
REC 가격예측 분석

2017. 12.

모두의 연구소 DataLab

I. 서론

1.1. 연구 배경

전력통계정보시스템(EPSIS)에 따르면, 우리나라의 2016년 발전원별 전력거래량 비중은 석탄이 40.6%로 가장 높고, 원자력이 30.3%, LNG가 22.0%, 그리고 태양광 등을 포함한 신재생에너지가 3.8%로 그 뒤를 잇는 것으로 나타났다. 최근 20년간의 추이를 살펴보면 신재생에너지는 2002년에 0.4%에서 그쳤으나, 2016년에는 3.8%까지 증가한 것으로 나타났으나, 석탄과 원자력이 꾸준히 3~40%대를 오르내리면서 우리나라 에너지 거래의 대부분을 차지하는 것으로 나타났다.

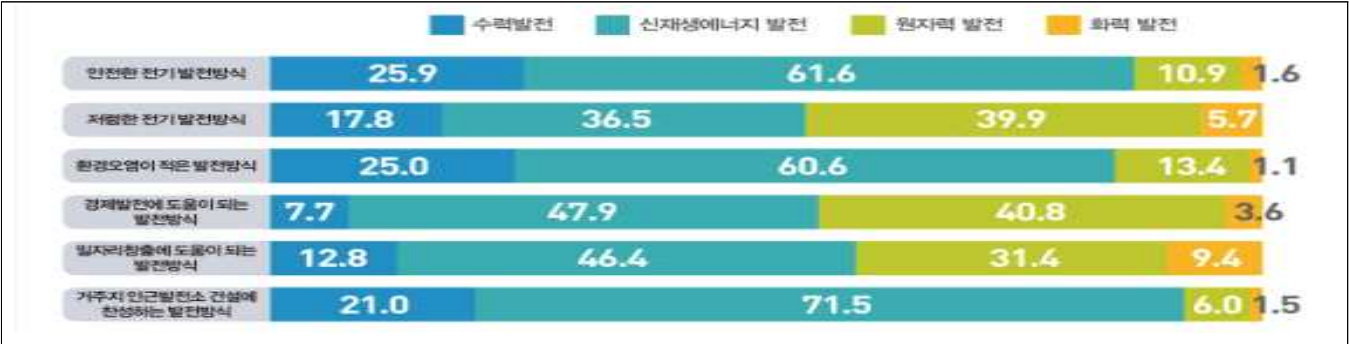
<표1> 발전원별 전력거래량

(단위:GWh, %)

연도	원자력	석탄	LNG	유류	양수	신재생
2001	80,528(40.5)	83,308(41.9)	17,040(8.6)	15,504(7.8)	1,528(0.8)	880(0.4)
2010	141,894(32.2)	191,007(43.3)	85,173(19.3)	11,841(2.7)	2,774(0.6)	8,177(1.9)
2016	154,310(30.3)	206,610(40.6)	111,814(22.0)	13,280(2.6)	3,618(0.7)	19,353(3.8)

(출처 : 연간 전력시장 통계, 전력통계정보시스템)

