAB	1million\$	month	ECOS
6	International price index	WTI oil price	OIL_W	US\$/bbl	month	ECOS
7		Dubai oil price	OIL_D	US\$/bbl	month	ECOS
8		Brent oil price	OIL_B	US\$/bbl	month	ECOS
9		Soybean Futures	BEAN	¢/bushel	month	ECOS
10		Gold price	GOLD	\$/OZ	month	ECOS
11		Nickel price	NIKEL	\$/MT	month	ECOS
12		Zinc price	ZINC	\$/MT	month	ECOS
13		Corn Futures	CORN	¢/bushel	month	ECOS
14		Wheat Futures	WHEAT	¢/bushel	month	ECOS
15		Copper price	COPPER	\$/MT	month	ECOS
16		Aluminium price	ALUMIN	\$/MT	month	ECOS

	Category	Variables	Variable(sum)	Unit	Frequency	Institution
17		Cotton Futures	COTTON	¢/£	month	ECOS
18	PPI	Producer Price Index(US)	PPI_US	Index	month	FRED
19		Producer Price Index(KR)	PPI_KR	Index	month	ECOS
20		Producer Price Index(JP)	PPI_JP	Index	month	FRED
21	Stock Price	Korea(KOSPI)	KOSPI	Index	month	ECOS
22		USA(Dow Jones)	DOW	Index	month	ECOS
23		USA(NASDAQ)	NASDAQ	Index	month	ECOS
24		Euro STOXX	STOXX	Index	month	ECOS
25		Germany(DAX)	DAX	Index	month	ECOS
26		Japan(NIKKEI)	NIKKEI	Index	month	ECOS
27		Sigapore(STI)	STI	Index	month	ECOS
28		Hong Kong(HSI)	HANG	Index	month	ECOS
29		China(SHCOMP)	SHCOMP	Index	month	ECOS
30	Exchange rate	KRW/USD	KRWUSD	Average	month	ECOS
31		KRW/JPY	KRWJPY	Average	month	ECOS
32		KRW/EUR	KRWEUR	Average	month	ECOS
33		JPY/USD	JPYUSD	Average	month	ECOS
34		USD/EUR	USDEUR	Average	month	ECOS
35	Finance	CBOE Volatility Index	VIX	Index	month	FRED
36	Economic cycle (KR)	Industrial Product Index	IIP	Index	month	ECOS
37		Operation ratio Index	OPER_IDX	Index	month	ECOS
38		Shipment Index	SHIP_IDX	Index	month	ECOS
39		Inventory Index	STOCK_IDX	Index	month	ECOS
40		Equipment Investment Index	PLANTIV_IDX	Index	month	ECOS
41		Composite Leading Indicator	ECO_B	Index	month	ECOS
42		Coincident Composite Index	ECO_R	Index	month	ECOS
43		Lagging composite Index	ECO_A	Index	month	ECOS
44		Energy Consumption	ENERGY	1,000TOE	month	ECOS
45	Employed	EMPLOY	number	month	ECOS	
46	Economic cycle (World)	Industrial Product Index(Japan)	IIP_JP	Index	month	ECOS
47		Industrial Product Index(Russia)	IIP_RU	Index	month	ECOS
48		Industrial Product Index(US)	IIP_US	Index	month	ECOS

	Category	Variables	Variable(sum)	Unit	Frequency	Institution
49		Composite Leading Indicator(OECD)	CLI	Index	month	OECD
50		Spread(US)	SPRATE_US	%	month	FRED
51		Effective Federal Funds Rate	EFFR_US	%	month	FRED
52		Amount of Imports(US)	IM_US	1million\$	month	KITA

Note: ECOS: Economic Statistics System, Bank of Korea, FRED: Federal Reserve Economic Data, St. Louis Fed, KITA: Korea International Trade Association.

수출입금액 변동을 가격요인과 물량요인으로 나누어 파악하기 위한 변수로 수출입물가지수는 수출의 채산성 변동 및 수입원가 부담 파악 등에 활용할 수 있다. 국제원자재의 가격변동은 이를 중간재로 투입하는 모든 상품 가격에 변동을 초래하기 때문에 수출입 예측에 필요한 변수로 설정하였다. 국내외 주식시장의 변동은 환율에 영향을 줄 뿐만 아니라 수출입 관련 제조업의 경기를 통해 수출입에 대한 정보를 제공하기 때문에 주요국의 주가지수들을 변수에 포함하였다. 통상 환율 상승(원화가치 하락)은 수출상품의 외화표기 가격을 하락시켜 수출을 증가시키는 요인으로 작용하나, 최근 한국 수출상품의 고도화에 따라 환율 영향이 과거보다 축소된 것으로 평가하는 연구 결과도 존재한다. 이에 원/달러 환율 외에 주요국 환율을 변수로 고려하였다. 한국의 경제성장은 수입의 변동에 미치는 영향이 크고, 실질소득의 변화는 수출입 변동에 인과관계가 있어 예측을 위한 변수로 활용하였다. 세계 경기 변동은 세계 수입 수요에 영향을 미쳐 우리나라 수출을 좌우하는 절대적 요인으로 세계경기와 한국의 수출은 대체로 유사한 방향성을 보이므로 주요국의 산업생산지수 등을 반영하였다.

설명변수로 사용된 시계열 데이터가 단위근을 가지고 있으면 이를 정상시계열로 변환하여 머신러닝 기법들을 적용하였는데, 단위근 검정 방법으로 PP(Phillips-Perron) 검정과 KPSS(Kwiatkowski-Phillips-Schmidt-Shin) 검정을 활용하였다. PP 검정의 귀무가설은 시계열이 단위근을 가지는 것이며, 대립가설은 정상시계열(stationary)인 검정 방법이다. 반면, KPSS 검정은 일반적인 단위근 검정법과 반대로 귀무가설이 '시계열이 정상시계열'로 설정되는 검정 방법이다. 단위근 검정 결과 <Table 2>와 같이 VIX 지수 외 다른 변수들이 모두 단위근을 가지는 것으로 확인되었으며 정상시계열로 만들기 위한 데이터 변환을 진행하였다. FRED-MD의 데이터 변환 방법 및 선행연구논문을 참고하여 지수형 변수는 로그차분, 가격변수들은 로그이차차분, 이자율 변수는 차분으로 변환하여 단위근을 제거하였다.

〈Table 2〉 Unit Root Test Results

	Variable	PP_Test (1)	PP_Test (2)	KPSS_Test (1)	KPSS_Test (2)	Transformation
-	EX_TOT	-1.0089	-3.8140	1.5778	0.2953	$\Delta \log(x_t)$
1	EX_INDEX_Q	-0.5588	-8.9949	1.7402	0.2292	$\Delta \log(x_t)$
2	IM_INDEX_Q	-0.5332	-12.1870	1.7290	0.0514	$\Delta \log(x_t)$
3	EX_INDEX_P	-2.5889	-2.5369	0.8810	0.0976	$\Delta \log(x_t)$
4	IM_INDEX_P	-1.4105	-1.5948	0.7633	0.3054	$\Delta \log(x_t)$
5	CAB	-6.0131	-9.2046	1.3727	0.1318	Δx_t
6	OIL_W	-2.4869	-2.5226	0.4686	0.3306	$\Delta^2 \log(x_t)$
7	OIL_D	-2.2579	-2.3079	0.5529	0.3161	$\Delta^2 \log(x_t)$
8	OIL_B	-2.2587	-2.2859	0.5167	0.3173	$\Delta^2 \log(x_t)$
9	BEAN	-2.1467	-2.6619	0.8752	0.2690	$\Delta^2 \log(x_t)$
10	GOLD	-0.6057	-1.7871	1.4601	0.2024	$\Delta^2 \log(x_t)$
11	NIKEL	-2.3731	-2.3698	0.2154	0.2204	$\Delta^2 \log(x_t)$
12	ZINC	-1.6261	-2.3844	0.7815	0.1013	$\Delta^2 \log(x_t)$
13	CORN	-2.1309	-2.5379	0.6611	0.2313	$\Delta^2 \log(x_t)$
14	WHEAT	-2.2023	-2.5743	0.6381	0.2542	$\Delta^2 \log(x_t)$
15	COPPER	-1.6444	-2.2598	0.9316	0.2749	$\Delta^2 \log(x_t)$
16	ALUMIN	-2.2539	-2.4152	0.2376	0.1861	$\Delta^2 \log(x_t)$
17	COTTON	-2.7333	-3.1616	0.5133	0.1348	$\Delta^2 \log(x_t)$
18	PPI_US	-0.5483	-1.9242	1.4996	0.2818	$\Delta^2 \log(x_t)$
19	PPI_KR	-0.4329	-1.4827	1.4999	0.3072	$\Delta^2 \log(x_t)$
20	PPI_JP	-2.2755	-2.7863	1.5349	0.1269	$\Delta^2 \log(x_t)$
21	KOSPI	-0.5644	-3.2801	1.5913	0.1314	$\Delta \log(x_t)$
22	DOW	2.3503	-0.3696	1.4536	0.3954	$\Delta \log(x_t)$
23	NASDAQ	5.1443	1.4722	1.2869	0.3720	$\Delta \log(x_t)$
24	STOXX	-2.3711	-2.1613	0.1835	0.1601	$\Delta \log(x_t)$
25	DAX	0.3116	-2.5288	1.5015	0.2509	$\Delta \log(x_t)$
26	NIKKEI	-0.2919	-2.3411	1.0106	0.2942	$\Delta \log(x_t)$
27	STI	-1.8616	-2.8317	1.1638	0.2010	$\Delta \log(x_t)$
28	HANG	-1.8576	-3.4348	1.4061	0.0816	$\Delta \log(x_t)$
29	SHCOMP	-2.4831	-3.1076	0.7977	0.0674	$\Delta \log(x_t)$
30	KRWUSD	-2.7432	-2.7338	0.1151	0.1142	$\Delta \log(x_t)$
31	KRWJPY	-2.0639	-2.0646	0.1757	0.1488	$\Delta \log(x_t)$

	Variable	PP_Test (1)	PP_Test (2)	KPSS_Test (1)	KPSS_Test (2)	Transformation
32	KRWEUR	-2.0226	-1.9239	0.3106	0.2645	$\Delta \log(x_t)$
33	JPYUSD	-1.7867	-1.7777	0.2995	0.2361	$\Delta \log(x_t)$
34	USDEUR	-1.9311	-1.7947	0.3801	0.3549	$\Delta \log(x_t)$
35	VIX	-5.2206	-5.3108	0.1897	0.0755	-
36	IIP	-0.8576	-2.5247	1.6788	0.3480	$\Delta \log(x_t)$
37	OPER_IDX	-4.2475	-5.1594	0.9025	0.1957	$\Delta \log(x_t)$
38	SHIP_IDX	-1.3436	-2.0620	1.6295	0.3831	$\Delta \log(x_t)$
39	STOCK_IDX	0.0472	-3.3737	1.7463	0.1297	$\Delta \log(x_t)$
40	PLANTIV_IDX	-0.6910	-5.3465	1.6658	0.0488	$\Delta \log(x_t)$
41	ECO_B	2.0942	-1.5173	1.7350	0.3860	$\Delta \log(x_t)$
42	ECO_R	-0.5839	-3.1852	1.7573	0.0629	$\Delta \log(x_t)$
43	ECO_A	1.8341	0.5820	1.7552	0.3868	$\Delta \log(x_t)$
44	ENERGY	-3.1035	-7.7731	1.6936	0.1947	$\Delta \log(x_t)$
45	EMPLOY	-1.1832	-3.2594	1.7526	0.1206	$\Delta \log(x_t)$
46	IIP_JP	-2.9571	-3.2646	0.4238	0.0921	$\Delta \log(x_t)$
47	IIP_RU	-0.2215	-2.6940	1.6758	0.1145	$\Delta \log(x_t)$
48	IIP_US	-2.1547	-2.6997	0.7654	0.0637	$\Delta \log(x_t)$
49	CLI	-3.3327	-3.3354	0.1280	0.0611	$\Delta \log(x_t)$
50	SPRATE_US	-1.9671	-2.0476	0.1760	0.1579	Δx_t
51	EFFR_US	-1.9870	-1.8794	0.7186	0.1255	Δx_t
52	IM_US	-1.5032	-4.7074	1.5536	0.1590	$\Delta \log(x_t)$

Note: PP test(1) includes only an intercept and critical values are 10%(-2.57), 5%(-2.86), 1%(-3.44), respectively. PP test(2) includes both an intercept and a linear trend and critical values are 10%(-3.13), 5%(-3.42), 1%(-3.97), respectively. KPSS test(1) includes only an intercept and critical values are 10%(0.347), 5%(0.463), 1%(0.739). KPSS test(2) includes both an intercept and a linear trend and critical values are 10%(0.119), 5%(0.146), 1%(0.216). The last column in the table shows how to transform each series into a stationary process.

다양한 머신러닝 모형들을 적용할 때 사용하는 설명변수로는 위에서 설명한 52개 거시경제/금융 변수와 함께 이 52개 거시경제/금융 변수에서 추출한 4개의 주성분요소(principal component factor)들도 포함하였다. 그리고 모든 설명변수의 1차에서 4차까지의 시차변수(from 1 lag up to 4 lags of all variables)들을 개별 설명변수로 취급하고 종속변수의 1차에서 4차까지의 시차변수도 설명변수로 추가하였다. 이는 Medeiros et al.(2019)가 채택한 방식을 그대로 따른 것이다. 따라서 각 머신러닝 모형에

사용되는 설명변수의 수는 총 228개가 되었다.

2. 예측 절차 및 예측력 평가 방법

본 연구는 구간이동 방법을 활용하여 총 13개 모형의 예측치를 생성하고, 예측오차로 RMSE(root mean squared error)와 MAE(mean absolute error)를 비교하였다. 전체 표본 크기는 총 264개월(2000.1월~2021.12월)이며, 구간이동 방법을 적용할 때 추정 기간의 크기(window size)는 144개월(12년)로 설정하고 예측 기간은 총 120개월(10년)로 정하였다. 예측 대상기간은 3개월 단위로 1개월, 3개월, 6개월, 9개월, 12개월로 설정하였다. 예측 대상기간이 3개월에서 12개월까지인 다단계 예측(multi-step ahead forecasting)에서는 Medeiros et al.(2019)와 Kim and Han(2022)과 같이 직접 예측(direct forecasting) 방식을 이용하였다. 이는 설명변수들의 예측치를 별도로 구할 필요가 없다는 면에서 편리할 뿐만 아니라 예측 관련 기존 연구에서 모형오지정(model misspecification)에 강건한(robust)하다는 장점이 있는 것으로 알려져 있기 때문이다.

각 모형의 예측치와 예측오차를 구한 다음 예측오차 간의 차이가 통계적으로 유의한지를 검정하기 위해 Giacomini and White(2006)와 Hansen et al.(2011)의 검정방법을 적용하였다. 예측오차가 가장 낮은 모형과 나머지 모형 각각에 대해 Giacomini and White(2006)의 무조건부 검정(unconditional test)을 적용하는데, 이 경우 귀무가설은 두 모형의 예측력이 동일하다는 것이다. Hansen et al.(2011)의 Model Confidence Set(MCS) 방법은 모든 모형을 한꺼번에 테스트하는 것으로, 테스트가 진행되면서 예측력이 열등한 모형들이 하나씩 제거되어 나가고 예측력이 우수한 모형들만 마지막까지 제거되지 않고 MCS에 남아 있게 된다. Medeiros et al.(2019)와 같이 유의수준(confidence level)은 50%로 설정하였다.

V. 예측 및 분석 결과

1. 수출증가율 예측 결과

<Table 3>과 <Table 4>는 각 모형의 예측 대상기간별 수출증가율 예측치에 대한 예측오차인 RMSE와 MAE를 제시한다. 모형별 RMSE와 MAE를 비교하면, 1개월에서 12개월까지의 모든 예측 대상기간에서 Boruta_selected_RF 모형의 예측오차가 가장 낮은 것으로 나타나, 보루타 알고리즘을 통해 중요 설명변수들의 순위를 도출하고 이들 중 교차검증을 통해 정해진 최적 변수의 수만큼만 사용하여 랜덤 포레스트를 적용하는 방법의 예측력이 가장 좋은 것을 알 수 있다. 두 번째로 낮은 예측오차를 지닌 모형은 예측 대상기간 또는 예측오차별로 다른 것으로 나타났는데, RMSE

기준으로는 XgBoost, Random Forest, AR, AdaLasso-RF 등으로 그리고 MAE 기준으로는 XgBoost, AR, CSR 등으로 확인되었다.