이러한 발전원별 에너지 정책 추세는 최근, 여러 사회·환경·경제적인 영향으로 인해 변화가능성이 감지되고 있다. 첫째, 가장 많은 에너지를 공급하고 있던 석탄 발전의 경우, 미세먼지 증가의 주범으로 인식되면서 정부 차원의 석탄 발전 감축이 예정되어 있다. 금년 5월에 한국전력이 개최한 미세먼지 대책회의에서는 석탄 발전으로 인한 미세먼지를 50% 감축하는 것으로 발표하였으며, 이를 위해 30년 이상 노후화된 석탄발전기 8기를 발전 중지하고, 30년 이상 노후석탄발전소 폐지시기를 단축하는 등 석탄 발전의 비중을 줄이는 대책을 발표하였다. 둘째, 약 30%의 전력을 공급하고 있던 원자력 발전의 경우에도, 체르노빌 사태, 후쿠시마 사태 등과 같은 방사능 유출로 인한 환경 재앙이 언제든지 발생할 수 있다는 위기감으로 인해 정부는 찬반논란에도 불구하고 중장기적인 탈원전 정책을 밀어붙이고 있다. 그 결과로 금년 6월, 고리원전1호기가 영구 정지되었고, 신고리 5·6호 원자로 건설을 재검토하는 수순을 밟고 있기도 하다. 셋째, 이와 같은 탈석탄, 탈원전 움직임으로 인해 주목받는 것은 신재생에너지이다. 신재생에너지는 발전방식이 안전하고, 환경오염이 적으며, 경제발전·일자리 창출에 긍정적인 영향을 미치는 등 사회·환경·경제적으로 긍정적인 면이 많은 것으로 인식되어 지고 있다. 이 중에서도 특히, 태양광 발전은 환경문제로 인한 위험성에서 벗어날 수 있는 중요한 대안으로 각광받고 있고, 이로 인해 서울시는 최근 “태양의 도시”, “원전하나 줄이기” 정책 등을 통해 태양광 발전을 도시의 자생적 에너지 발전 정책으로 내세우고 있는 상황이다.



[그림1] 발전원별 이용선호도

(출처 : 한국원자력문화재단)

이러한 가운데, 한국에너지공단은 태양광 발전 및 신재생에너지 거래 활성화를 위해 고정가격계약 경쟁입찰 제도를 통한 태양광 에너지 입찰을 금년부터 실시하였다. 올해 처음 도입된 고정가격계약 경쟁입찰 제도는 태양광 발전사업자들에게 계통한계가격(SMP) 변동에 따른 가격 안정성을 확보할 수 있게 하고, 안정적인 투자환경을 제공받을 수 있도록 하기 위한 목적으로 마련되었다.

◇ 신재생에너지 현물시장과 고정가격계약 경쟁입찰시장

신재생에너지 인증서(REC) 시장은 크게 현물시장, 고정가격계약 경쟁입찰 시장, 그리고 수의계약 시장이 있다. 현물시장의 경우 매월 많은 횟수에 걸쳐 거래되고 있는 REC 거래시장을 의미하며, 최근에는 마치 주식시장에서 거래를 하는 것처럼 빠르게 가격이 변동되는 특징을 보인다.

고정가격계약 경쟁입찰 시장은 이와는 다르게, 1년에 2차례에 걸쳐 정부의 주도를 통해 대량의 거래가 한꺼번에, 그리고 장기간 물량이 한번 계약되는 형태로 이루어진다. 또한, 계통한계가격(SMP)과 신재생 공급인증서(REC) 합산가를 기반으로 신재생 발전 단가를 책정하는 방식으로 특정 발전소와의 계약을 통해 그 수익을 20년간 계통한계가격(SMP)과 함께 고정시켜 거래를 진행한다.

다만, 정부의 기대와 달리 태양광 발전 사업자들의 기대치보다 낮은 입찰상한가와 20년이라는 다소 긴 계약기간으로 인해 사업자들로부터 외면받아 역대 최저 경쟁률로 마감되고 말았다. 지난 4월에 실시하였던 입찰에서 최종 입찰경쟁률은 역대 최저였던 2012년 상반기 입찰(2.5대1)보다 낮은 1.9대1로 집계된 것이다. 당시, 입찰 평균가격은 상한가(19만1330원)보다 불과 약 1만원 낮은 18만1595원이었고, 이것은 향후 20년간의 물가상승률을 고려하였을 때 전혀 매력적이지 않은 가격이라는 것이 중론이었다. 긴 계약기간으로 인한 안정적 투자환경 마련이라는 당근이 낮은 가격임에도 많은 사업자가 입찰에 참여하게 하여 시장이 활성화될 것이라는 정부의 기대와 달리 사업자들이 고정가격계약 경쟁입찰 시장을 외면하고 현물시장으로 떠나게 하는 결과를 낳고 말았다. 탈석탄, 탈원전의 대안으로 신재생에너지 거래의 활성화를 도모했던 정부 입장은 매우 곤혹스러워졌음이 분명하다.