〈Table 3〉 RMSEs for Exports Forecasting

Model	1Month	3Month	6Month	9Month	12Month
AR	0.0739	0.0797	0.0789	0.0744	0.0789
Ridge	0.0829	0.0831	0.0830	0.0831	0.0831
LASSO	0.0738	0.0814	0.0922	0.0754	0.0799
Adaptive Lasso	0.0723	0.0823	0.0817	0.0738	0.0786
Elastic Net	0.0754	0.0831	0.0827	0.0765	0.0799
Adaptive Elasticnet	0.0732	0.0832	0.0825	0.0745	0.0785
Target Factor	0.0775	0.0799	0.0817	0.0813	0.0768
CSR	0.0671	0.0803	0.0808	0.0734	0.0743
Radom Forest	0.0700	0.0739	0.0813	0.0734	0.0738
AdaLasso / RF	0.0713	0.0812	0.0805	0.0723	0.0817
Neural Network	0.0794	0.0871	0.0979	0.0905	0.0930
XgBoost	0.0667	0.0747	0.0801	0.0801	0.0743
Boruta_selected_RF	0.0612	0.0677	0.0706	0.0657	0.0683

Note: Numbers in bold represents the minimum of forecast errors for each forecast horizon.

〈Table 4〉 MAEs for Exports Forecasting

Model	1Month	3Month	6Month	9Month	12Month
AR	0.0587	0.0630	0.0625	0.0560	0.0624
Ridge	0.0650	0.0652	0.0651	0.0651	0.0650
LASSO	0.0581	0.0639	0.0705	0.0585	0.0620
Adaptive Lasso	0.0571	0.0650	0.0639	0.0565	0.0612
Elastic Net	0.0590	0.0651	0.0647	0.0591	0.0619
Adaptive Elasticnet	0.0577	0.0651	0.0644	0.0568	0.0608
Target Factor	0.0622	0.0627	0.0626	0.0640	0.0616
CSR	0.0517	0.0633	0.0632	0.0550	0.0570
Radom Forest	0.0547	0.0572	0.0632	0.0563	0.0571
AdaLasso / RF	0.0565	0.0650	0.0630	0.0555	0.0633
Neural Network	0.0605	0.0685	0.0783	0.0692	0.0708
XgBoost	0.0508	0.0568	0.0677	0.0601	0.0564
Boruta_selected_RF	0.0473	0.0526	0.0565	0.0507	0.0525

Note: Numbers in bold represents the minimum of forecast errors for each forecast horizon.

다음 단계로는 Boruta_selected_RF 모형의 예측력이 다른 모형들에 비해 통계적으로 유의하게 우월한지를 Giacomini-White 검정과 MCS 검정을 이용하여 검정하였다. <Table 5>는 Boruta_selected_RF 모형과 기타 모형 간의 Giacomini-White 검정 결과(예측오차 절대값 기준)를 제시한다. <Table 5>에서 통계치들이 모두 음(-)의 값을 가지는 것은 Boruta_selected_RF 모형의 예측오차가 테스트하는 상대 모형보다 낮기 때문이다. Giacomini-White 검정의 귀무가설은 두 모형의 예측력이 같다는 것인데, 대부분의 예측 대상기간에서 귀무가설을 기각하는 것으로 나타났다. 이는 Boruta_selected_RF 모형이 다른 모형들에 비해 통계적으로 유의하게 예측력이 우수하다는 것을 나타낸다. 다만 예측 대상기간이 1개월, 3개월, 12개월이 경우 XgBoost 모형에 대해서 그리고 예측 대상기간이 9개월인 경우 CSR과 AdaLasso-RF 모형에 대해서는 귀무가설을 기각할 수 없는 것으로 나타났다. <Table 6>은 MCS 검정 결과를 제시하는데, 이 검정 방법에 따르면 예측력이 우수한 모형들이 MCS에 포함된다. <Table 6>에 따르면 모든 예측 대상기간에서 Boruta_selected_RF 모형만 MCS에 포함되는 것으로 나타나 Boruta_selected_RF 모형이 예측력이 우수한 모형임을 보인다.

<Table 5> Giacomini-White Test Results for Exports Forecasting

Model	1Month	3Month	6Month	9Month	12Month
AR	-4.29***	-4.26***	-2.17**	-1.36	-3.62***
Ridge	-5.50***	-4.14***	-3.26***	-6.13***	-5.34***
Lasso	-4.55***	-3.88***	-3.22***	-5.01***	-3.62***
Adaptive Lasso	-3.69***	-3.76***	-3.22***	-3.20***	-2.33**
Elastic Net	-4.87***	-4.04***	-3.22***	-5.45***	-3.64***
Ada-Elasticnet	-4.07***	-4.03***	-3.22***	-4.44***	-2.50**
Target Factor	-4.38***	-2.42**	-2.60***	-5.03***	-4.18***
CSR	-1.87*	-4.82***	-2.14**	-1.22	-1.65*
Radom Forest	-4.17***	-1.87*	-2.38**	-2.45**	-2.09**
AdaLasso - RF	-3.96***	-4.26***	-2.39**	-1.47	-2.74***
Neural Network	-3.14***	-3.83***	-4.04***	-4.33***	-3.86***
XgBoost	-1.59	-1.56	-2.8***	-2.49**	-1.19

Note: The table provides unconditional Giacomini-White test statistics between the Boruta_selected_RF model and each competing model for absolute errors. A negative statistic implies that the Boruta_selected_RF model has a lower forecast loss than a competing model. *, **, *** indicate rejection of equal predictive ability at the 10%, 5%, 1% significance level, respectively.

〈Table 6〉 MCS(Model Confidence Set) Test Results for Exports Forecasting

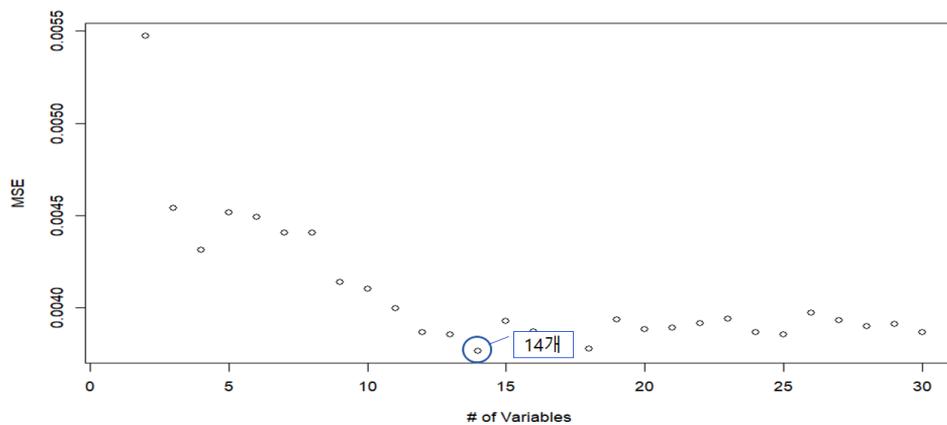
Model	1Month	3Month	6Month	9Month	12Month
AR					
Ridge					
LASSO					
Adaptive Lasso					
Elastic Net					
Adaptive Elasticnet					
Target Factor					
CSR					
Radom Forest					
AdaLasso / RF					
Neural Network					
XgBoost					
Boruta_selected_RF	O	O	O	O	O

Note: The table provides model confidence set test results with significance level 50%. If a model is included in model confidence set, it means that the model exhibits best predictive ability.

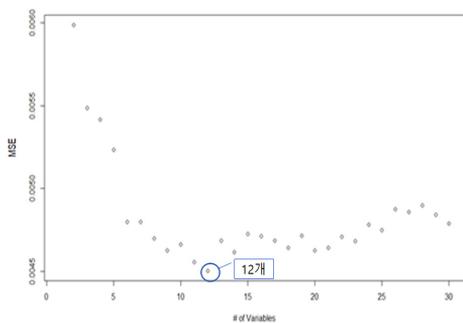
〈Figure 3〉은 Boruta_selected_RF 모형에서 각 예측 대상기간별로 교차검증을 통해 최적 변수의 수를 산출하는 과정을 보여준다. 3장에서 설명한 바와 같이 보루타 알고리즘을 통해 중요한 변수들로 구분된 변수들을 중요도 순위에 따라 추가해 가면서 MSE를 계산하였고, 예측 대상기간별로 각각 14개, 12개, 16개, 19개, 25개의 변수가 선택되었다. Boruta_selected_RF 모형은 이렇게 선택된 변수들만을 랜덤 포레스트에 사용하는데, 이 변수들이 수출증가율 예측에 중요한 요인임을 의미한다.

〈Table 7〉은 각 예측 대상기간별 선택된 변수들을 제시한다. 수출입 금액, 물량 및 물가지수 증가율의 과거값들이 포함된 것은 시계열의 의존성(dependence)을 고려할 때 당연한 것으로 보이고, 이들을 제외한 나머지 변수들을 살펴볼 때 가장 큰 특징은 ‘에너지사용량’ 변수가 모든 예측 대상기간에 포함된 것이다. 이는 에너지사용량이 무역 및 산업생산에 대한 선행지표의 역할을 한다는 기존 연구 결과에 대응한다. 에너지사용량과 무역 및 GDP의 관련성에 관한 연구에 따르면 각각의 변수는 서로 그레인저 인과(Granger cause)하는 것으로 나타났으며(백문영·김우환(2012), Sanli(2019), Shakeel et al.(2014) 참조), 산업연구원의 박성근(2020)에 따르면 전력 사용량은 생산통계가 발표되기 전 최근 2개월간의 생산 동향에 대한 유용한 정보를 제공한다고 한다.

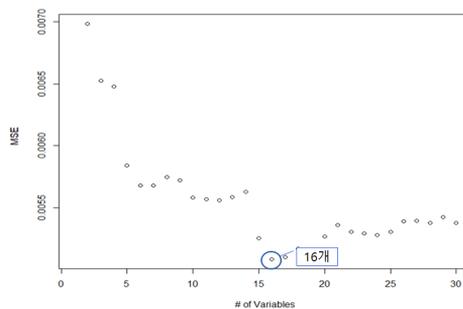
〈Figure 3〉 Number of Variables and In-Sample OOB MSE for Exports Forecasting



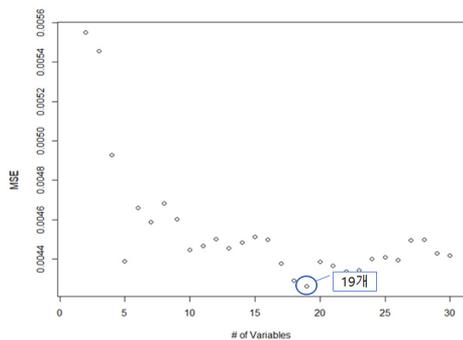
1 month ahead



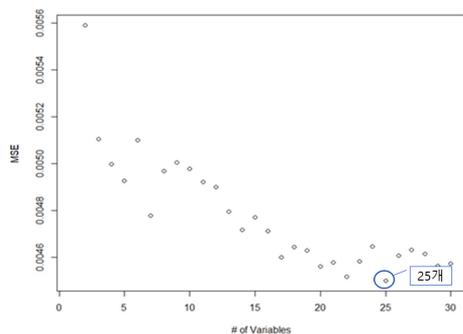
3 month ahead



6 month ahead



9 month ahead



12 month ahead

Note: For each forecasting horizon, the changes in in-sample out-of-bag (OOB) mean squared error (MSE) are plotted as variables that are added to the dataset based on the order of the rankings decided by the Boruta algorithm.

〈Table 7〉 List of Selected Variables by Boruta Algorithm and Cross-Validation for Exports Forecasting

1 Month		3 Month	
X45	Energy Consumption(t-1)	X45	Energy Consumption(t-1)
X1	Amount of Export(KR)(t-1)	X3	Import Quantity Index((t-1)
X151	Energy Consumption(t-3)	X161	Export Quantity Index(t-4)
X109	Import Quantity Index(t-3)	X160	Amount of Export(KR)(t-4)
X3	Import Quantity Index((t-1)	X153	Industrial Product Index(JP)(t-3)
X2	Export Quantity Index(t-1)	X204	Energy Consumption(t-4)
X36	VIX Index(t-1)	X98	Energy Consumption(t-2)
X21	Producer Price Index(JP)(t-1)	X55	Export Quantity Index(t-2)
X55	Export Quantity Index(t-2)	X162	Import Quantity Index((t-4)
X56	Import Quantity Index((t-2)	X106	Amount of Imports(US)(t-2)
X5	Import Price Index(t-1)	X212	Amount of Imports(US)(t-4)
X204	Energy Consumption(t-4)	X165	Current account balance(t-4)
X206	Industrial Product Index(JP)(t-4)	9 Month	
X161	Export Quantity Index(t-4)	X212	Amount of Imports(US)(t-4)
6 Month		X160	Amount of Export(KR)(t-4)
X1	Amount of Export(KR)(t-1)	X161	Export Quantity Index(t-4)
X2	Export Quantity Index(t-1)	X55	Export Quantity Index(t-2)
X3	Import Quantity Index((t-1)	X151	Energy Consumption(t-3)
X204	Energy Consumption(t-4)	X56	Import Quantity Index((t-2)
X53	Amount of Imports(US)(t-1)	X54	Amount of Export(KR)(t-2)
X112	Current account balance(t-3)	X162	Import Quantity Index((t-4)
X6	Current account balance(t-1)	X106	Amount of Imports(US)(t-2)
X161	Export Quantity Index(t-4)	X45	Energy Consumption(t-1)
X28	Sigapore(STD)(t-1)	X165	Current account balance(t-4)
X86	KRW/EUR(t-2)	X98	Energy Consumption(t-2)
X157	Spread(US)(t-3)	X112	Current account balance(t-3)
X38	Operation ratio Index(t-1)	X147	Equipment Investment Index(t-3)
X160	Amount of Export(KR)(t-4)	X3	Import Quantity Index((t-1)
X107	Amount of Export(KR)(t-3)	X59	Current account balance(t-2)
X26	Germany(DAX)(t-1)	X207	Industrial Product Index(RU)(t-4)
X60	WTI oil price(t-2)	X1	Amount of Export(KR)(t-1)
		X172	Zinc price(t-4)

12 Month			
X53	Amount of Imports(US)(t-1)	X106	Amount of Imports(US)(t-2)
X2	Export Quantity Index(t-1)	X174	Wheat Futures(t-4)
X98	Energy Consumption(t-2)	X74	Producer Price Index(JP)(t-2)
X204	Energy Consumption(t-4)	X144	Operation ratio Index(t-3)
X1	Amount of Export(KR)(t-1)	X59	Current account balance(t-2)
X3	Import Quantity Index(t-1)	X108	Export Quantity Index(t-3)
X55	Export Quantity Index(t-2)	X91	Operation ratio Index(t-2)
X56	Import Quantity Index(t-2)	X22	Korea(KOSPI)(t-1)
X126	Producer Price Index(KR)(t-3)	X32	KRW/JPY(t-1)
X151	Energy Consumption(t-3)	X109	Import Quantity Index(t-3)
X112	Current account balance(t-3)	X133	Japan(NIKKEI)(t-3)
X6	Current account balance(t-1)	X159	Amount of Imports(US)(t-3)
X81	Singapore(STI)(t-2)		

Note: The table provides explanatory variables used for the Boruta_selected_RF model. For each forecast horizon, we obtain a ranking of variables that are confirmed to be important by using the Boruta algorithm. Next, a cross validation procedure is conducted to determine the optimal number of variables to be used in a random forest model. This procedure provides selected variables that are used for a random forest model which is the Boruta_selected_RF model.

2. 수입증가율 예측 결과

<Table 8>과 <Table 9>는 각 모형의 예측 대상기간별 수입증가율 예측치에 대한 예측오차인 RMSE와 MAE를 제시한다. 수출증가율 예측 결과와 같이 수입증가율 예측에서도 Boruta_selected_RF 모형이 모든 예측 대상기간에서 가장 작은 RMSE와 MAE를 나타내었다. 두 번째로 낮은 예측오차를 지닌 모형은 예측 대상기간 또는 예측오차별로 다른 것으로 나타났는데, XgBoost, Random Forest, Adaptive Lasso 등으로 확인되었다.