1.2. 연구 목적

앞서 설명한 바와 같이, 고정가격계약 경쟁입찰제도는 발전사업자들의 외면으로 시작부터 난관에 부딪히고 말았다. 신재생에너지가 석탄과 원자력의 대체제로 주목받고 있는 이 시점에서 보다 면밀한 접근이 필요한 시점이며, 고정가격계약 경쟁입찰 시장의 합리적 가격을 산출하고, 안정적인 수급을 가능하게 하는 것이 신재생에너지 및 태양광 발전시장의 거래 활성화를 위해 중요함이 분명하다. 특히, 상한가 결정을 위한 합리적인 방법의 도입과 가격의 투명한 공개가 태양광 발전사업자를 다시 고정가격계약 입찰시장으로 유입시킬 수 있는 중요한 요인이 될 것임에 틀림없다.

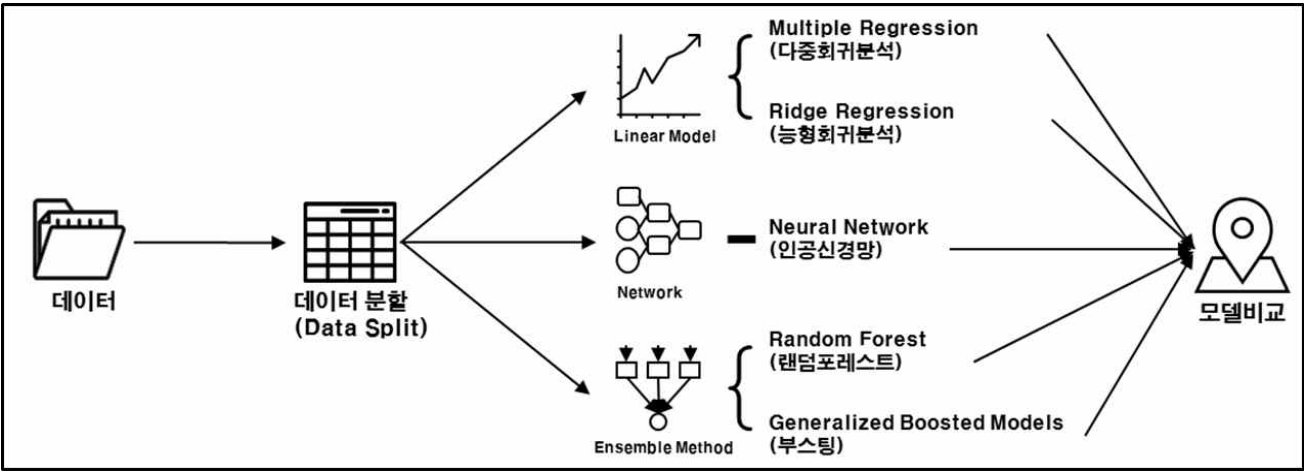
이에, 본 연구에서는 이러한 문제 인식을 토대로 REC 가격 예측을 통해 고정가격 계약입찰 시장에서 적용해야 할 상한가 적용 모형을 제시하여 신재생에너지 및 태양광 발전의 거래 활성화 정책에 시사점을 제시하고자 한다. 이를 위해 신재생에너지 공급인증서인 REC(Renewable Energy Certificate) 가격 변동을 가장 잘 예측할 수 있는 예측 모형 구축을 목표로 하며, 이것은 고정가격계약 경쟁입찰제도의 상한가격을 합리적으로 도출하기 위한 모형이 되어 사업자들을 시장으로 불러올 계기를 마련할 수 있을 것이다.

본 연구에서 차별적인 점은 기존 관행이나 일반적인 통계 분석 방법에서 벗어나 머신러닝 방법을 새롭게 도입하여 가장 우수하게 가격 예측이 가능한 모형을 찾고자 하였다는 점이다. 이것은 정부의 탈석탄, 탈원전 정책을 성공시키는 데에 중요한 시발점이 될 것으로 기대한다.

1.3. 연구 방법

본 연구의 절차로는 먼저 EDA(Exploratory Data Analysis)를 통해 데이터의 구조를 파악한 뒤, 다양한 통계모형과 머신러닝 모형을 적용하여 가장 적합한 모델을 찾아낸다. 모델 선정기준은 예측력을 나타내는데 많이 사용하는 RMSE(Root Mean Square Error)값을 활용하였다. 데이터에 적용한 모형 중 가장 RMSE가 작은 모형을 선택하고 찾아낸 모형의 적합성을 판정한다.

이에, 예측성능이 가장 우수한 모형을 찾는 과정은 데이터 분할, 모델링, 평가 순으로 진행될 수 있다. 각 방법에 대한 그림은 아래 [그림2]과 같이 표현할 수 있다.



[그림2] 예측 모형 개발 과정

본 연구를 위해 신재생에너지센터, 전력통계정보시스템에서 관련 데이터를 수집하였다. 사용한 데이터는 date를 제외한 한가지의 반응 변수와 28가지 설명변수로 이루어져 있다. 데이터의 샘플 수는 277개이다. 변수들의 구성은 아래의 <표2>와 같다. 먼저 반응변수는 REC 가격을 뜻하는 price이고, 연속형 변수이다. price 변수는 58,700부터 174,267의 범위를 가지며 평균값은 122,791이고 표준편차는 18,268.42이다. 28가지 설명변수는 3개의 범주형 변수와 18개의 연속형 변수와 6개의 비연속형 변수로 이루어진다. 범주형 변수를 살펴보면, type변수는 육지(location_in), 제주(location_out) 두 가지의 요인을 가진다. date_end는 12월에 거래가 일어났으면 1, 아니면 0의 값을 가진다. agenda1은 정책이 바뀐 시점에는 1, 아니면 0의 값을 가진다. 서울 열린데이터 광장에서 수집된 물가지수 데이터는 시간적 특성을 반영하고 있기 때문에 물가지수 상승률이라는 변수를 새로 생성하였다. 또한 케이웨더에서 수집된 날씨 데이터는 전라남도에 국한하여 수집하였다. 전라남도는 전국 최고 수준의 일사량과 해상에서 불어오는 양질의 바람, 진도 울돌목 등의 빠른 조류, 풍부한 농수산물 부산물 등은 태양광·해상풍력·조류·바이오매스 등 신재생에너지 개발에 최적의 입지 조건을 갖추고 있다. 대부분의 태양광 발전소는 전라남도에 위치하고 있기 때문에 전라남도 날씨 데이터를 사용하였다.

<표2> 변수의 조작적 정의

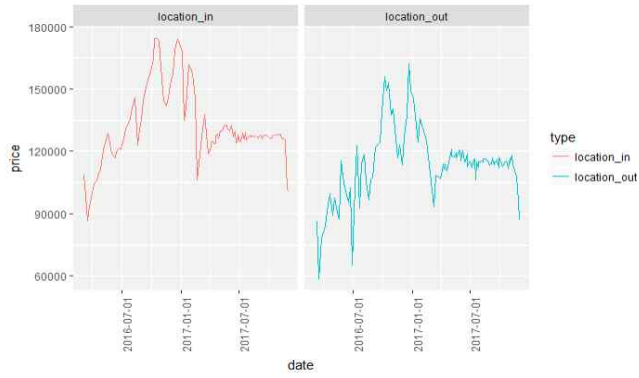
변수 이름	설명	타입	출처
date	2016-03-04 ~ 2017-11-30	날짜	-
date_end	매년의 마지막달	범주형	-
agenda1	정책의 변화 유무		-
type	육지(in)/제주(out)		신재생에너지센터

price	rec 인증서 가격(원)	연속형	전력통계정보시스템
before_count	한 시점 전 체결수량		
SMP	전력시장 가격(원)		
생활물가지수	물가지수		서울 열린데이터 광장 (서울시 생활물가지수 통계)
식품			
식품_이외			
전월세			
전월세포함_생활물가지수			
생활물가지수_up	물가지수의 상승률		한국석유공사
식품_up			
식품_이외_up			
전월세_up			
전월세포함_생활물가지수_up			
Dubai	원유가격		케이웨더
Brent			
WTI			
min_Temperature_sum	이전 거래부터 현 거래일 사이의 전라남도 최저온도	비연속형	
max_Temperature_sum	이전 거래부터 현 거래일 사이의 전라남도 최고온도		
Precipitation_sum	이전 거래부터 현 거래일 사이의 전라남도 강수량		
snow_sum	이전 거래부터 현 거래일 사이의 전라남도 적설량		
구름많음	이전 거래부터 현 거래일 사이의 전라남도 날씨		
구름조금			
흐림			
눈			
맑음			
비			