<Table 8> RMSEs for Imports Forecasting

Model	1Month	3Month	6Month	9Month	12Month
AR	0.0634	0.0674	0.0695	0.0642	0.0643
Ridge	0.0703	0.0704	0.0704	0.0705	0.0705
LASSO	0.0642	0.0702	0.0704	0.0640	0.0646
Adaptive Lasso	0.0615	0.0702	0.0704	0.0615	0.0626
Elastic Net	0.0669	0.0705	0.0704	0.0660	0.0655
Adaptive Elasticnet	0.0649	0.0707	0.0704	0.0626	0.0627

Model	1Month	3Month	6Month	9Month	12Month
Target Factor	0.0660	0.0672	0.0691	0.0710	0.0655
CSR	0.0582	0.0663	0.0683	0.0622	0.0636
Radom Forest	0.0572	0.0613	0.0658	0.0611	0.0618
AdaLasso / RF	0.0610	0.0662	0.0698	0.0683	0.0682
Neural Network	0.0638	0.0745	0.0806	0.0766	0.0790
XgBoost	0.0523	0.0592	0.0697	0.0647	0.0637
Boruta_selected_RF	0.0498	0.0566	0.0563	0.0534	0.0542

Note: Numbers in bold represents the minimum of forecast errors for each forecast horizon.

<Table 9> MAEs for Imports Forecasting

Model	1Month	3Month	6Month	9Month	12Month
AR	0.0497	0.0536	0.0556	0.0513	0.0523
Ridge	0.0564	0.0564	0.0565	0.0565	0.0565
LASSO	0.0506	0.0558	0.0565	0.0516	0.0525
Adaptive Lasso	0.0481	0.0558	0.0565	0.0497	0.0507
Elastic Net	0.0534	0.0564	0.0565	0.0535	0.0530
Adaptive Elasticnet	0.0514	0.0566	0.0565	0.0511	0.0509
Target Factor	0.0514	0.0533	0.0557	0.0572	0.0528
CSR	0.0437	0.0526	0.0549	0.0495	0.0531
Radom Forest	0.0450	0.0492	0.0523	0.0492	0.0508
AdaLasso / RF	0.0485	0.0520	0.0541	0.0517	0.0569
Neural Network	0.0506	0.0569	0.0657	0.0608	0.0648
XgBoost	0.0415	0.0486	0.0558	0.0513	0.0521
Boruta_selected_RF	0.0394	0.0452	0.0446	0.0437	0.0430

Note: Numbers in bold represents the minimum of forecast errors for each forecast horizon.

<Table 10>은 Boruta_selected_RF 모형과 기타 모형 간의 Giacomini-White 검정 결과(예측오차 절대값 기준)를 제시한다. Boruta_selected_RF 모형의 예측오차가 가장 낮아서 Giacomini-White 통계치들이 모두 음(-)의 값을 나타낸다. 예측 대상기간이 1개월과 3개월인 경우의 XgBoost 모형과의 검정을 제외한 모든 경우에 예측력이 같다는 귀무가설을 기각하고 있는데, 이는 Boruta_selected_RF 모형이 다른 모형들에 비해 통계적으로 유의하게 예측력이 우수하다는 것을 나타낸다. <Table 11>은 MCS 검정 결과를 제시한다. 예측 대상기간이 12개월을 제외한 모든 경우에는 Boruta_selected_RF 모형만 model confidence set에 포함되는 것으로 나타났으며, 예측 대상기간이 12개월인 경우에는 Boruta_selected_RF과 함께 AR, Adaptive Lasso, Elastic Net, Random Forest, Xgboost 등도 MCS에 포함되었다.

〈Table 10〉 Gioacomini-White Test Results for Imports Forecasting

Model	1Month	3Month	6Month	9Month	12Month
AR	-4.99 ^{***}	-3.64 ^{***}	-3.88 ^{***}	-2.88 ^{***}	-4.32 ^{***}
Ridge	-6.21 ^{***}	-4.62 ^{***}	-5.33 ^{***}	-4.73 ^{***}	-5.67 ^{***}
Lasso	-5.06 ^{***}	-4.38 ^{***}	-5.33 ^{***}	-3.93 ^{***}	-5.55 ^{***}
Adaptive Lasso	-3.70 ^{***}	-4.41 ^{***}	-5.33 ^{***}	-2.68 ^{***}	-3.70 ^{***}
Elastic Net	-5.94 ^{***}	-4.55 ^{***}	-5.33 ^{***}	-4.42 ^{***}	-5.84 ^{***}
Ada-Elasticnet	-5.13 ^{***}	-4.58 ^{***}	-5.33 ^{***}	-3.35 ^{***}	-4.09 ^{***}
Target Factor	-4.93 ^{***}	-2.96 ^{***}	-3.93 ^{***}	-4.42 ^{***}	-4.00 ^{***}
CSR	-2.15 ^{**}	-2.76 ^{***}	-3.61 ^{***}	-2.55 ^{**}	-5.57 ^{***}
Radom Forest	-3.74 ^{***}	-2.17 ^{**}	-3.37 ^{***}	-2.85 ^{***}	-4.40 ^{***}
AdaLasso - RF	-4.02 ^{***}	-2.41 ^{**}	-3.38 ^{***}	-2.81 ^{***}	-5.33 ^{***}
Neural Network	-3.58 ^{***}	-3.39 ^{***}	-4.42 ^{***}	-4.74 ^{***}	-6.08 ^{***}
XgBoost	-1.26	-1.28	-3.96 ^{***}	-3.70 ^{***}	-4.14 ^{***}

Note: See <Table 5>.

〈Table 11〉 MCS(Model Confidence Set) Test Results for Imports Forecasting

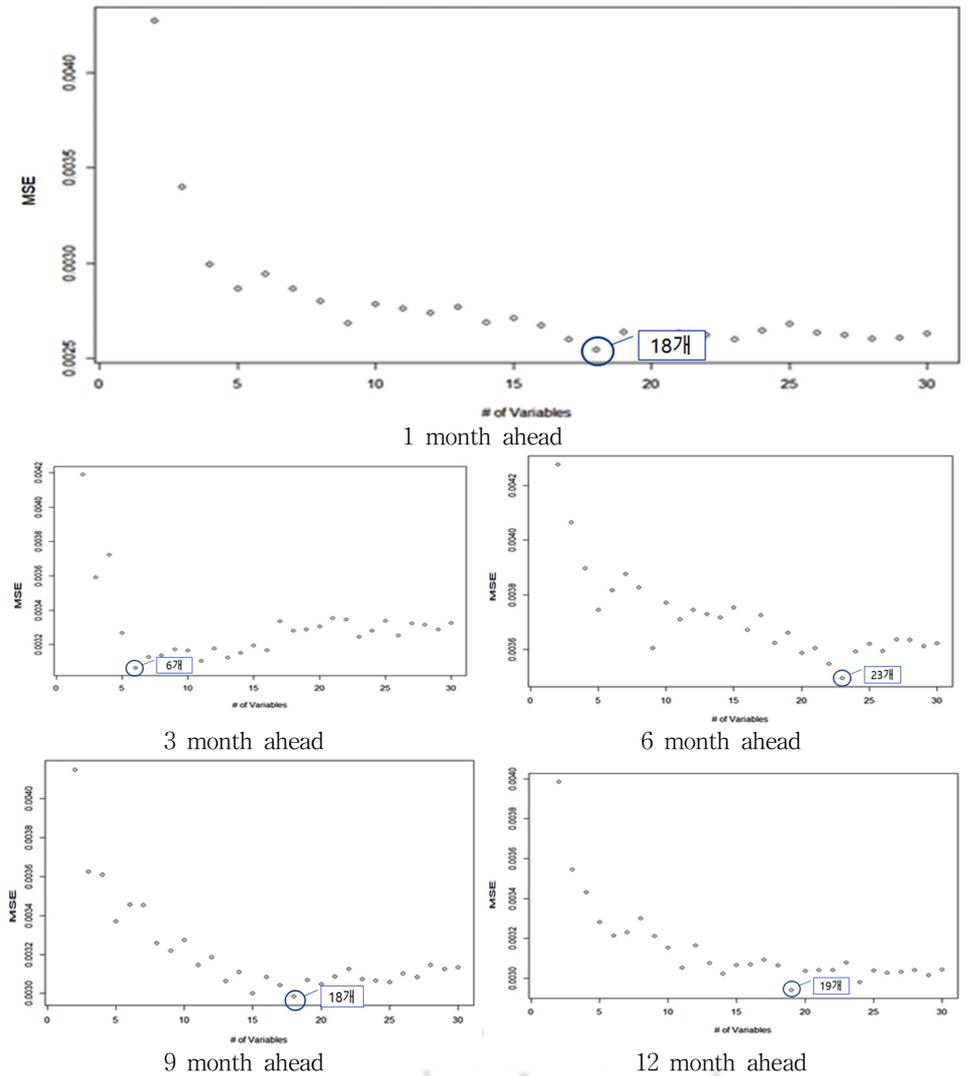
Model	1Month	3Month	6Month	9Month	12Month
AR					O
Ridge					
LASSO					
Adaptive Lasso					O
Elastic Net					O
Adaptive Elasticnet					
Target Factor					
CSR					
Random Forest					O
AdaLasso / RF					
Neural Network					
XgBoost					O
Boruta_selected_RF	O	O	O	O	O

Note: See <Table 6>.