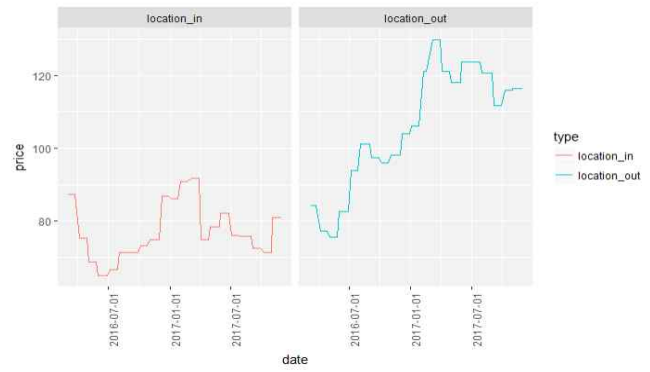
II. 연구 결과

2.1. EDA(Exploratory Data Analysis)

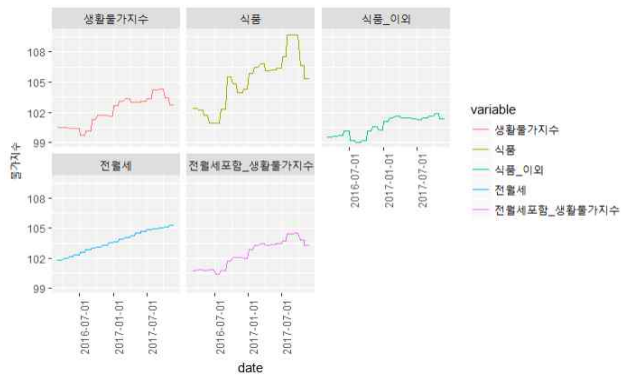
REC 거래 데이터를 본격적으로 분석하기 전에 EDA를 통해 데이터가 어떤 구조를 가지고 있는지 시간별 경과는 어떠한지 알아보았다. 먼저 각각의 설명변수와 반응변수와의 경향을 살펴보기 위해 각 변수마다의 시간별 추세를 그려보았다. [그림3]은 육지(location_in)와 제주(location_out)지역의 REC 가격 추이를, [그림4]는 육지와 제주지역의 SMP(계통한계가격)의 가격 추이를 그리고 있다. [그림5]는 물가지수, [그림6]은 물자지수의 상승률, [그림7]은 원유, [그림8]은 체결수량, [그림9]는 날씨, [그림10]은 기상 추이를 나타내고 있다. [그림3]을 보면 육지지역과 제주지역의 경향성이 비슷하고 일자마다 동일한 가격 차이를 가지고 있다. [그림3]과 [그림4]를 살펴보면 입찰시장의 고정가격의 영향으로 SMP값이 내려갈 때 REC값이 내려가고, SMP값이 오를 때 REC값이 오르는 것을 알 수 있다. 둘 사이에 음의 상관관계가 존재한다. 물가지수와 원유가격은 상승 추이를 보이고, 기상 데이터는 1년마다 비슷한 추이를 보이는 것을 알 수 있다.



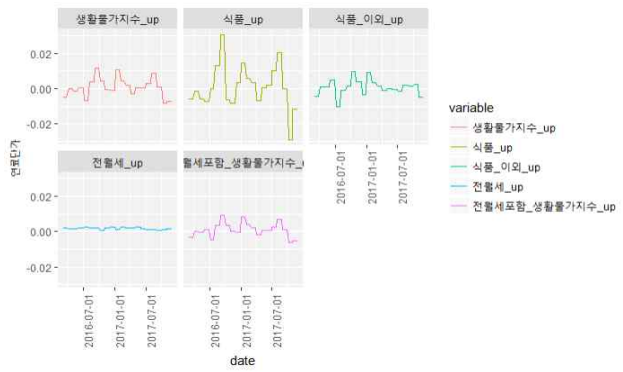
[그림3] type별 rec 가격 추이



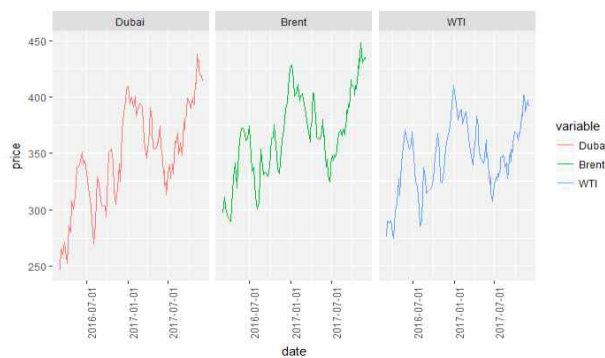
[그림4] type별 smp 가격 추이



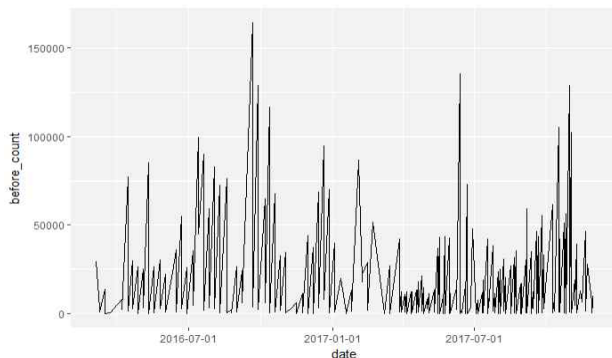
[그림5] 물가지수 추이



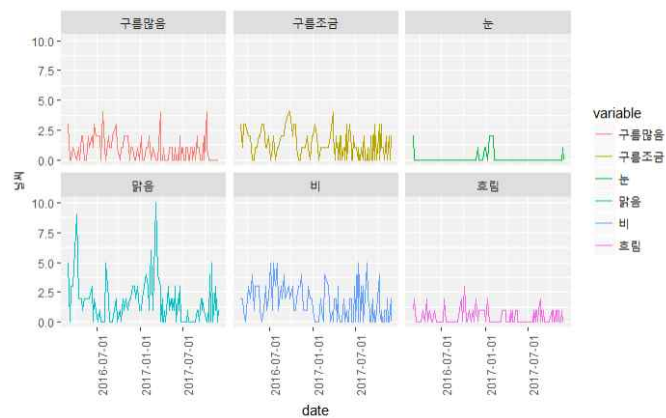
[그림6] 물가지수 상승률 추이



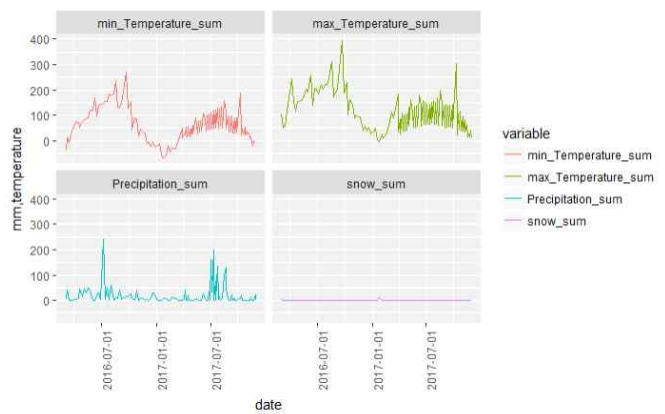
[그림7] 원유 추이



[그림8] 체결수량 추이



[그림9] 날씨 추이



[그림10] 기상 추이

2.2. 최적의 모형 구축

2.2.1 데이터 분할 (Data Split)

올바른 예측력을 평가하기 위해서 데이터를 모형 적합용 데이터(Training data)와 성능 평가용 데이터(Test data)로 나눴다. 본 데이터의 경우, 2016년 3월 4일부터 2017년 11월 30일까지의 기간 내 자료로 시계열 특성을 가지고 있다. 그러므로 데이터를 분할할 때, 시간적 특성을 반영하기 위해 2017년 6월 1일을 기준으로 이전 시점의 데이터는 모형 적합용 데이터로, 이후의 데이터는 평가용 데이터로 사용하였다.