〈Figure 4〉는 Boruta_selected_RF 모형에서 각 예측 대상기간별로 교차검증을 통해 최적 변수의 수를 산출하는 과정을 보여준다. 예측 대상기간별로 각각 18개, 6개, 23개, 18개, 19개의 변수가 선택되었고, 〈Table 12〉는 선택된 변수들의 목록을 제시한다. 수입증가율 예측에 중요한 변수 중 수출입 금액, 물량 및 물가지수 증가율의

과거값들을 제외하고 살펴보면, 모든 예측 대상기간에서 ‘에너지사용량’ 변수가 포함된 것을 확인할 수 있다. 에너지경제연구원의 『2021 에너지통계월보』에 따르면 우리나라의 에너지 수입의존도는 92.8%이며 우리나라의 10대 수입품목 중 원유는 2019년 1위, 2020년 2위를 차지하고 있다. 따라서 에너지사용량이 수입금액 증가율 예측에 중요 변수로 선택되는 것은 자연스러운 것으로 보인다.

<Figure 4> Number of Variables and In-Sample OOB MSE for Imports Forecasting



Note: See <Figure 3>.

〈Table 12〉 List of Selected Variables by Boruta Algorithm and Cross-Validation for Imports Forecasting

1Month		6Month	
X45	Energy Consumption(t-1)	X45	Energy Consumption(t-1)
X3	Import Quantity Index(t-1)	X212	Amount of Imports(US)(t-4)
X151	Energy Consumption(t-3)	X2	Export Quantity Index(t-1)
X107	Amount of Import(KR)(t-3)	X6	Current account balance(t-1)
X98	Energy Consumption(t-2)	X38	Operation ratio Index(t-1)
X109	Import Quantity Index(t-3)	X39	Shipment Index(t-1)
X1	Amount of Import(KR)(t-1)	X55	Export Quantity Index(t-2)
X21	Producer Price Index(JP)(t-1)	X3	Import Quantity Index(t-1)
X204	Energy Consumption(t-4)	X187	Sigapore(STI)(t-4)
X5	Import Price Index(t-1)	X37	Industrial Product Index(t-1)
X2	Export Quantity Index(t-1)	X126	Producer Price Index(KR)(t-3)
X55	Export Quantity Index(t-2)	X132	Germany(DAX)(t-3)
X108	Export Quantity Index(t-3)	X157	Spread(US)(t-3)
X29	Hong Kong(HSI)(t-1)	X123	Aluminium price(t-3)
X54	Amount of Import(KR)(t-2)	X122	Copper price(t-3)
X100	Industrial Product Index(JP)(t-2)	X28	Sigapore(STI)(t-1)
X28	Sigapore(STI)(t-1)	X54	Amount of Import(KR)(t-2)
X46	Employed(t-1)	X70	Aluminium price(t-2)
3Month		X151	Energy Consumption(t-3)
X45	Energy Consumption(t-1)	X112	Current account balance(t-3)
X1	Amount of Import(KR)(t-1)	X179	Producer Price Index(t-4)
X204	Energy Consumption(t-4)	X204	Energy Consumption(t-4)
X3	Import Quantity Index(t-1)	X78	Euro STOXX(t-2)
X159	Amount of Imports(US)(t-3)		
X165	Current account balance(t-4)		
9Month		12Month	
X212	Amount of Imports(US)(t-4)	X53	Amount of Imports(US)(t-1)
X162	Import Quantity Index(t-4)	X204	Energy Consumption(t-4)
X204	Energy Consumption(t-4)	X3	Import Quantity Index(t-1)
X112	Current account balance(t-3)	X106	Amount of Imports(US)(t-2)
X59	Current account balance(t-2)	X151	Energy Consumption(t-3)
X160	Amount of Import(KR)(t-4)	X45	Energy Consumption(t-1)
X161	Export Quantity Index(t-4)	X1	Amount of Import(KR)(t-1)
X106	Amount of Imports(US)(t-2)	X98	Energy Consumption(t-2)
X159	Amount of Imports(US)(t-3)	X2	Export Quantity Index(t-1)

9Month		12Month	
X151	Energy Consumption(t-3)	X108	Export Quantity Index(t-3)
X98	Energy Consumption(t-2)	X200	Equipment Investment Index(t-4)
X53	Amount of Imports(US)(t-1)	X56	Import Quantity Index(t-2)
X187	Sigapore(STI)(t-4)	X28	Sigapore(STI)(t-1)
X56	Import Quantity Index(t-2)	X155	Industrial Product Index(US)(t-3)
X108	Export Quantity Index(t-3)	X59	Current account balance(t-2)
X54	Amount of Import(KR)(t-2)	X81	Sigapore(STI)(t-2)
X3	Import Quantity Index(t-1)	X74	Producer Price Index(JP)(t-2)
X143	Industrial Product Index(t-3)	X12	Nickel price(t-1)
		X91	Operation ratio Index(t-2)

Note: See <Table 7>.

3. 에너지사용량 포함 여부에 따른 예측력 비교

수출입증가율 예측시 Boruta_selected_RF 모형에 에너지사용량이 설명변수로 사용되는데, 이를 제외한 경우 모형의 예측력이 어떻게 변화하는지를 살펴보았다. <Table 13>은 에너지사용량이 설명변수로 포함된 기존 Boruta_selected_RF 모형과 에너지사용량을 제외한 Boruta_selected_RF 모형의 RMSE와 MAE 그리고 Giacomini-White 검정 결과를 제시한다. 수출/수입, RMSE/MAE, 예측 대상기간을 고려한 모든 경우에 에너지사용량을 포함할 때 예측력이 우수한 것으로 나타났으며, Giacomini-White 검정은 이러한 예측력의 우수성이 수출 6개월, 수입 3개월과 12개월 제외하고는 모두 통계적으로 유의한 것을 보였다. 한편, 자기회귀 모형에 에너지사용량을 추가할 때의 예측력 변화도 살펴보았는데, 대체로 예측력이 향상되었으나 일부 예측 대상기간에는 예측력이 오히려 나빠지는 경우도 있는 것으로 나타났다. 따라서 에너지사용량이 수출입증가율 예측에 중요한 변수이지만, 모형에 따라 예측력을 향상시키는 정도는 상이하다는 점을 주의해야 할 것이다.

<Table 13> Comparison of Forecast Errors for the Boruta_selected_RF Model with or without Energy Consumption Variable

Boruta_selected_RF Model		1Month	3Month	6Month	9Month	12Month	
E X P O R T	RMSE	with energy consumption	0.0612	0.0677	0.0706	0.0657	0.0683
		without energy consumption	0.0651	0.0718	0.0739	0.0701	0.0715
	MAE	with energy consumption	0.0473	0.0526	0.0565	0.0507	0.0525
		without energy consumption	0.0485	0.0538	0.0575	0.0515	0.0530

Boruta_selected_RF Model			1Month	3Month	6Month	9Month	12Month
		without energy consumption	0.0509	0.0575	0.0583	0.0552	0.0551
	GW test	statistic	-2.46**	-2.67***	-1.55	-2.70***	-2.10**
I M P O R T	RMSE	with energy consumption	0.0498	0.0566	0.0563	0.0534	0.0542
		without energy consumption	0.0554	0.0591	0.0624	0.0634	0.0573
	MAE	with energy consumption	0.0394	0.0452	0.0446	0.0437	0.0430
		without energy consumption	0.0423	0.0456	0.0489	0.0502	0.0443
	GW test	statistic	-2.18**	-0.20	-2.95***	-3.25***	-0.91

Note: The table provides forecast errors for the Boruta_selected_RF model with or without the energy consumption variable. GW test means the Giacomini-White test. ** and *** indicate 5% and 1% significance level, respectively.