2.2.2 모델링 및 평가 (Modeling & Assessment)

최종 예측 모형의 후보로 통계적 모형인 다중회귀모형(Multiple Regression)과 능형회귀모형(Ridge Regression) 그리고 머신러닝 모형인 인공신경망(neural network), 랜덤 포레스트(random forest), 부스팅(Generalized Boosted Model) 등을 고려하였다. 각 모형의 정의 및 특징은 다음과 같다.

통계적 모형 중, 다중회귀모형은 변수의 선형결합을 통해 식을 추정하는 모형이다. 그렇기에 수치적으로 변수의 영향을 파악할 수 있고, 회귀식을 통해 직관적인 해석이 가능하다는 장점을 가지고 있다. 다만, 머신러닝 모형에 비해 일반적으로 예측성능이 떨어진다고 알려져 있다. 능형회귀모형은 각 회귀식의 계수에 대해 일정한 패널티를 부과한 모형으로, 일반적으로 다중회귀모형에 비해 높은 예측력을 보이는 모형으로 알려져 있다.

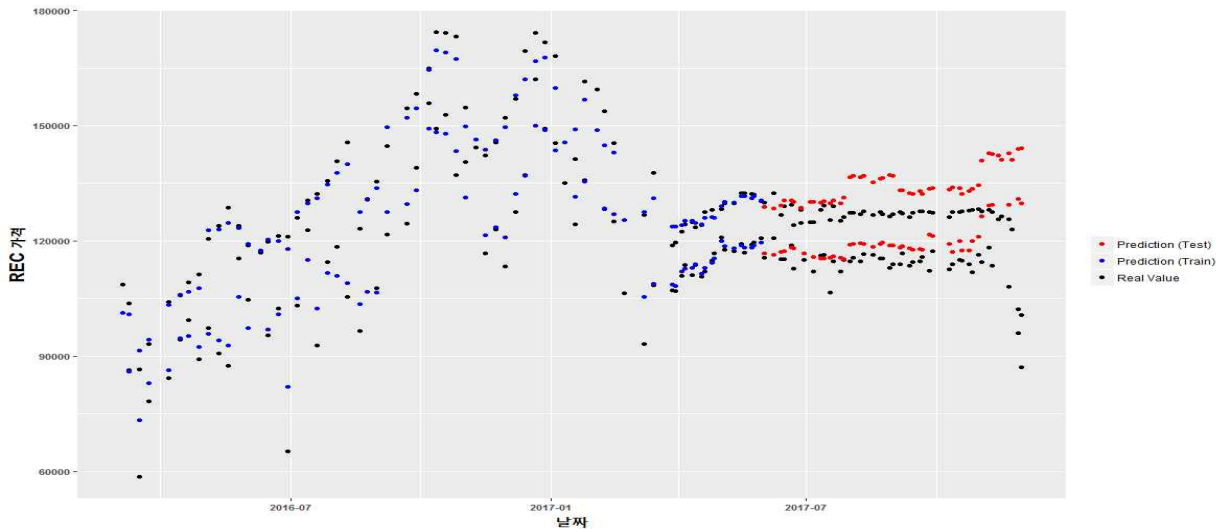
머신러닝 모형 중, 인공신경망은 다중회귀모형과 유사하게 변수의 선형 결합을 통해 추정된 모형이다. 하지만, 활성화함수와 은닉층을 통해 변수 간 비선형 관계를 추정하는 장점을 가지고 있다. 반면, 랜덤 포레스트와 부스팅은 앞선 모형들과는 달리 여러 모형의 결과를 바탕으로 예측을 하는 앙상블 기법이다. 즉, 여러 트리 기반의 모형(Tree based Model) 혹은 선형 모형(Linear Model)의 결과값의 평균을 통해 예측을 함으로 단일 모형 보다 우수한 예측 성능을 보인다는 장점을 가지고 있다.