VI. 결론 및 시사점

본 연구에서는 수출입 증가율 예측을 위해 다양한 머신러닝 방법들을 활용하여 모형별 예측력을 분석하였다. 예측오차들을 비교한 결과, 보루타 알고리즘으로 변수 중요도의 순위를 도출한 후 교차검증을 통해 최적 변수의 개수를 결정하고 중요도가 높은 변수들을 정해진 개수만큼만 사용하여 랜덤 포레스트를 실시하는 모형이 모든 예측 대상기간에 있어 가장 예측력이 우수한 것으로 나타났다. 또한 Giacomini and White 검정과 Model confidence set 검정은 대부분 이 모형이 통계적으로 유의하게 예측력이 우수함을 보였다. 이처럼 이 모형의 예측력이 우수한 것은 이미 Medeiros et al.(2019)이 강조한 것처럼 랜덤 포레스트 모형이 변수들 간의 비선형(nonlinear) 관계를 모형화하는 데 뛰어난 점과 보루타 알고리즘이 변수 중요도의 순위를 도출하는 데 우수한 점 그리고 랜덤 포레스트에 사용할 중요도가 높은 변수들의 수를 교차검증을 통해 체계적으로 선택하는 방법의 장점 등이 결합하여 있기 때문으로 보인다.

수출입 증가율 예측에 중요한 변수로 채택된 변수 중 수출입 금액, 물량 및 물가지수 증가율의 과거값들을 제외하고 살펴보면, 수출과 수입 예측에서 동일하게 에너지 사용량의 과거값들이 선택되었다. 우리나라의 에너지 수입의존도가 높은 것을 고려할 때 수입증가율 예측에서 에너지사용량이 중요한 변수로 선택되는 것은 놀랍지 않은 것이고, 수출증가율 예측에도 에너지사용량이 중요한 변수로 선택되는 것은 에너지

사용량이 산업생산 동향을 통해 향후 수출 동향에 대한 유용한 정보를 제공하기 때문으로 보인다. 따라서 수출입 모니터링에 에너지사용량을 적극적으로 활용하는 방안을 검토할 필요가 있을 것이다.

참 고 문 헌

- 강기천(2008), “대일본·대중국 무역수지 결정요인 분석과 시사점”, 『월간 KIET 산업경제』, 산업연구원, 67-77.
- 거시건전성감독국(2019), “우리나라 수출에 영향을 미치는 주요 요인”, 조사연구자료 분야별 감독제도, 금융감독원.
- 고동환 외 12인(2020), “ICT 산업 중장기 전망(2021~2025년) 및 ICT 수출 대응전략”, 『방송통신정책연구』, 2020-0-01401, 정보통신정책연구원.
- 김도완·김동훈·이진기(2021), “우리나라 주식시장의 실물경제 대표성 분석: 산업별 비교를 중심으로”, 『조사통계월보』, 제75권 제5호, 한국은행, 16-43.
- 김민우(2022), “2022년 3/4분기 수출산업경기전망조사(EBSI)”, 『Trade Focus』, 2022년 14호, 한국무역협회.
- 김수이(2017), “국내 제조업부문의 에너지소비, 생산, 수출간의 인과관계 분석”, 『한국자원경제학회연구』, 제26권 제1호, 한국환경경제학회, 37-56.
- 김신중·최정일(2019), “수출물동량과 수출물가지수, 국고채금리가 수출금액에 미치는 영향”, 『융합정보논문지』, 제9권 제9호, 중소기업융합학회, 133-140.
- 김정우(2020), “머신러닝 기법 기반의 예측조합 방법을 활용한 산업 부가가치율 예측 연구”, 『한국콘텐츠학회논문지』, 제20권 제12호, 49-57.
- 김종성·김별화(2005), “우리나라의 제조업 해외직접투자가 수출에 미치는 영향 분석”, 『수은해외경제』, 7월호, 한국수출입은행, 4-18.
- 김치호·최요철·신현열(2000), “한국의 산업내무역: 변동추이와 결정요인”, 『경제분석』, 제6권 4호, 한국은행, 120-163.
- 박성근(2020), “전력사용량을 활용한 실시간 경기 모니터링 방안”, 『월간 KIET 산업경제』, 제260호, 산업연구원, 7-19.
- 백문영·김우환(2012), “한국의 경제성장과 전력수요간의 인과성에 관한 연구: 분기별 자료를 이용하여”, 『응용통계연구』, 제25권 제1호, 한국통계학회, 89-99.
- 변창욱·최용재·이상호·신현수(2009), “수출입 함수의 추정을 통한 우리나라 국별·산업별 수출입 결정요인 분석”, 『산업연구원 연구정책세미나』, 산업연구원, 1-257.

- 송민기(2022), “수출 급변동 조기예측모형: 코로나19 사례를 중심으로”. 『KIF 금융분석 보고서』, 한국금융연구원, 1-79.
- 안성배(2021), “2022년 세계경제 전망”, 『오늘의 세계경제』, 제21권 제18호, 대외경제정책연구원, 1-143.
- 조용성(2021), 『에너지통계월보 2021/04』, 제37권 4호, 에너지경제연구원, 1-129.
- 오형나(2011), “제조업부문 에너지 수요에 대한 연구”, 『정책연구시리즈』 2011-15, 한국개발연구원, 1-73.
- 유정호(1984), “상품군별 수출입함수의 추정”, *Journal of Economic Policy*, 제6권 제3호, 한국개발연구원, 101-119.
- 이동원 · 임성운(2019), “수출의 기초적 흐름 판단을위한 수출상황지수(ET-COIN) 개발”, 『조사통계월보』, 2019년 7월호, 한국은행, 16-31.
- 이재열 · 한희준(2001), “품목별 수출입 행태분석과 시사점”, 『조사통계월보』, 한국은행, 2001년 2월호, 한국은행.
- 이환호(1995), “엔달러환율 변동이 우리나라 지역별 수출입에 미치는 효과”, 『국제경제연구』, 제1권 제2호, 129-155.
- 임준형(2008), “한국의 무역과 경제성장의 인과관계 분석”, 『상업교육연구』, 제22권, 한국상업교육학회, 399-408.
- 정규철 외(2022), 『KDI 경제전망, 2022 상반기』, 제39권 제1호, 한국개발연구원
- 조재영 · 주지환 · 한인구(2021), “기계학습을 이용한 수출신용보증 사고예측”, 『지능정보연구』 제27권 제1호, 83-102.
- 최문정 · 김경근(2019), “선진국 수입수요가 우리나라 수출에 미치는 영향”, 『경제분석』, 제25권 제1호, 한국은행, 34-65.
- 한국경제연구원 경제정책팀(2022), “2022년 수출 증가율, 작년보다 크게 둔화될 전망“, 한국경제연구원 보도자료(1월 3일자).
- 홍성욱 · 신현수 · 박성근 · 김정현(2017), “국가별 · 경제권별 수출결정요인 분석을 통한 수출전망모형 구축 연구”, 『연구보고서』, 2017-837, 한국산업연구원, 3-139.
- 홍지상 외 6인(2021), “2021년 수출입 평가 및 2022년 전망”, 『Trade Focus』, 2021년 35호, 한국무역협회 국제무역통상연구원, 1-33.
- 홍지상 외 5인(2022), “2022년 상반기 수출입 평가 및 하반기 전망”, 『Trade Focus』, 2022년 13호, 한국무역협회 국제무역통상연구원, 1-33.
- Giacomini, R. and White, H.(2006), “Tests of conditional predictive ability”, *Econometrica*, 74(6), 1545-1578.
- Goldstein, M. and Khan, M. S.(1978), “The Supply and Demand for Exports : A

- Simultaneous Approach”, *The Review of Economics and Statistics*, 60(2), 275-286.
- Greenhalgh, C.(1990), “Innovation and Trade Performance in the United Kingdom”, *The Economic Journal*, 100(400), 105-118.
- Hansen, P. R., Lunde, A., and Nason, J. M.(2011), “The model confidence set”, *Econometrica*, 79(2), 453-497.
- Helkie, W. and Hooper, P.(1988), “The U.S. External Deficit in the 1980’s: An Empirical Analysis”, In R. Bryant, G. Holtham, and P. Hooper (eds.), *External Deficits and the Dollar: The Pit and the Pendulum*, Washington, D.C.: Brookings Institution.
- Kim, B. Y. and Han, H.(2022), “Multi-step-ahead forecasting of the CBOE volatility index in a data-rich environment: Application of random forest with Boruta algorithm”, *The Korean Economic Review*, 38(3), 541-569.
- Kursa, M. B. and Rudnicki, W. R.(2010), “Feature selection with the Boruta Package”, *Journal of Statistical Software*, 36, 1-13.
- Kreinin, M. E.(1967), “Price Elasticities in International Trade”, *Review of Economics and Statistics*, 49(4), 510-516.
- Landesman, M. and Pfaffermayer, M.(1997), “Technological Competition and Trade Performance”, *Applied Economics*, 29, 179-196.
- Medeiros, M. C., Vasconcelos, G. F., Veiga, A., and Zilberman, E.(2019), “Forecasting Inflation in a Data-rich Environment: The Benefits of Machine Learning Methods”, *Journal of Business & Economic Statistics*, 1-45.
- Kursa, M. B.(2010), “Boruta: A System for Feature Selection”, *Fundamenta Informaticae*, 101(4), 271-285.
- Muscatelli, V. A., Stevenson, A., and Montagna, C.(1995), “Modelling aggregate manufactured exports from some Asian newly industrialised economies”, *The Review of Economics and Statistics*, 77(1), 147-155.
- Sanli, D.(2019), “The long-run relationship between energy consumption and export sophistication in OECD countries”, *Ankara Haci Bayram Veli University Journal of the Faculty of Economics and Administrative Sciences*, 21, 342-361.
- Shakeel, M, Iqbal, M. M., and Majeed, M. T.(2014), “Energy consumption, trade and GDP: A case study of south Asian countries”, *The Parkistan Development Review*, 53(4), 461-476.

〈부록 2〉 추정 기간의 크기(window size)에 대한 강건성 검토

본문에서는 구간이동 방법을 적용하여 예측치를 생성할 때 추정 기간의 크기(window size)를 144개월(12년)로 설정하였는데, 추가로 추정 기간의 크기가 다른 경우 모형의 예측력이 어떻게 달라지는지 검토하였다. AR 모형과 몇 가지 주요 모형에 대해 추정 기간의 크기를 120개월(10년)과 96개월(8년)로 설정하여 예측력을 비교하였다. 예측 기간 및 예측치의 수는 본문의 경우와 같다.

아래 <Table A>는 수출증가율 예측시 추정 기간의 크기를 120으로 설정한 경우의 RMSE와 MAE를 나타낸다. 본문의 경우와 같이 Boruta_selected_RF 모형이 모든 예측 대상기간에서 가장 작은 RMSE와 MAE를 나타내었다. <Table B>는 수출증가율 예측시 추정 기간의 크기를 96으로 설정한 경우이고, 여전히 Boruta_selected_RF 모형이 모든 예측 대상기간에서 가장 작은 RMSE와 MAE를 나타내었다. <Table C>와 <Table D>는 수입증가율 예측시 추정 기간의 크기를 각각 120과 96으로 설정한 경우이고, 이들 경우에도 여전히 Boruta_selected_RF 모형이 모든 예측 대상기간에서 가장 작은 RMSE와 MAE를 나타내었다.

〈Table A〉 Forecast Errors for Exports: Window Size 120

	Model	1Month	3Month	6Month	9Month	12Month
RMSE	AR	0.0760	0.0804	0.0794	0.0742	0.0788
	Adaptive Lasso	0.0763	0.0847	0.0841	0.0772	0.0802
	Radom Forest	0.0728	0.0764	0.0821	0.0746	0.0734
	XgBoost	0.0712	0.0761	0.0896	0.0785	0.0725
	Boruta_selected_RF	0.0638	0.0702	0.0680	0.0688	0.0713
MAE	AR	0.0595	0.0632	0.0627	0.0567	0.0631
	Adaptive Lasso	0.0593	0.0668	0.0656	0.0594	0.0637
	Radom Forest	0.0568	0.0594	0.0641	0.0582	0.0571
	XgBoost	0.0544	0.0607	0.0712	0.0609	0.0554
	Boruta_selected_RF	0.0492	0.0548	0.0524	0.0533	0.0546

Note: Numbers in bold represents the minimum of forecast errors for each forecast horizon.

〈Table B〉 Forecast Errors for Exports : Window Size 96

	Model	1Month	3Month	6Month	9Month	12Month
RMSE	AR	0.0751	0.0807	0.0792	0.0745	0.0795
	Adaptive Lasso	0.0747	0.0950	0.0996	0.0926	0.0999
	Radom Forest	0.0749	0.0771	0.0825	0.0769	0.0763
	XgBoost	0.0722	0.0776	0.0869	0.0855	0.0829

	Model	1Month	3Month	6Month	9Month	12Month
	Boruta_selected_RF	0.0659	0.0689	0.0675	0.0685	0.0719
MAE	AR	0.0586	0.0639	0.0622	0.0573	0.0635
	Adaptive Lasso	0.0588	0.0716	0.0767	0.0746	0.0750
	Radom Forest	0.0592	0.0604	0.0650	0.0597	0.0594
	XgBoost	0.0558	0.0612	0.0705	0.0670	0.0648
	Boruta_selected_RF	0.0511	0.0543	0.0520	0.0539	0.0548

<Table C> Forecast Errors for Imports : Window Size 120

	Model	1Month	3Month	6Month	9Month	12Month
RMSE	AR	0.0642	0.0679	0.0697	0.0646	0.0646
	Adaptive Lasso	0.0637	0.0706	0.0712	0.0633	0.0628
	Radom Forest	0.0596	0.0647	0.0687	0.0631	0.0613
	XgBoost	0.0586	0.0634	0.0713	0.0679	0.0661
	Boruta_selected_RF	0.0530	0.0565	0.0566	0.0569	0.0547
MAE	AR	0.0502	0.0541	0.0557	0.0523	0.0524
	Adaptive Lasso	0.0500	0.0564	0.0574	0.0511	0.0496
	Radom Forest	0.0468	0.0518	0.0556	0.0508	0.0501
	XgBoost	0.0469	0.0511	0.0571	0.0561	0.0538
	Boruta_selected_RF	0.0413	0.0446	0.0445	0.0450	0.0434

<Table D> Forecast Errors for Imports : Window Size 96

	Model	1Month	3Month	6Month	9Month	12Month
RMSE	AR	0.0627	0.0679	0.0701	0.0640	0.0644
	Adaptive Lasso	0.0651	0.0742	0.0951	0.0831	0.0833
	Radom Forest	0.0610	0.0669	0.686	0.0628	0.0624
	XgBoost	0.0606	0.0663	0.7545	0.0658	0.0668
	Boruta_selected_RF	0.0515	0.0597	0.0591	0.0562	0.0564
MAE	AR	0.0491	0.0539	0.0559	0.0518	0.0521
	Adaptive Lasso	0.0499	0.0590	0.0711	0.0635	0.0583
	Radom Forest	0.0475	0.0532	0.0550	0.0506	0.0515
	XgBoost	0.0473	0.0525	0.0582	0.0528	0.0532
	Boruta_selected_RF	0.0414	0.0471	0.0458	0.0448	0.0438

KUKJE KYUNGJE YONGU
Volume 28 Number 4
December 2022

Forecasting Exports and Imports of South Korea Using Machine Learning Methods*

Nawon Jang** · Heejoon Han***

Abstract

This paper considers exports and imports forecasting in South Korea using various machine learning methods in a data-rich environment. The samples are from January 2000 to December 2021 (total 264 observations) and the forecast period is set to be the last ten years of the samples. Among thirteen models, one model outperforms the rest models for all forecast horizons, which was proposed by Kim and Han (2022). The method consists of the following three steps: 1) Using the Boruta algorithm, obtain the rankings of the covariates in high-dimensional data, 2) Choose the optimal number of the most important covariates via cross-validation, 3) Using only those selected covariates, implement the random forest method and produce a forecast. The tests by Giacomini and White (2006) and Hansen et al. (2009) show that the model provides significantly better forecasts. In particular, the energy consumption is selected as an important variable for exports and imports forecasting.

Key words: Exports, Imports, Forecasting, Machine Learning, Random Forest, Boruta Algorithm

JEL Classification Number: C32, C50, F16

Received 26 October 2022; Received in revised form 17 December 2022; Accepted 20 December 2022

* This paper is based on the first author's 2022 research project paper in the Department of Quantitative Applied Economics at Sungkyunkwan University.

** First Author, Assistant Manager, KOTRA, E-mail : jangnw@kotra.or.kr

*** Corresponding Author, Professor, Department of Economics/Quantitative Applied Economics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul 03063, E-mail: heejoonhan@skku.edu