각 모형의 예측력은 평균 제곱근 오차(Root Mean Square Error; RMSE)을 통해 평가하였다. RMSE는 예측한 값과 실제 값 사이의 차이를 정의하는 측도로, 평균 차이라는 의미를 가지고 있다. 이에 대한 결과는 아래 <표3>과 같다.

<표3> 모형 비교

Model	RMSE (Training set)	RMSE (Test set)
다중회귀모형	8710.3	24789.02
능형회귀모형	10230.9	22095.22
인공신경망	9215.947	30438.27
랜덤 포레스트	4871.11	11806.13
부스팅	9268.206	12177.16

<표3>을 보면, 랜덤 포레스트가 Test set의 RMSE가 가장 작은 것을 알 수 있다. 즉, 랜덤 포레스트가 최적의 예측 모형이라고 결론 내릴 수 있다. 랜덤 포레스트의 예측 결과 그림은 아래 [그림11]를 통해 확인 가능하다.



[그림11] 예측한 값과 실제 값 비교 그림

[그림11]에서 검은 점은 실제 값을 파란 점은 모형 적합용 데이터에 대한 적합 값, 빨간 점은 검증용 데이터에 대한 예측 값이다. 검증용 데이터에 대한 예측 값을 보면, 전반적으로 실제 값의 트렌드를 잘 따르고 있는 것을 알 수 있다. 하지만, 마지막 부분은 예측을 잘 못하는 것으로 확인되는데, 이는 신재생에너지 투자의 인기로 인해 투자자가 몰려 발전량 자체가 많아졌기 때문에 REC가격이 하락하였기 때문이다. 이런 현상은 데이터화하지 못하였기에 해당 파트에 대한 예측이 좋지 않은 것으로 확인 할 수 있다.

Ⅲ. 결론 및 시사점

본 연구는 신재생에너지 공급인증서인 REC 가격변동 예측을 통해 고정가격계약 경쟁입찰 제도가 안착되기 위한 가격정책을 수립하고자 하였다. 즉, 합리적인 REC 상한가격을 제시할 수 있는 최적의 모형을 도출하고자 하였다. 이를 위해 전력통계정보시스템, 서울열린데이터광장 등의 공공데이터베이스에서 추출한 REC 가격, 물가지수, 물가지수상승률, 원유가격, 날씨 등의 공공데이터를 활용하였다. 이 공공데이터를 활용하여 인공신경망, 다중회귀모형과 같이 REC 가격변동을 예측하는 여러 모형을 구축하여 실험하였고, 각 모형의 RMSE값을 비교하여 최적의 모형을 선택하였다. 이 중, 랜덤포레스트 모형의 예측력 RMSE값이 11806.13으로 가장 작아 REC 가격변동을 가장 잘 예측하는 모형은 랜덤포레스트인 것으로 밝혀졌다.

본 연구는 신재생에너지의 확대를 반대하는 전문가들의 주장이 주로 신재생에너지 생산비용이 높다는 데에 초점이 맞춰져 있다는 측면에서 시사점을 제시하고 있다. 즉, 위 랜덤포레스트 모형을 활용하여 신재생에너지 가격을 합리적으로 예측하여 고정가격계약 경쟁입찰 제도의 안정화를 이끌어 낼 수 있다면, 고비용의 문제도 해결할 수 있을 것이라는 점이다. 이에, 보다 적극적으로 본 모형을 활용하여 정책에 활용할 필요가 있다.

다만, 본 연구의 실험을 위한 데이터수가 부족하였다는 점은 보다 정확하고 신뢰성있는 연구를 하는 데에 한계로 여겨지며, 이를 보완할 수 있는 추가적인 연구가 필요할 것으로 판단된다.